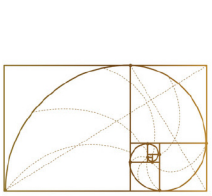
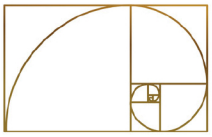
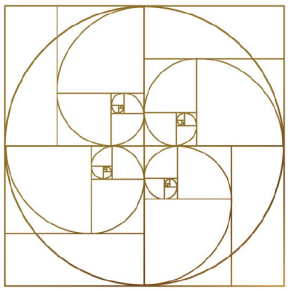
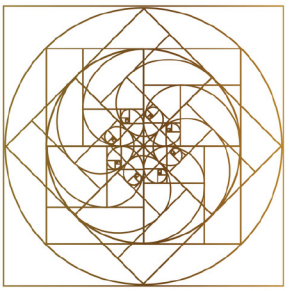
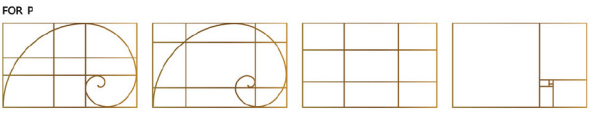
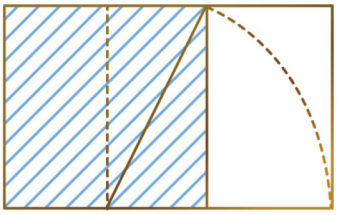






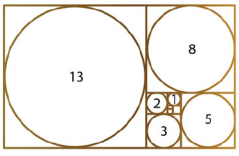


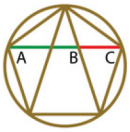


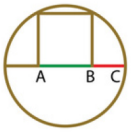


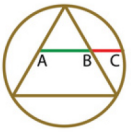
Джулиан Хэйвилл

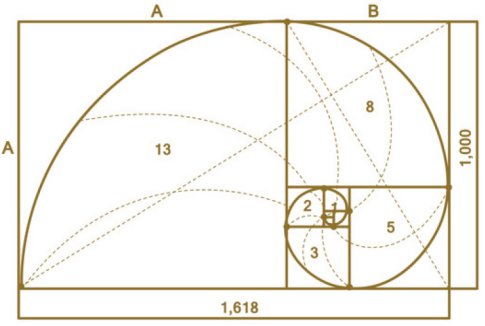
ЗАМЕЧАТЕЛЬНЫЕ МАТЕМАТИЧЕСКИЕ КРИВЫЕ

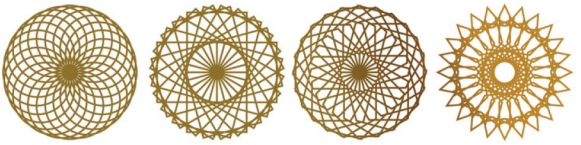












ДМК ПРЕСС
ИЗДАТЕЛЬСТВО

Джулиан Хэйвил

Замечательные математические кривые

Curves for the Mathematically Curious

An anthology of the unpredictable, historical, beautiful and romantic

Julian Havil

PRINCETON
UNIVERSITY
PRESS

Замечательные математические кривые

Антология непредсказуемого, исторического, чарующего и романтического

Джулиан Хэйвил

ДМК ПРЕСС
ИЗДАТЕЛЬСТВО
Москва, 2025

УДК 512.77+514.75
ББК 22.147+22.151.1
Х99

Джулиан Хэйвил

Х99 Замечательные математические кривые: антология непредсказуемого, исторического, чарующего и романтического / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2025. – 244 с.: ил.

ISBN 978-5-93700-241-9

В этой книге собраны описания десяти математических кривых, тщательно отобранных за их значимость, интересность и красоту. В каждой главе читатель найдет историю и определение кривой, а также узнает о красивой и часто неожиданной математической основе, связанной с ее созданием и эволюцией. Книга построена так, что все желающие могут превратиться в исследователей, просто вооружившись карандашом и бумагой.

Издание адресовано широкому кругу любителей математики и может быть полезно преподавателям и руководителям математических кружков.

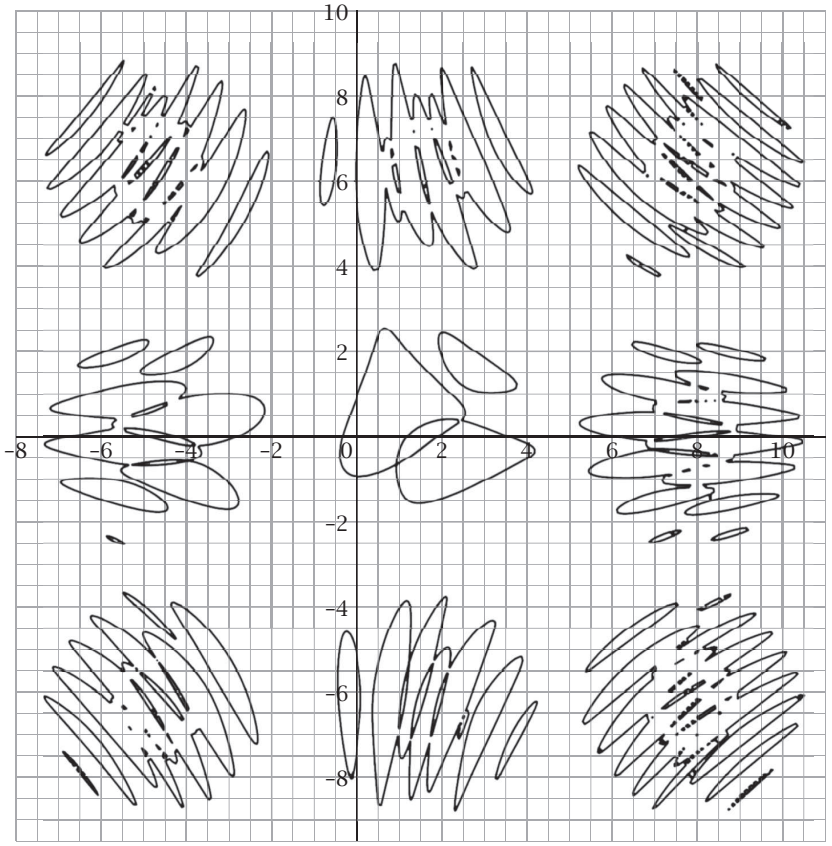
УДК 512.77+514.75
ББК 22.147+22.151.1

ISBN (анг.) 978-0691180052
ISBN (рус.) 978-5-93700-241-9

183153372066978579413813161120260490562039261666571156496660784216681464987082595826896046168627213434548372679840038786077305207518343443634540572248370231663948491083610315772942321071840819721770562297162281857627692632972691882708869450959589825219731434884638041327897169895079816448554544832720638356333727176101707225620202965299323169703097145697089728882897417655843882915475422113296794313979058634917269805716255157589738999733963072304148150107992987929819107147019732608436103300513924880753098551972506167401269270937671266188410480659227201446263656645880093188441974569407252078778357045344519701744886288892504495215623349724873006156070667898686586113351446361025922643828994675060373878142299994781453388798220455149189828149573769262183570084156175842373190888117948110340670167576139867939940026365080561899096527318016921837200333962991129556876915437179446547088918221394605276007079940311919423102484304241791034442158371922678254716491446006249954027937067017930177496148955132974758394021214827419334231139279560498382225097164485227341414478

Посвящается тем, кто придет, в память о тех, кто ушел

Математические закорючки



$\sin(\sin x + \cos y) = \cos(\sin xy + \cos x)$

Festina lente

Hâtez-vous lentement, et sans perdre courage
Vingt fois sur le métier votre ouvrage:
Polissez-le sans cesse, et le repolissez;
Ajoutez quelquefois, et souvent effacez.

Спешите медленно

Спешите медленно и, мужество утроя,
Отделывайте стих, не ведая покоя,
Шлифуйте, чистите, пока терпенье есть:
Добавьте две строки и вычеркните шесть[1].

Никола Буало-Депрео (1636–1711)

[1] Перевод Э. Л. Линецкой.

Здесь тоже мерзко: гнусный шум и пьянка, гольф, бургундское, охота, математика, Ньюмаркет, состязания и разные бесчинства…

Лорд Байрон, из письма к мисс Элизабет Пигон
с описанием Тринити колледжа, Кембридж,
26 октября 1807

Оглавление

Предисловие от издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте *www.dmkpress.com*, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу *dmkpress@gmail.com*; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу *http://dmkpress.com/authors/publish_book/* или напишите в издательство по адресу *dmkpress@gmail.com*.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг – возможно, ошибку в основном тексте или программном коде, – мы будем очень благодарны, если вы сообщите нам о ней. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Если вы найдете какие-либо ошибки в коде, пожалуйста, сообщите о них главному редактору по адресу *dmkpress@gmail.com*, и мы исправим это в следующих тиражах.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательство «ДМК Пресс» очень серьезно относится к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу *dmkpress@gmail.com*.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

Предисловие

Я здесь только собрал чужие цветы,
а от меня самого – только нитка, которой они связаны.

— Мишель де Монтень[1]

В 1943 г. прославленный британский военачальник, фельдмаршал сэр Арчибальд Уэйвелл, опубликовал антологию стихотворений под названием «Чужие цветы». При этом почитатель поэзии, солдат – обладатель изумительной памяти и разностороннего образования, заметил, что слово «антология» происходит от греческого *anthos*, означающего «цветок», и *logia*, производного от *legein*, означающего «собрание»: стало быть, «собрание цветов». Быть может, букет. В этом контексте слово «цветы» употребляется как метафора подборки образцов чего-то, обычно, но не всегда связанного с поэзией, составленного одним человеком и неизбежно отражающего его вкусы. Лорд Уэйвелл составлял свой букет из образцов поэзии XIX в., мы же – из длинной истории математики, а конкретно – математических кривых.

Любая кривая, имеющая название, имеет и историю, и мы поставили себе цель рассказать историю некоторых кривых в манере, которая, как мы надеемся, понравится читателям с математическим складом ума. Но каких и почему именно их? Составляя первый вариант списка кривых, мы прошерстили множество записных книжек, статей, книг и сайтов, как сорока, выбирая все блестящее только для того, чтобы впоследствии признать, что не все то золото, что блестит, и отказаться от некоторых находок. В результате осталось десять кривых – и нет сомнения, что кто-то другой выбрал бы иные.

Составитель антологии наслаждается ничем не стесненной свободой личного выбора, но и несет ответственность за целостность результата; выбор не должен быть случайным, но и полностью предсказуемым он тоже быть не может, а должен покоиться на надежном основании разумного критерия отбора. Для нас самым важным критерием была двумерность кривой, что исключает пространственные кривые и кривые, расположенные в абстрактных математических мирах, – быть может, мы вернемся к ним в другой раз. Кроме того, у кривой должна быть история – интересная и, быть может, неожиданная, – и она должна быть важной в своем роде, по какому бы критерию эту важность ни оценивали. Это может быть ее форма, или угловой коэффициент, или площадь под ней, или еще какая-то значимая характеристика. А быть может, это была первая кривая такого вида. И еще хорошо бы, чтобы она была красивой. Иногда фортуна становилась на нашу сторону, и по закону неожиданных последствий на свет выплывали дополнительные интересные кривые. В част-

[1] Перевод Н. Я. Рыковой.

ности, квадратриса Гиппия сразу привлекла наше внимание и прошла последующий отбор, потому что дает возможность вернуться в мир Древней Греции и красивых построений с помощью циркуля и линейки – умение, в значительной степени утраченное в перегруженных программах математики средней школы. Мы надеемся, что, прочитав эту главу, хотя бы некоторые читатели обратят свои взоры к счастливым воспоминаниям об этих задачах, и, чем черт не шутит, даже захотят порешать еще какие-то.

Но антологии часто судят в большей степени не по тому, что включено, а по тому, что опущено, и мы опустили многое из того, что другим может показаться существенным, – прежде всего конические сечения с их долгой историей и вездесущностью. Но нам кажется, что о них уже было написано так много, так подробно и так по-разному, а их история столь величественна, что наш вклад либо занял бы слишком много страниц, заполненных рабским повторением чужих работ, либо оказался бы слишком поверхностным. Да и то сказать – не каждая же антология стихотворений включает сонеты Шекспира. Однако же конические сечения упоминаются не раз, не в одной главе и по разным поводам. Мы также признаем другие упущения и сожалеем о них, но надеемся, что читатели нас поймут.

У антологий есть особенность – их редко читают подряд. Так и с этой книгой. Главы можно читать в любом порядке, независимо друг от друга, но изредка можно встретить наложение методов, и мы надеемся, что читатель простит вытекающие отсюда незначительные повторы. Не придумав никакого разумного критерия, мы решили расположить главы в порядке возрастания количества слов.

Итак, мы приглашаем читателя присоединиться к нам в этом прихотливом и эклектичном математическим приключении, в котором мы близко познакомимся с Пабло Пикассо, Георгом II, принцем Альбертом, супругом и консортом королевы Виктории, святой инквизицией, императором Священной Римской империи (Фридрихом II) и многими математиками, жившими на протяжении тысячелетий, а также одним никогда не жившим, профессором Онесимом Дюраном. И неизбежно мы встретимся с Леонардом Эйлером. Не считая одной главы, мы не собирались касаться работ этого разностороннего гения из восемнадцатого столетия, но его имя само собой всплывало куда чаще, чем мы могли предвидеть в начале проекта. Ссылки на его работы, начинающиеся буквой «E», взяты из *индекса Энестрёма*, названного в честь шведского математика Густава Энестрёма, составившего авторитетный список наиболее важных работ Эйлера.

Мы берем на себя ответственность за все прокравшиеся в текст ошибки, приносим свои извинения за них и заверяем читателей, что сделали все, что в наших силах, чтобы их устранить. Имея это в виду, мы и поместили в начало книги французский оригинал и английский (русский) перевод отрывка из поэмы Никола Буало «Поэтическое искусство»: оно довольно хорошо итожит процесс написания книг.

Благодарности

Говорят, что книги сами себя пишут. Это ложь. Книги сами себя не пишут. Вам даже трудно представить себе, сколько размышлений, исследований, боли в пояснице, времени и работы требуется для их написания.

Нил Гейман «Дым и зеркала», АСТ, 2023

Именно так. А еще нужна существенная помощь тех, кто немало сделал, чтобы книга увидела свет, но, как правило, остается в тени; имен этих людей вы не найдете на обложке, но без них книга, которую вы сейчас держите в руках, не появилась бы. В этом небольшом разделе я хочу хотя бы отчасти исправить это положение.

Прежде всего огромное спасибо Вики Кирн, которая недавно ушла с поста исполнительного редактора отдела математики и компьютерных наук в издательстве Принстонского университета. Вики была редактором этой и четырех моих предыдущих книг, всегда невозмутимая, доброжелательная, вдохновляющая и обаятельная. Желаю ей как можно дольше наслаждаться заслуженным отдыхом и исполнения всех желаний. Ее место заняла Сюзанна Шумейкер, которая курировала последние этапы производства и следила, чтобы не возникло никаких проблем, о чем мечтает каждый автор. Также различные немаловажные обязанности выпали на долю Доран Бучча, Пэм Шниттер, Кэтлин Чиоффи и Жаклин Пуарье и, без сомнения, многих других, кого я забыл упомянуть. Спасибо вам всем.

Джон Уэйнрайт и С. Кларк из компании T&T Productions Ltd снова взмахнули своими волшебными палочками и превратили «оконченную» рукопись в готовую книгу, разница между которыми куда больше, чем мне хотелось бы признать. Спасибо им за терпение, учтивость и очевидный опыт общения с авторами, которые знают, чего хотят, но не всегда понимают, как этого добиться. Да и можно ли этого добиться вообще.

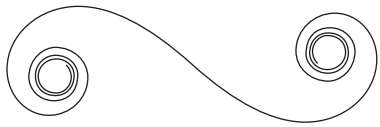
Наконец, спасибо семье и друзьям, основной вклад которых заключался просто в «присутствии». Понадобилась бы целая страница, чтобы перечислить всех, но во главе двух партий стоят моя жена, Энн, и мой давний друг, Колин. «Человек не остров», как заметил Джон Донн, и все они являются частью континента, на котором я живу, – к моему великому счастью и радости.

Что такое кривая? Всякий знает, что такое кривая, пока не выучится математике настолько, что вконец запутается в бесконечных исключениях.

— Феликс Клейн (1958)

Глава 1

Спираль Эйлера



Почему именно эта кривая?

С этой кривая связана кое-какая любопытная математика, у нее есть различные неожиданные приложения. А еще она такая восхитительно элегантная. И наша любимая.

1.1. Необычная параметризация...

Стандартная практика выражения кривой в параметрической форме $x = g(t)$, $y = h(t)$ дает формулы для обычных характеристик: углового коэффициента, площади под кривой, длины дуги и кривизны. В общепринятых обозначениях:

- угловой коэффициент: $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{y'}{x'}$;
- площадь под кривой: $\displaystyle\int_{t_1}^{t_2} yx'\,\mathrm{d}t$;
- длина дуги $\displaystyle\int_{t_1}^{t_2} \sqrt{x'^2 + y'^2}\,\mathrm{d}t$;
- кривизна $\dfrac{x'y'' - y'x''}{(x'^2 + y'^2)^{3/2}}$,

где штрих обозначает производную по параметру t. Если даны дифференцируемые функции $g(t)$ и $h(t)$, то основная проблема – можно ли вычислить интегралы в замкнутой форме.

Нас интересует то, что на первый взгляд может показаться весьма экзотическим примером параметризации:

$$x = x(t) = \int \cos f(t)\,\mathrm{d}t = \int_0^t \cos f(u)\,\mathrm{d}u,$$

$$y = y(t) = \int \sin f(t)\,\mathrm{d}t = \int_0^t \sin f(u)\,\mathrm{d}u,$$

где начальное значение параметра равно 0, а $f(u)$ – произвольная дифференцируемая функция от u. Способная поначалу испугать форма интеграла на самом деле упрощает все формулы, кроме площади под кривой:

$$\begin{aligned} x' &= x'(t) = \cos f(t) \\ y' &= y'(t) = \sin f(t) \end{aligned} \quad \text{и} \quad \begin{aligned} x'' &= -f'(t)\sin f(t) \\ y'' &= f'(t)\cos f(t), \end{aligned}$$

что дает

- $\dfrac{dy}{dx} = \dfrac{\sin f(t)}{\cos f(t)} = \tan f(t);$
- $s = \displaystyle\int_0^t \sqrt{\cos^2 f(u) + \sin^2 f(u)}\, du = t;$
- $\kappa(t) = \dfrac{f'(t)\cos^2 f(t) + f'(t)\sin^2 f(t)}{[\cos^2 f(t) + \sin^2 f(t)]^{3/2}} = f'(t).$

Как видим, параметр t – в точности длина дуги, а кривизна в точке, соответствующей параметру t, равна $f'(t)$ или, после интегрирования,

$$f(t) = \int \kappa(t)\, dt = \int^t \kappa(u)\, du.$$

Поэтому мы можем заменить абстрактный параметр t длиной дуги кривой s и переписать ее в параметрическом виде следующим образом:

$$x = x(s) = \int_0^s \cos\left(\int^u \kappa(t)\, dt\right) du,$$

$$y = y(s) = \int_0^s \sin\left(\int^u \kappa(t)\, dt\right) du.$$

Читатель может убедиться, что эти уравнения описывают прямую линию (ось x), когда $\kappa(t) = 0$, и окружность ($x^2 + (y - 1)^2 = 1$), когда $\kappa(t) = 1$. Следующий естественный шаг – взять $\kappa(t) = t$ или эквивалентно $\kappa(s) = s$, что приводит к кривой, кривизна которой линейно возрастает с ростом длины дуги. Простейшие параметрические уравнения этой кривой имеют вид:

$$x = x(s) = \int_0^s \cos \tfrac{1}{2}u^2\, du,$$

$$y = y(s) = \int_0^s \sin \tfrac{1}{2}u^2\, du.$$

Такая кривая должна спирально завиваться внутрь, потому что ее кривизна монотонно возрастает. В итоге мы получаем *спираль Эйлера*, показанную на рис. 1.1 и являющуюся предметом настоящей главы.

При наличии потворствующего вашим целям графопостроителя можно с удовольствием потратить кучу времени на эксперименты с другими вариантами $\kappa(t)$. Например, кривая на рис. 1.2 получена при $\kappa(t) = t^2$, а на рис. 1.3 – при $\kappa(t)\cos t - t\sin t$ (и, следовательно, $f(t) = t\cos t$).

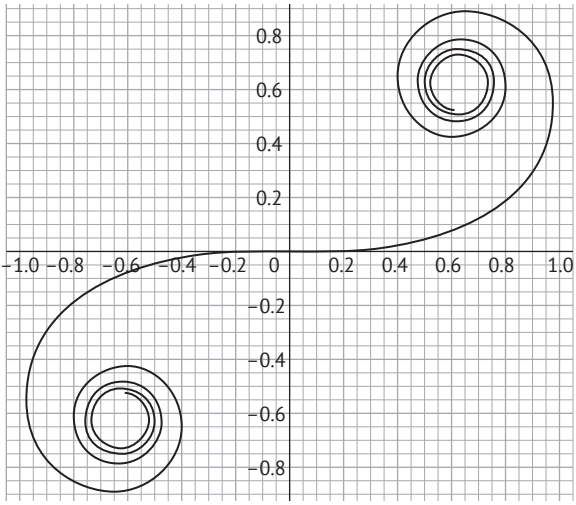


Рис. 1.1. Спираль Эйлера

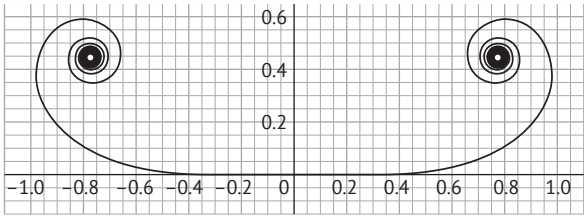


Рис. 1.2. Шезлонг

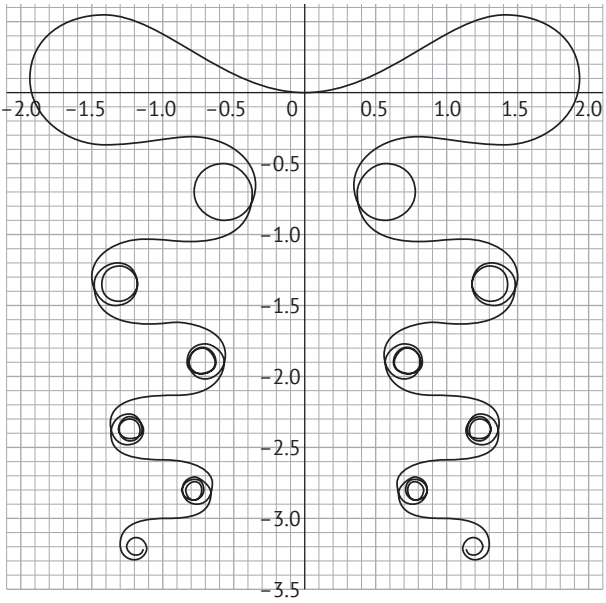


Рис. 1.3. Элегантное безумие

Таким образом, кривая, являющаяся предметом этой главы, дает пример довольно экзотической, на первый взгляд, параметризации с участием длины дуги и кривизны, но, как мы теперь видим, такое построение вовсе не выглядит странным. И даже напротив.

1.2 ... НО ПРИ ЭТОМ ЕСТЕСТВЕННАЯ

Если математическому результату присваивают название, значит считают его важным; если названию предшествует прилагательное «фундаментальный» или «основной», значит важность многократно увеличивается, так как результат занимает центральное место в отдельной теории или даже в математике в целом. Так обстоит дело с *основной теоремой о плоских кривых*, для строгой формулировки которой необходимо привлечь некоторые абстракции, но суть состоит в том, что «кривая определяется своей кривизной». То есть если задана начальная точка на плоскости и функция кривизны, то кривая определена однозначно.

Итак, предположим, что задана функция кривизны $\kappa(s)$, параметризованная длиной дуги, s. Тогда мы имеем две примитивные величины, т. е. не зависящие от таких внешних обстоятельств, как система координат или репер. Они называются *внутренними* переменными, и, помимо длины дуги, есть еще одна: *тангенциальный угол*, ψ.

Это угол, который каждая касательная к кривой составляет с некоторым фиксированным направлением. Обычно его нормируют, так что $s = 0$ при $\psi = 0$, а в качестве направления принимают положительное направление оси x. Из рис. 1.4 ясно, что $\mathrm{d}y/\mathrm{d}x = \operatorname{tg}\psi$. Более того, – и этот стандартный результат легко доказывается, – имеем $\mathrm{d}\psi(s) / \mathrm{d}s = \kappa(s)$; на самом деле это даже не *результат*, а *определение* кривизны как (нормированной) скорости вращения касательной к кривой.

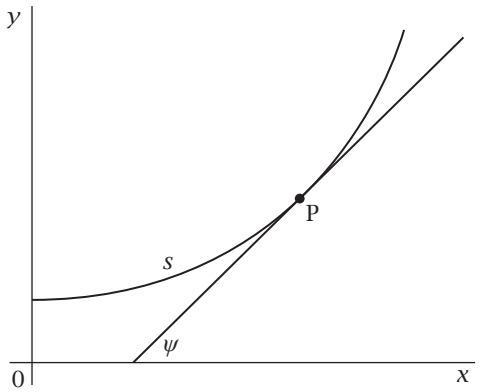


Рис. 1.4. Внутренние координаты

Далее, по определению

$$\frac{\mathrm{d}s}{\mathrm{d}x} = \sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2} = \sqrt{1 + \operatorname{tg}^2\psi} = \sec\psi,$$

поэтому $\mathrm{d}x / \mathrm{d}s = \cos\psi$, и, так как

$$\frac{dy}{dx} = \frac{dy}{ds} \times \frac{ds}{dx} = \operatorname{tg}^2 \psi,$$

должно иметь место равенство $dy / ds = \sin \psi$.

Теперь у нас имеется необходимый аппарат для достижения цели: вывести уравнение кривой, зная ее кривизну.

Пусть кривая параметризована длиной дуги:

$$x = g(s)$$
$$y = h(s).$$

Напишем

$$\frac{dx}{ds} = g'(s) \quad \text{и} \quad \frac{dy}{ds} = h'(s).$$

Тогда имеем следующую систему дифференциальных уравнений, описывающую кривую в терминах внутренних координат:

$$\kappa(s) = \frac{d\psi(s)}{ds},$$
$$g'(s) = \cos \psi(s),$$
$$h'(s) = \sin \psi(s).$$

Таким образом, зная $\kappa(s)$, мы сначала находим $\psi(s)$, а затем параметрические уравнения кривой.

Мы уже переходили от определенных интегралов к неопределенным и продолжим это занятие, снова используя переменную как предел и не забывая о произвольной постоянной. Тогда общее решение первого уравнения имеет вид:

$$\psi(s) = \int \kappa(s)\, ds = \int_0^s \kappa(t)\, dt + \psi_0.$$

Это означает, что

$$g'(s) = \cos\left(\int_0^s \kappa(t)\, dt + \psi_0 \right) \quad \text{и} \quad h'(s) = \sin\left(\int_0^s \kappa(t)\, dt + \psi_0 \right),$$

и, стало быть, общее решение таково:

$$g(s) = \int \cos\left(\int_0^s \kappa(t)\, dt + \psi_0 \right) ds = \int_0^s \cos\left(\int_0^u \kappa(t)\, dt + \psi_0 \right) du + x_0,$$

$$h(s) = \int \sin\left(\int_0^s \kappa(t)\, dt + \psi_0 \right) ds = \int_0^s \sin\left(\int_0^u \kappa(t)\, dt + \psi_0 \right) du + y_0.$$

Теперь сделаем явным влияние этих произвольных постоянных интегрирования, воспользовавшись хорошо известными тождествами из элементарной тригонометрии:

$$\cos(\theta + \varphi) = \cos\theta \cos\varphi - \sin\theta \sin\varphi,$$
$$\sin(\theta + \varphi) = \sin\theta \cos\varphi + \cos\theta \sin\varphi.$$

Получаем

$$\begin{aligned} g(s) &= \int_0^s \cos\left(\int_0^u \kappa(t)\,\mathrm{d}t + \psi_0\right)\mathrm{d}u + x_0 \\ &= \int_0^s \cos\left(\int_0^u \kappa(t)\,\mathrm{d}t\right)\cos\psi_0 - \sin\left(\int_0^u \kappa(t)\,\mathrm{d}t\right)\sin\psi_0\,\mathrm{d}u + x_0 \\ &= \cos\psi_0 \int_0^s \cos\left(\int_0^u \kappa(t)\,\mathrm{d}t\right)\mathrm{d}u \\ &\quad - \sin\psi_0 \int_0^s \sin\left(\int_0^u \kappa(t)\,\mathrm{d}t\right)\mathrm{d}u + x_0, \end{aligned}$$

$$\begin{aligned} h(s) &= \int_0^s \sin\left(\int_0^u \kappa(t)\,\mathrm{d}t + \psi_0\right)\mathrm{d}u + y_0 \\ &= \int_0^s \sin\left(\int_0^u \kappa(t)\,\mathrm{d}t\right)\cos\psi_0 + \cos\left(\int_0^u \kappa(t)\,\mathrm{d}t\right)\sin\psi_0\,\mathrm{d}u + y_0 \\ &= \cos\psi_0 \int_0^s \sin\left(\int_0^u \kappa(t)\,\mathrm{d}t\right)\mathrm{d}u \\ &\quad + \sin\psi_0 \int_0^s \cos\left(\int_0^u \kappa(t)\,\mathrm{d}t\right)\mathrm{d}u + y_0, \end{aligned}$$

что удобно переписать в матричной форме:

$$\begin{pmatrix} g(s) \\ h(s) \end{pmatrix} = \begin{pmatrix} \cos\psi_0 & -\sin\psi_0 \\ \sin\psi_0 & \cos\psi_0 \end{pmatrix} \begin{pmatrix} \displaystyle\int_0^s \cos\left(\int_0^u \kappa(t)\,\mathrm{d}t\right)\mathrm{d}u \\ \displaystyle\int_0^s \sin\left(\int_0^u \kappa(t)\,\mathrm{d}t\right)\mathrm{d}u \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

Первая матрица в произведении описывает вращение против часовой стрелки вокруг начала координат на угол ψ_0, а последний вектор – параллельный перенос. Мы показали, что с точностью до этих двух изометрий наша кривая однозначно определена параметрическими уравнениями:

$$\begin{aligned} x = g(s) &= \int_0^s \cos\left(\int_0^u \kappa(t)\,\mathrm{d}t\right)\mathrm{d}u, \\ y = h(s) &= \int_0^s \sin\left(\int_0^u \kappa(t)\,\mathrm{d}t\right)\mathrm{d}u, \end{aligned}$$

в которых параметром является длина дуги. Эти уравнения – зримое воплощение основной теоремы о плоских кривых, они подтверждают, что кривизна действительно определяет кривую. Памятуя об этом, мы приходим к выводу, что спираль Эйлера – кривая, естественно определяемая простейшим нетривиальным соотношением между кривизной и длиной дуги: $\kappa = s$. И тем не менее открыта эта кривая была совсем не так – и не Эйлер первым занялся ею.

1.3. Проблема

Когда в семье два Якоба, три Иоганна, два Даниила и пять Николаев (и это не считая вариантов написания имен), очень легко перепутать одного члена семейства Бернулли с другим. Эта математическая династия существовала на протяжении целого столетия, и девять входивших в нее сыновей и внуков знамениты своим вкладом в различные разделы математики и связанных с ней наук. Но первое появление нашей кривой нужно искать в работах Якоба I (или Джеймса I, или Жака I) – патриарха династии, хотя не он первым оценил ее по-настоящему.

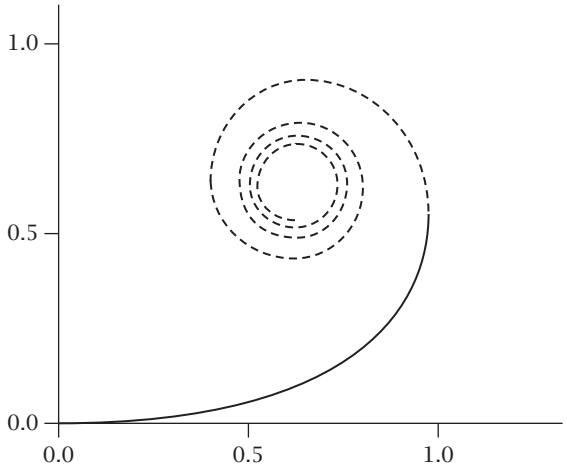


Рис. 1.5. Кривая Бернулли

Основополагающая публикация 1694 г. (Bernoulli 1694) стала кульминацией предпринятого Якобом Бернулли исследования того, что сегодня мы называем *проблемой консольной балки*: тонкая горизонтальная балка пренебрежимо малой массы, закрепленная на одном конце и нагруженная на другом, принимает форму кривой – но какой кривой? Данный им ответ мы теперь называем *эластикой*: эта кривая имеет самостоятельное значение в различных областях и тесно связана с нашей спиралью. В конце своей работы автор по своему обыкновению поставил ряд задач для читателей, одну из которых можно считать обратной к уже решенной:

> «Найти, какую кривизну должна иметь тонкая полоска, чтобы она распрямилась горизонтально под действием веса, приложенного к одному концу».

Балка остается закрепленной на одном конце, но теперь имеет искривленную, загнутую вверх форму, которая превращается в горизонтальную прямую под действием силы, приложенной к другому концу. В заметке, датированной тем же годом, Бернулли показал, что решил свою задачу, и привел внутреннее уравнение кривой в виде $a^2 = sR$, где a – постоянная, а R – радиус кривизны кривой, определенный как величина, обратная кривизне κ; таким образом, кривизна этой кривой пропорциональна длине дуги. Его рассуждение очень

лаконично, оно больше напоминает набросок подхода к решению, чем доказательство, а потому не слишком вразумительно – эту точку зрения выразил в 1744 г. его племянник Николай, который готовил работу дяди к публикации и отметил, что «я не считаю этот факт установленным». На рис. 1.5 кривая Бернулли изображена сплошной линией, доходящей до точки, в которой становится вертикальной. К ее продолжению он не проявил интереса, предоставив неподражаемому Леонарду Эйлеру возможность привести убедительные аргументы, показавшие, что ее продолжением является пунктирная спираль.

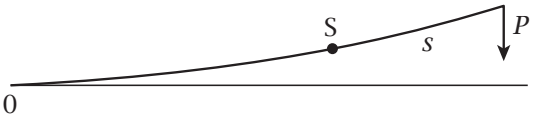


Рис. 1.6. Рассуждение Эйлера

В том же 1744 г. Эйлер опубликовал работу (см. Эйлер 1744, (E65)), которая даже по его высоким стандартам является необычайной по охвату предмета. В ней есть два приложения, в первом из которых, *Additamentum 1*, речь идет об упругих кривых, и в разделах 51 и 52 рассматривается эта задача. Использованное Эйлером сочетание чистой математики и физики является более ясной версией того, что, как нам представляется, было доказательством Бернулли.

Рассмотрим рис. 1.6 и предположим, что кривизна балки в принадлежащей ей точке S до и после приложения изгибающей силы равна $\kappa_1(s)$ и $\kappa_2(s)$ соответственно. Также предположим, что S отстоит на расстояние s вдоль балки от точки приложения силы. В современной терминологии влияние силы описывается уравнением $M = \kappa EI$, где M, E и I – момент силы, модуль Юнга и второй момент площади балки (относительно ее нейтральной оси) соответственно. Положим $M = Ps$, мы вслед за Эйлером измерим вклад кривизны балки:

$$\kappa = \frac{Ps}{EI} = \frac{s}{a^2}.$$

Это означает, что $\kappa_2(s) = \kappa_1(s) - \kappa$, а так как требуется, чтобы в конечный момент балка была прямой линией, то $\kappa_2(s) = 0$, и потому

$$0 = \kappa_1(s) - \frac{s}{a^2} \quad \text{и} \quad \kappa_1(s) = \frac{1}{r} = \frac{s}{a^2}.$$

Из этого внутреннего уравнения Эйлер, опуская подробности, вывел параметрические уравнения кривой[1]:

$$x = \int \cos\frac{s^2}{2a^2}\,\mathrm{d}s = \int_0^s \cos\frac{u^2}{2a^2}\,\mathrm{d}u$$
$$y = \int \sin\frac{s^2}{2a^2}\,\mathrm{d}s = \int_0^s \sin\frac{u^2}{2a^2}\,\mathrm{d}u.$$

При этом он, по-видимому, рассуждал следующим образом:

[1] Мы поменяли местами sin и cos, чтобы записать уравнения в их современной форме.

$$\int \frac{1}{r}\,\mathrm{d}s = \int \frac{s}{a^2}\,\mathrm{d}s \quad \Rightarrow \quad \int \frac{\mathrm{d}\psi}{\mathrm{d}s}\,\mathrm{d}s = \frac{s^2}{2a^2} \quad \Rightarrow \quad \psi = \frac{s^2}{2a^2}.$$

Полагая

$$\frac{\mathrm{d}x}{\mathrm{d}s} = \cos\psi \quad \text{и} \quad \frac{\mathrm{d}y}{\mathrm{d}s} = \sin\psi,$$

приходим к нужному нам решению.

Параметрические уравнения позволяют продолжить кривую после точки обращения в вертикаль, и мы видим, что она завивается в бесконечную спираль.

Эйлер признавал, что ни один из приведенных выше интегралов нельзя вычислить в замкнутой форме, поэтому использовал разложение в ряд и почленное интегрирование и получил полезные представления в виде бесконечных рядов:

$$x = s - \frac{s^5}{2! \times 5} + \frac{s^9}{4! \times 9} - \frac{s^{13}}{6! \times 13} + \cdots,$$
$$y = \frac{s^3}{1! \times 3} - \frac{s^7}{3! \times 7} + \frac{s^{11}}{5! \times 11} - \frac{s^{15}}{7! \times 15} + \cdots.$$

Он также сделал следующее наблюдение:

> «Теперь из того факта, что радиус кривизны монотонно убывает с увеличением дуги, ясно следует, что кривая не может простираться бесконечно, даже если длина дуги стремится к бесконечности. Поэтому кривая принадлежит классу спиралей и после бесконечного числа витков сворачивается в некоторую точку, являющуюся ее центром, каковую точку очень трудно найти из этого построения. Таким образом, следует признать, что анализ не дает ни малейшего преимущества тому, кто захочет найти метод, позволяющий хотя бы приближенно вычислить значения интегралов в случае, когда s бесконечна. Эта задача представляется достойной того, чтобы математики испытали на ней свои силы».

Итак, спиральная кривая стремится к двум своим предельным точкам:

$$x = \pm \int_0^\infty \sin\frac{u^2}{2a^2}\,\mathrm{d}u,$$
$$y = \pm \int_0^\infty \cos\frac{u^2}{2a^2}\,\mathrm{d}u.$$

На рис. 1.7 изображен график (для произвольно выбранного значения $a = 0.7$),

$$x(t) = \int_0^t \sin\frac{u^2}{2a^2}\,\mathrm{d}u \quad \text{(пунктирная)},$$
$$y(t) = \int_0^t \cos\frac{u^2}{2a^2}\,\mathrm{d}u \quad \text{(сплошная)},$$

для $t > 0$, и очень хотелось бы знать точные значения пределов, к которым эти кривые стремятся.

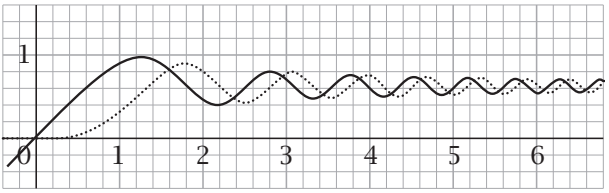


Рис. 1.7. Предельная точка

Мы не знаем о вкладах других математиков, но спустя 38 лет, в 1781 г., Эйлер сам опубликовал ответ на им же поставленный вопрос[1]:

$$x = \pm\frac{a}{\sqrt{2}}\sqrt{\frac{\pi}{2}},$$
$$y = \pm\frac{a}{\sqrt{2}}\sqrt{\frac{\pi}{2}}.$$

Его подход – типичный пример мастерства Эйлера в манипуляции символами: комбинация интегралов, комплексных степеней и гамма-функции (которую он же ввел в рассмотрение в 1729 г.); сам он описывал этот метод как *сингулярный*. Этим сингулярным методом были получены и другие результаты, самым значительным из которых является формула:

$$\int_0^\infty \frac{\sin x}{x}\,\mathrm{d}x = \frac{\pi}{2}.$$

На первоначальный вопрос Бернулли был дан ответ, и кривая, являющаяся решением, была тщательно изучена. Но это лишь крохотная частица, почти затерявшаяся в огромном наследии Эйлера.

1.4. Кривая одна, названий много

Потом эта кривая была надолго забыта. Точнее, до 1814 г., когда французский физик Огюстен Френель вывел выражение для интенсивности освещения в любой точке дифракционной картины, которое (при некоторых упрощающих предположениях) имеет вид:

$$I_v = \left[\int_0^v \cos\tfrac{1}{2}\pi t^2\,\mathrm{d}t\right]^2 + \left[\int_0^v \sin\tfrac{1}{2}\pi t^2\,\mathrm{d}t\right]^2.$$

Так что если и не сама спираль, то ее компоненты снова всплыли, и в 1818 г. в письме Французской академии наук Френель представил таблицу значений обоих интегралов для значений v с шагом 0.1 в диапазоне от $v = 0.1$ до $v = 5.1$, впоследствии расширенном до $v = 5.5$, с четырьмя знаками после запятой. Конечно, потом было много других таблиц, но эти две определяющие компоненты спирали Эйлера почти повсеместно называются *интегралами Френеля*. Сама спираль возродилась в 1874 г., когда французский ученый Мари Альфред Корню, следуя по стопам Френеля, построил кривую и осознал ее полез-

[1] О значениях интегралов при изменении переменной от $x = 0$ до $x = \infty$ (E675).

ность для вычислений, связанных с задачами дифракции. В знак признания заслуг этого выдающегося ученого кривой присвоено его имя – приведенные ниже слова были произнесены на похоронах Корню в 1902 г. его учеником Анри Пуанкаре:

> «Сегодня для предсказания влияния произвольного экрана на луч света все пользуются спиралью Корню».

И вот еще выдержка из одного некролога (Ames 1902):

> «...а метод изучения проблем дифракции с помощью спиралей Корню знаком каждому».

Термин *спираль Корню* в ходу до сих пор. Начиная с конца XIX и даже в XX в. математические свойства этой кривой исследовал итальянский математик Эрнесто Чезаро, ему она напоминала форму, которую принимает нить, намотанная на катушку. Исходя из этого образа, Чезаро в 1886 г. предложил назвать спираль *клотоидой* в часть *Клото*, одной из трех *сестер-мойр*, которая прядет нить человеческой жизни, наматывая ее на веретено. Итальянский романтизм уравновешивается геометрическими открытиями Чезаро, относящимися к этой кривой; их было много, но их природа далека от современной математики. В качестве примера приведем формулировки двух из них, оставляя читателю возможность поискать непонятные термины в интернете:

> «Если клотоида катится по прямой, то геометрическим местом центров кривизны, соответствующих точке контакта, является равнобочная гипербола с асимптотой, совпадающей с рассматриваемой прямой».

> «Клотоида – единственная кривая, обладающая тем свойством, что центр тяжести любой дуги совпадает с центром подобия окружностей, соприкасающихся с дугой в ее крайних точках».

Термин *клотоида* также применяется и поныне.

Тем временем поезда стали двигаться быстрее. Современные игрушечные железные дороги комплектуются фрагментами рельсов разной формы, но поворотные имеют форму дуг окружностей разной длины и радиуса. Однако, как известно любому серьезному любителю моделей железных дорог, такая форма не отражает реальность. Поезд, движущийся с постоянной скоростью v по прямому участку пути, будет испытывать мгновенное изменение ускорения от 0 до центростремительного ускорения $v^2/r = \kappa v^2$, когда направление движения изменяется по дуге окружности радиуса r (и, следовательно, кривизны $\kappa = 1/r$). При этом и поезд, и его пассажиры испытывают неприятное ощущение, которое технически называется *толчок* и которого следует избегать. С первых лет развития железнодорожного транспорта стремление избежать толчков побуждало к проектированию различных *переходных кривых*, заменяющих дугу окружности, но лучшей из всех является та, для которой кривизна, а следовательно, и ускорение линейно возрастает вдоль пути, начиная с 0, – а это и есть спираль Эйлера. Вероятно, спираль была впервые применена таким образом в 1881 г. (Talbot 1890–91):

> «По-видимому, переходная спираль впервые была использована компанией Pan Handle Railroad в 1881 г. Эллиотом Холбруком. Основная часть представленного здесь исследования была выполнена еще до того, как вниманию автора было предложено использование кривой м-ром Холбруком, и есть основания полагать, что большинство формул и методов излагаются здесь впервые».

Наглядная иллюстрация приведена на рис. 1.8. Часть (a) – вид железнодорожных путей сверху: две прямые линии соединены полуокружностями, но та, что справа, заменена двумя участками спирали Эйлера (показана пунктиром). На рисунке (b) показано ускорение поезда, который сначала движется с постоянной скоростью по верхней колее, затем с ускорением по спиралям Эйлера, затем снова по нижней колее и, наконец, завершает движение по полуокружности: толчок заменен линейно возрастающим ускорением, которое по абсолютной величине чуть больше, но все равно намного предпочтительнее.

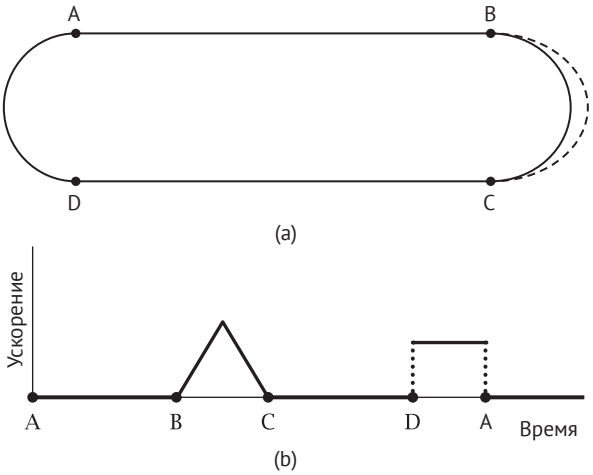


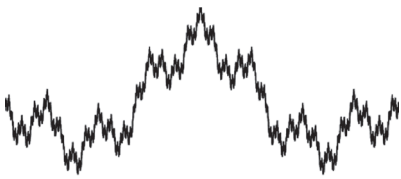
Рис. 1.8. Переходная кривая в действии

Как ни странно, не существует термина *спираль Холбрука*, хотя есть *спираль Гловера*, названная в часть Джеймса Гловера, который в 1900 г. раструбил на весь свет о полезности этой кривой в роли переходной кривой для железнодорожных путей. В этом качестве американцам она может быть хорошо знакома под названием *спирали AREMA*, она применяется и по сей день в любой ситуации, когда необходимо гладкое соединение прямых и кривых, – в частности, на автотрассах и головокружительных спусках на американских горках.

Но вне зависимости от контекста, теоретического или практического применения и названия, эта кривая явилась на свет как точно определенная и хорошо изученная сущность благодаря вездесущему Леонарду Эйлеру и потому, безусловно, является *спиралью Эйлера*.

Глава 2

Кривая Вейерштрасса



Почему именно эта кривая?

Это первая всюду непрерывная, но нигде не дифференцируемая кривая, в этом качестве она в немалой степени послужила стимулом к формулировке точного определения предела. Также это первый признанный пример того, что впоследствии назовут фракталом.

2.1. Наивные мысли

Мы начнем эту главу с обращения к логически разумной интуиции. Если мы считаем кривую чем-то таким, что можно нарисовать, то разумно предположить, что она должна удовлетворять следующим критериям:

- она должна быть непрерывной или, по крайней мере, состоять из непрерывных участков, поскольку линия, проведенная между любыми двумя точками, по необходимости должна проходить через все точки на выбранной траектории;
- в каждой точке у нее должна существовать касательная, потому что в каждой точке траектории должно быть определенное направление движения;
- путь между любыми двумя точками должен иметь конечную длину, потому что он нарисован за конечное время.

Так думаем мы, и так думало большинство математиков в начале XIX в., когда точная природа «кривой» была предметом оживленных споров. Это могла быть только что описанная нами траектория движения, или физическое проявление функции, выраженной аналитической формулой, или бесконечный степенной ряд, не имеющий непосредственного геометрического воплощения, возможно, даже не сходящийся. К тому времени стало ясно, что удобная, хотя и наивная концепция *кривой*, которая так долго служила математикам, нуждается в срочной переоценке. Первое из приведенных выше предположе-

ний было подвергнуто сомнению в 1829 г., когда в конце работы, посвященной рядам Фурье, Лежен Дирихле (Dirichlet 1829) предъявил пример всюду разрывной функции для $c \neq d$:

$$f(x) = \begin{cases} c: \text{ если } x \text{ рационально} \\ d: \text{ если } x \text{ иррационально,} \end{cases}$$

хотя можно ли назвать ее кривой, было сомнительно тогда и остается таковым по сей день. Но не это, а атака на второе предположение породила кривую, являющуюся предметом настоящей главы. Цель дифференцирования – найти градиенты касательных к кривым. Правда, касательная может быть вертикальной, например в крайней левой и крайней правой точке окружности или к кривой $\sqrt{x}$ в начале координат, где наличие касательной физически очевидно, но производная бесконечна. Важнее, впрочем, другое: простота обеих «кривых» на рис. 2.1 маскирует их значимость, когда мы пытаемся придать смысл касательной в точке, где направление резко изменяется.

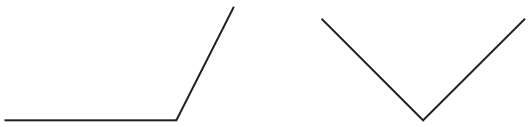


Рис. 2.1. Простая, но недифференцируемая

Аналитически кривую на рис. 2.1(b) можно описать с привлечением функции модуля:

$$|x| = \begin{cases} x: & x \geqslant 0 \\ -x: & x < 0, \end{cases}$$

с помощью которой легко построить бесконечное число странных точек, воспользовавшись периодичностью, например $|\sin x|$, как показано на рис. 2.2. Тогда отклонение от поведения касательных к непрерывным кривым будет многократно усилено, но по крайней мере проблематичные точки – исключения, между которыми кривая ведет себя как положено, – остаются изолированными.

Короче говоря, то, что непрерывность – обязательное условие дифференцируемости, было признано при самом создании дифференциального исчисления, но вопрос о том, является ли это условие достаточным (по большей части), оставался открытым. Однако общее мнение склонялось к тому, что так оно и есть, и это утверждение даже «доказывалось» во многих основных тогдашних учебниках. Важный и очень длинный мемуар Андре-Мари Ампера (с отнюдь не кратким названием, см. Ampère (1806)) содержал рассуждение, которое многие сочли его собственным «доказательством» этого результата и которое было положено в основу других; однако нечеткость языка бросает тень сомнения на истинную природу вывода из его вычислений. Но что бы ни думал сам Ампер по поводу своего доказательства, в математическом сообществе царила уверенность, что (отвлекаясь от сомнительных примеров

типа приведенного Дирихле) как бы ни определяли функцию, если с ней ассоциирована кривая, то функция не может быть настолько своенравной, чтобы эту кривую нельзя было нарисовать.

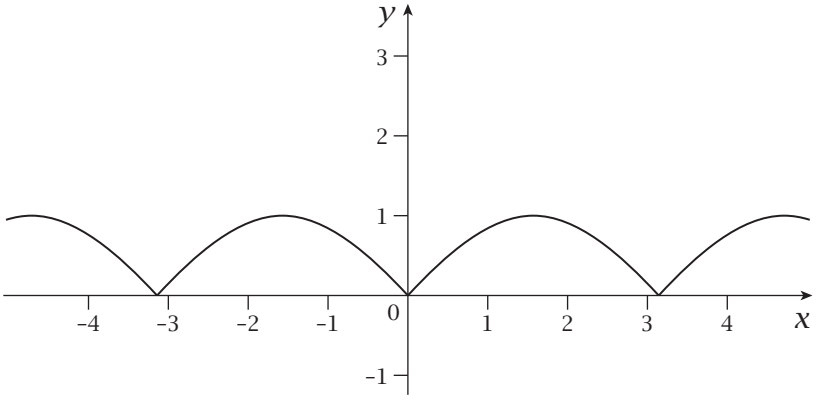


Рис. 2.2. Более сложный случай недифференцируемости

2.2. Глубокие мысли

Убежденность в том, что кривые порождены известными функциями, которые комбинируются разумными способами, нависла над всей территорией, где был рожден раздел математики, который мы сегодня называем вещественным анализом. Чтобы исследовать ее ландшафт должным образом, потребовались не столь комфортные построения:

> «Всякий понимает глубокую разницу, которая существует между фактом, полученным в условиях, специально сконструированных с определенной целью, не имея никакого другого намерения и интереса, кроме как продемонстрировать возможность, – своего рода экспонатом в кунсткамере, и тем же фактом, встретившимся в процессе разработки теории, уходящей корнями в наиболее востребованные и насущные проблемы анализа».

Эти слова Жака Адамара (1921, стр. 212) относятся к тому, что к 1921 г. превратилось в стаю монстров из математической бездны, построенных специально для того, чтобы поставить под сомнение идеи, которые долго считались само собой разумеющимися: функция Дирихле и ее вариант, непрерывный во всех иррациональных точках, но разрывный во всех рациональных; монотонные кривые, производная которых почти всюду равна нулю; ограниченные, но не интегрируемые функции и т. д. Но самой шокирующей была кривая, которая всюду непрерывна, но нигде не дифференцируема. Кривая, которую невозможно нарисовать, потому что ни в одной точке не определено направление пера. Для краткости будем называть такие конструкции CND-кривыми.

Хотя первая известная попытка построить CND-кривую должна быть отнесена на счет богемского ученого-энциклопедиста Бернарда Больцано, его рукопись «Functionenlehre», написанная около 1830 г., была опубликована только в 1930 г. Кроме того, его доказательство непрерывности содержало дефекты, а недифференцируемость была установлена лишь для плотного подмножества

вещественной оси. Тем не менее впоследствии было доказано, что бесконечная последовательность отрезков прямых, описанная в его построении, действительно является CND-кривой. Похожая судьба постигла работу швейцарского математика Шарля Селлерье, который приблизительно в 1860 г. предложил в качестве CND-кривой функцию

$$\sum_{n=1}^{\infty} \frac{1}{a^n} \cos(a^n x),$$

где $a > 1000$ – четное число. И эта рукопись осталась неопубликованной, а ее появление в 1890 г. вряд ли сильно поспособствовало признанию ее важности. Тем не менее сама форма этой функции в иных руках привела к открытию того, что единодушно и уверенно признавалось невозможным: всюду непрерывной, но нигде не дифференцируемой кривой.

18 июля 1872 г. Карл Вейерштрасс решился вообразить невообразимое и с успехом явил миру новую истину. Именно тогда он представил Королевской академии наук в Берлине статью – вторую из когда-либо поданных им на рассмотрение академии статей, – в которой описал свой пример CND-кривой и строго доказал это. Вот что он писал:

> «Еще совсем недавно все полагали, что производная непрерывной функции вещественной переменной может быть не определена или бесконечна лишь в нескольких изолированных точках. Даже в работах Гаусса, Коши и Дирихле – математиков, склонных критически относиться ко всему в своих областях, – мы не находим (насколько мне известно) следов иного мнения. От учеников Римана я узнал, что он первым выразил убежденность (в 1861 г. или даже раньше) в неверности этого утверждения, например для функции, описываемой бесконечным рядом:
>
> $$\sum_{n=1}^{\infty} \frac{1}{n^2} \sin(n^2 x).$$
>
> К сожалению, доказательство Риманом этого факта не было опубликовано и не сохранилось ни в его бумагах, ни в устном изложении. Еще более достойно сожаления, что я не знаю точно, что говорил Риман по поводу этого примера. Те математики (по крайней мере большинство из них), которые занялись этой проблемой после того, как гипотеза Римана стала известна в широких кругах, кажется, придерживаются мнения, что достаточно доказать существование функций таких, что в любом сколь угодно малом интервале изменения их аргумента существуют точки, где они не дифференцируемы. Существование таких функций доказать очень просто, поэтому я полагаю, что Риман имел в виду только те функции, которые не имеют производной ни в какой точке своей области определения».

Похоже, это первое упоминание функции Римана, и из слов Вейерштрасса мы можем заключить, что ему не удалось доказать этот факт: неудивительно, поскольку впоследствии было доказано, что она всюду непрерывна, но лишь «почти» всюду не дифференцируема (у нее имеется производная, равная $-½$, в точках вида $x = (p/q)\pi$, где p, q – нечетные числа)[1]. Функция, которую Вей-

[1] Этот вопрос окончательно решен в следующих работах: Gerver (1970, 1971) и Hardy (1916, стр. 322–325).

ерштрасс представил на той лекции и которая является предметом настоящей главы, называется *функцией Вейерштрасса*:

$$W(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x),$$

где $0 < a < 1$, $ab > 1 + 3/2\pi$ и $b > 1$ – нечетное число; она непрерывна всюду на вещественной оси, но нигде не дифференцируема.

В этой странной функции и еще более странных условиях в ее определении слышатся отголоски работы Селлерье. Она стала известна широкому математическому сообществу благодаря усилиям математика Поля Дюбуа-Реймона (Paul du Bois-Reymond, 1875) и повторно опубликована в престижном *Crelle's Journal*[1]. Судить о том, как было воспринято существование подобной функции, можно по примечанию самого Дюбуа-Реймона на стр. 29 статьи:

> «Мне кажется, что метафизика функции Вейерштрасса все еще таит в себе много загадок, и я не могу избавиться от мысли, что более глубокое проникновение в предмет в конечном итоге приведет к пределам нашего интеллекта».

На рисунке в начале этой главы показана сумма нескольких первых членов ряда, определяющего кривую Вейерштрасса с параметрами $a = 0.5$ и $b = 5$[2]. Очень грубо идея построения состоит в том, что амплитуды a^n тригонометрических членов бесконечного ряда быстро уменьшаются (поэтому кривая непрерывна), но амплитуды производных членов $(ab)^n$ быстро возрастают, из-за чего ряд, состоящий из производных, расходится. Примером может служить ряд $\sum_{n=0}^{\infty}(\frac{1}{2})^n \cos(4^n x)$, для которого уменьшающиеся амплитуды имеют вид $(½)^n$, а амплитуды членов ряда, составленного из производных, $\sum_{n=0}^{\infty} -2^n \sin(4^n x)$, равны 2^n.

Но мы должны понять дух построения: оно задумывалось как строгое, и именно строгость нам предстоит оценить.

2.3. Дифференцируемость

Поскольку тема этой главы – понятия непрерывности и дифференцируемости, а история предостерегает против нечеткости формулировок, мы приведем стандартное определение в варианте, свободном от формализма ε-δ.

Функция $f(x)$ называется *непрерывной* в точке x_0 своей области определения, если

$$\lim_{x \to x_0} f(x) = f(x_0).$$

Она называется *дифференцируемой* в точке x_0, если предел

---

[1] Настоящее название журнала «Für die reine und angewandte Mathematik». В то время он был известен как *журнал Борхардта*, по имени редактора Карла Борхардта. Это старейшее математическое периодическое издание, выходящее до сих пор.

[2] Заметим, что оригинальные условия выполнены не полностью, но об этом ниже.

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

существует, и в таком случае этот предел называется *производной* $f(x)$ в точке $x = x_0$; производная определена как угловой коэффициент касательной к кривой в точке. В обоих случаях требуется, чтобы предел не зависел от направления приближения к x_0.

Например, стандартное упражнение по элементарному анализу – показать, что определенная выше функция

$$f(x) = |x| = \begin{cases} x: & x \geqslant 0 \\ -x: & x < 0 \end{cases},$$

которая, очевидно, всюду непрерывна, не является дифференцируемой в начале координат. Доказательство следует из того, что

$$\lim_{x \to 0^+} \frac{|x| - |0|}{x - 0} = \lim_{x \to 0^+} \frac{x}{x} = 1,$$

тогда как

$$\lim_{x \to 0^-} \frac{|x| - |0|}{x - 0} = \lim_{x \to 0^-} \frac{-x}{x} = -1;$$

т. е. пределы существуют, но различны.

Существует, однако, формулировка непрерывности и дифференцируемости в терминах сходящихся последовательностей, и в случае дифференцируемости у нас скоро появятся причины ими воспользоваться. Поэтому мы включили следующее определение.

Функция $f(x)$ называется *дифференцируемой* в точке x_0, если для любой последовательности $\{x_n\} \to x_0$ последовательность

$$\left\{ \frac{f(x_n) - f(x_0)}{x_n - x_0} \right\}$$

сходится, и в этом случае предел называется *производной* $f(x)$ в точке $x = x_0$. Заметим, что и в этом случае предел должен существовать независимо от того, с какой стороны последовательность приближается к x_0, и оба предела должны быть равны. Чтобы доказать недифференцируемость, мы должны лишь предъявить последовательность или последовательности, для которых с пределом возникают проблемы.

Снова рассмотрим функцию $f(x) = |x|$ и две подпоследовательности последовательности:

$$x_n = \frac{(-1)^n}{n} \xrightarrow[n \to \infty]{} 0;$$

в одной все n четные, так что $y_n = 1/n$, а в другой все n нечетные, так что $z_n = -1/n$. Тогда

$$\frac{f(y_n) - f(0)}{y_n - 0} = \frac{1/n}{1/n} = 1 \quad \text{и} \quad \frac{f(z_n) - f(0)}{z_n - 0} = \frac{1/n}{-1/n} = -1.$$

Оба предела существуют, но не равны. Далее мы встретимся с более суровыми условиями применения этого определения, когда будем разбирать предложенное Вейерштрассом доказательство недифференцируемости его пресловутой функции.

2.4. Доказательство Вейерштрасса

Формулировка

Функция Вейерштрасса

$$W(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x),$$

где $0 < a < 1$, $ab > 1 + 3/2\pi$ и $b > 1$ — нечетное число, непрерывна на всей вещественной оси, но нигде не дифференцируема.

Доказать требуется две вещи: что функция непрерывна в каждой точке и что она нигде не дифференцируема, и подходы к доказательству этих двух фактов совершенно различны.

Непрерывность

Это следствие совокупности трех результатов из вещественного анализа.

- Предел равномерно сходящейся последовательности или ряда непрерывных функций сам является непрерывной функцией.
- Ряд функций $\sum_{n=0}^{\infty} f_n(x)$ равномерно сходится, если существуют такие постоянные M_n, что $|f_n(x)| < M_n$ для любого n и для всех x, и ряд $\sum_{n=0}^{\infty} M_n$ сходится.
- $|f_n(x)| = |a^n \cos(b^n \pi x)| \leqslant a^n = M_n$, и сумма сходящегося геометрического ряда $\sum_{n=0}^{\infty} a^n = 1/(1-a)$.

Очень правильно, что второй результат известен под названием *признака Вейерштрасса*[1].

Недифференцируемость

Мы возьмем за основу и обобщим доказательство, предложенное Полем Дюбуа-Реймоном в 1875 г., которое разобьем на части. По ходу дела станет ясно, откуда взялись странные ограничения на параметры. Хотя на первый взгляд так не кажется, доказательство элементарно и требует только школьных знаний о суммировании, бесконечной геометрической прогрессии и тригонометрии, а также о неравенстве треугольника – деталей много, но они почти не представляют затруднений. Тем не менее при первом чтении доказательство может показаться устрашающим, однако после повторного внимательного изучения это чувство исчезает. Поучительно наблюдать, что опытный математик может сотворить со столь ограниченными средствами.

[1] В англоязычной литературе принято название Weierstrass M-test. – *Прим. перев.*

Идея

Строятся две последовательности, каждая из которых сходится к произвольной фиксированной точке x_0, – одна сверху, другая снизу, – но при этом пределы отношения разностей различны; с одной стороны, предел равен плюс бесконечности, а с другой стороны, минус бесконечности.

Последовательности

Для фиксированного целого положительного числа b и произвольного целого положительного числа m выберем $\alpha_m \in \mathbb{Z}$ такое, что $-½ < b^m x_0 - \alpha_m \leqslant ½$. Это означает, что

$$\left| \frac{\alpha_m}{b^m} - x_0 \right| < \frac{1}{2b^m} \quad \text{и, следовательно,} \quad \lim_{m\to\infty} \frac{\alpha_m}{b^m} = x_0.$$

Теперь определим промежуточную последовательность $x_{m+1} = b^m x_0 - \alpha_m$ (тогда $-½ < x_{m+1} \leqslant ½$), а также две интересующие нас последовательности:

$$y_m = \frac{\alpha_m - 1}{b^m} \quad \text{и} \quad z_m = \frac{\alpha_m + 1}{b^m}.$$

Тогда

$$y_m - x_0 = \frac{\alpha_m - 1}{b^m} - x_0 = \frac{\alpha_m - 1 - b^m x_0}{b^m} = -\frac{1 + x_{m+1}}{b^m} < 0,$$

откуда следует, что $y_m < x_0$. И также

$$z_m - x_0 = \frac{\alpha_m + 1}{b^m} - x_0 = \frac{\alpha_m + 1 - b^m x_0}{b^m} = \frac{1 - x_{m+1}}{b^m} > 0,$$

откуда следует, что $z_m > x_0$. Так как $\lim_{m\to\infty} y_m = x_0$ слева и $\lim_{m\to\infty} z_m = x_0$ справа, то мы имеем требуемые сходящиеся последовательности.

Инструментарий

По ходу доказательства нам понадобится несколько стандартных элементарных результатов, которые мы перечислим в порядке использования.

1. $\cos x - \cos y = -2 \sin \frac{1}{2}(x + y) \sin \frac{1}{2}(x - y)$.
2. $|\sum_{r=1}^{n} a_r| \leqslant \sum_{r=1}^{n} |a_r|$ (неравенство треугольника).
3. $|(\sin x)/x| \leqslant 1$.
4. $\sum_{n=0}^{m-1} r^n = (r^m - 1)/(r - 1)$.
5. $\cos(\theta + \varphi) = \cos\theta \cos\varphi - \sin\theta \sin\varphi$.

Предел слева

Отношение разностей разбивается на две части, которые рассматриваются по отдельности:

$$\begin{aligned}\frac{W(y_m)-W(x_0)}{y_m-x_0} &= \sum_{n=0}^{\infty} a^n \frac{\cos(b^n\pi y_m)-\cos(b^n\pi x_0)}{y_m-x_0}\\ &= \sum_{n=0}^{m-1} (ab)^n \frac{\cos(b^n\pi y_m)-\cos(b^n\pi x_0)}{b^n(y_m-x_0)}\\ &\quad + \sum_{n=0}^{\infty} a^{m+n} \frac{\cos(b^{m+n}\pi y_m)-\cos(b^{m+n}\pi x_0)}{y_m-x_0}\\ &= S_1 + S_2.\end{aligned}$$

Используя тождество (1) выше, имеем

$$\begin{aligned}S_1 &= \sum_{n=0}^{m-1} (ab)^n \frac{\cos(b^n\pi y_m)-\cos(b^n\pi x_0)}{b^n(y_m-x_0)}\\ &= \sum_{n=0}^{m-1} (ab)^n \frac{-2\sin(\frac{1}{2}(b^n\pi y_m + b^n\pi x_0))\sin(\frac{1}{2}(b^n\pi y_m - b^n\pi x_0))}{b^n(y_m-x_0)}\\ &= \sum_{n=0}^{m-1} (-\pi)(ab)^n \sin\left(\frac{b^n\pi(y_m+x_0)}{2}\right)\frac{\sin(\frac{1}{2}b^n\pi(y_m-x_0))}{\frac{1}{2}b^n\pi(y_m-x_0)}.\end{aligned}$$

Используя (2), затем (3) и затем (4), получаем

$$\begin{aligned}|S_1| &= \left|\sum_{n=0}^{m-1} (-\pi)(ab)^n \sin\left(\frac{b^n\pi(y_m+x_0)}{2}\right)\frac{\sin(\frac{1}{2}b^n\pi(y_m-x_0))}{\frac{1}{2}b^n\pi(y_m-x_0)}\right|\\ &\leqslant \pi \sum_{n=0}^{m-1} (ab)^n \left|\sin\left(\frac{b^n\pi(y_m+x_0)}{2}\right)\right|\left|\frac{\sin(\frac{1}{2}b^n\pi(y_m-x_0))}{\frac{1}{2}b^n\pi(y_m-x_0)}\right|\\ &\leqslant \pi \sum_{n=0}^{m-1} (ab)^n = \pi\frac{(ab)^m-1}{ab-1} < \frac{\pi(ab)^m}{ab-1},\end{aligned}$$

откуда следует, что можно записать

$$S_1 = (ab)^m \frac{\pi}{ab-1}\varepsilon_1,$$

где $\varepsilon_1 \in (-1, 1)$.

С S_2 мы поступаем по-другому, потому что b^m сокращается.

Сначала рассмотрим обе компоненты числителя:

$$\begin{aligned}\cos(b^{m+n}\pi y_m) &= \cos\left(b^{m+n}\pi\frac{\alpha_m-1}{b^m}\right)\\ &= \cos(b^n(\alpha_m-1)\pi) = (-1)^{b^n(\alpha_m-1)}\\ &= [(-1)^{b^n}]^{\alpha_m-1} = (-1)^{\alpha_m-1} = -(-1)^{\alpha_m},\end{aligned}$$

где мы воспользовались тем, что

$$\cos N\pi = \begin{cases} 1: & N \text{ четное} \\ -1: & N \text{ нечетное}, \end{cases}$$
$$= (-1)^N$$

а затем – тем, что b нечетное.

Точно так же, пользуясь тем, что

$$x_{m+1} = b^m x_0 - \alpha_m \quad \Rightarrow \quad x_0 = \frac{\alpha_m + x_{m+1}}{b^m}$$

и тождеством (5) выше, получаем

$$\begin{aligned} &\cos(b^{m+n}\pi x_0) \\ &\quad = \cos\left(b^{m+n}\pi\frac{\alpha_m + x_{m+1}}{b^m}\right) = \cos(b^n\pi\alpha_m + b^n\pi x_{m+1}) \\ &\quad = \cos(b^n\pi\alpha_m)\cos(b^n\pi x_{m+1}) - \sin(b^n\pi\alpha_m)\sin(b^n\pi x_{m+1}) \\ &\quad = [(-1)^{b^n}]^{\alpha_m}\cos(b^n\pi x_{m+1}) - 0 = (-1)^{\alpha_m}\cos(b^n\pi x_{m+1}), \end{aligned}$$

так как $\alpha_m \in \mathbb{Z}$.

Отсюда и из определения y_m имеем

$$\begin{aligned} S_2 &= \sum_{n=0}^{\infty} a^{m+n}\frac{-(-1)^{\alpha_m} - (-1)^{\alpha_m}\cos(b^n\pi x_{m+1})}{-(1+x_{m+1})/b^m} \\ &= (ab)^m(-1)^{\alpha_m}\sum_{n=0}^{\infty} a^n\frac{1+\cos(b^n\pi x_{m+1})}{1+x_{m+1}}. \end{aligned}$$

Но

$$\begin{aligned} \sum_{n=0}^{\infty} a^n\frac{1+\cos(b^n\pi x_{m+1})}{1+x_{m+1}} &\geqslant \frac{1+\cos(\pi x_{m+1})}{1+x_{m+1}} \\ &\geqslant \frac{1+0}{1+\frac{1}{2}} = \frac{2}{3}. \end{aligned}$$

Это неравенство просто отражает тот факт, что бесконечная сумма положительных членов ($0 < a < 1$ и $-\frac{1}{2} < x_{m+1} \leqslant \frac{1}{2}$, так что $-\frac{1}{2}\pi < \pi x_{m+1} \leqslant \frac{1}{2}\pi$) не может быть меньше первого члена ряда. Итак,

$$S_2 = (ab)^m(-1)^{\alpha_m} \times \tfrac{2}{3} \times \eta_1 \quad \text{где } \eta_1 > 1.$$

Объединяя оба результата, можно написать:

$$\begin{aligned} \frac{W(y_m) - W(x_0)}{y_m - x_0} &= (ab)^m\frac{\pi}{ab-1}\varepsilon_1 + (ab)^m(-1)^{\alpha_m} \times \frac{2}{3} \times \eta_1 \\ &= (-1)^{\alpha_m}(ab)^m\eta_1\left(\frac{\pi}{ab-1}\frac{\varepsilon_1}{\eta_1} + \frac{2}{3}\right). \end{aligned}$$

Предел справа

Рассуждение точно такое же и дает почти такой же результат, но присутствие начального знака минус приводит к принципиально другому выражению:

$$\frac{W(z_m) - W(x_0)}{z_m - x_0} = -(-1)^{\alpha_m}(ab)^m \eta_2 \left(\frac{\pi}{ab-1} \frac{\varepsilon_2}{\eta_2} + \frac{2}{3} \right),$$

где $\varepsilon_2 \in (-1, 1)$ и $\eta_2 > 1$.

Развязка

Прежде всего заметим, что $|\varepsilon_1/\eta_1| < 1$ (и то же самое верно для $|\varepsilon_2/\eta_2|$). Условие $ab > 1 + 1\frac{1}{2}\pi$ эквивалентно $\pi/(ab - 1) < \frac{2}{3}$, поэтому левое и правое отношение разностей гарантированно разного знака; они не стремятся к нулю, и, поскольку $\lim_{m\to\infty}(ab)^m = \infty$, у функции нет производной в произвольной точке x_0.

Таким образом, мы получили искомый результат: кривую, которую можно нарисовать теоретически, но невозможно практически.

2.5. Отголоски

После открытия кривой Вейерштрасса искомая цель была достигнута, но путешествие еще далеко не завершилось. Выбирая параметры, Вейерштрасс стремился к функциональности, а не к оптимальности, и несравненный математический эстет Г. Х. Харди пожелал навести порядок. Вот что он писал (Hardy 1916):

> «Очевидно, что эти условия искусственны. Трудно поверить, что хотя бы одно из них действительно отражает какую-то существенную особенность обсуждаемой проблемы. Они возникли просто из-за ограниченности использованных методов. Есть только одно условие, которое кажется естественным и, по видимости, имеющим отношение к делу: $ab \geqslant 1$».

Он не упомянул необходимость условия $0 < a < 1$, – которое обеспечивает сходимость коэффициентов тригонометрического ряда, гарантирующую непрерывность, – чтобы подчеркнуть требование к расходимости производного ряда. Применяя гораздо более сложные методы, чем Вейерштрасс, он доказал следующую теорему.

Теорема 1.31. *Ни одна из функций* $C(x) = \sum a^n \cos(b^n \pi x)$, $S(x) = \sum a^n \sin(b^n \pi x)$, *где* $0 < a < 1$, $b > 1$ *и* $ab \geqslant 1$, *не имеет конечной производной ни в одной точке.*

Читатель, возможно, заметил, что параметры, выбранные нами для кривой, изображенной в начале главы, не удовлетворяют оригинальным условиям Вейерштрасса, но удовлетворяют условиям Харди.

Тем временем другие люди активно пополняли кунсткамеру Адамара. Варианты тригонометрических рядов, варианты диапазонов параметров и варианты конечных выводов следовали один за другим на протяжении многих лет. Например, было показано, что:

- функция Дарбу (1873, опубликовано в 1875)

$$\sum_{n=1}^{\infty} \frac{1}{n!} \sin((n+1)!x),$$

- класс функций Дини (1877–1878), примером которого служит функция

$$\sum_{n=1}^{\infty} \frac{a^n}{1.3.5 \cdot \cdots \cdot (2n-1)} \cos(1.3.5 \cdot \cdots \cdot (2n-1)\pi x), \text{ где } |a| > 1 + 1\tfrac{1}{2}\pi,$$

- функция Герца (1879), $\sum_{n=1}^{\infty} a^n \cos^p(b^n \pi x)$, где $a > 1$, $p \in \mathbb{N}$ – нечетное число и $ab > 1 + \frac{2}{3}\pi$,

являются CND-функциями. Такие примеры вызвали глубокое замешательство и негодование. За 12 лет Шарль Эрмит и Томас Стилтьес обменялись 432 письмами, и в одном из них, датированном 20 мая 1893 г., Эрмит (Hermite 1905, стр. 318) писал:

> «Но на этих элегантных построениях лежит печать проклятия; их производные <...> являются бессмысленными рядами. Одной рукой анализ отнимает то, что дает другой. Со страхом и ужасом взираю я на эту прискорбную пагубу непрерывных функций, не имеющих производной».

В 1899 г. человек, которого принято считать последним математиком-энциклопедистом[1], Пуанкаре (Poincaré 1899), писал:

> «В течение половины столетия на свет появилось полчище странных функций, построенных так, чтобы как можно меньше походить на честные функции, служащие какой-то цели. Исчезла непрерывность, или непрерывность осталась, но без производных и т. д. и т. п. <...> В былые времена, человек изобретал новую функцию для практической цели; ныне же их придумывают, только чтобы продемонстрировать изъяны в рассуждениях отцов, и это все, что можно из них извлечь».

Шло время, но шлюзы упрямо оставались открытыми:

- функция Такаги (1903):

$$\sum_{n=1}^{\infty} \frac{1}{2^n} \inf_{m \in \mathbb{Z}} (2^n x - m);$$

- функция ван дер Вардена (1930):

$$\sum_{n=1}^{\infty} \frac{1}{10^n} \inf_{m \in \mathbb{Z}} (10^n x - m);$$

- функции Фабера (1907, 1908):

$$\sum_{n=1}^{\infty} \frac{1}{10^n} \inf_{m \in \mathbb{Z}} (2^{n!} x - m) \quad \text{и} \quad \sum_{n=1}^{\infty} \frac{1}{n!} \inf_{m \in \mathbb{Z}} (2^{n!} x - m);$$

- функции Кноппа (1918):

$$\sum_{n=0}^{\infty} a^n \inf_{m \in \mathbb{Z}} (b^n x - m),$$

где $a \in (0, 1)$, $ab > 4$ и $b > 1$ – четное число;

[1] В том смысле, что он оставил след во всех известных в то время разделах математики.

- функция Маккарти (1953):

$$\sum_{n=1}^{\infty} \frac{1}{2^n} g(2^{2^n} x),$$

где

$$g(x) = \begin{cases} 1 + x: & x \in [-2, 0] \\ 1 - x: & x \in [0, 2] \end{cases} \text{ и } g(_x + 4) = g(x), x \in \mathbb{R};$$

- функция Вена (2002):

$$\prod_{n=1}^{\infty} (1 + a_n \sin(b_n \pi x)),$$

где сумма и произведение

$$\sum_{n=1}^{\infty} a_n < \infty \quad \text{и} \quad b_n = \prod_{k=1}^{n} p_k$$

берутся по четным числам, для которых $\lim_{n\to\infty} 2^n / a_n p_n = 0$.

Все это CND-функции. Этот перечень вряд ли исчерпывающий, а если мы добавим также кривые, для которых аномалия существует только в единичном интервале, или те, которые определены сугубо геометрически, то он еще увеличится.

2.6. Заключительные мысли

Сначала мы отметили, что третье из *наивных предположений*, сформулированных в начале этой главы, также поставлено под сомнение кривой Вейерштрасса. Принимая во внимание стандартную формулу длины дуги

$$\int_a^b \sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2}\, \mathrm{d}x,$$

вряд ли стоит удивляться тому, что для функции, не имеющей производной, скорее всего, нельзя корректно определить длину дуги, но этот факт лучше рассматривать как следствие более общей теории, в тонкие детали которой мы вдаваться не будем. Для наших целей достаточно следующего.

У термина *ограниченная вариация* есть точное определение, но интуитивно вариация кривой ограничена, если она колеблется не слишком часто и амплитуды не слишком велики.

- Кривая называется *спрямляемой*, если для нее имеет смысл понятие длины.
- Кривая спрямляема тогда и только тогда, когда она имеет ограниченную вариацию.
- Если вариация функции ограничена, то она дифференцируема почти всюду.
- Контрапозиция последнего утверждения означает, что ни на каком интервале для кривой невозможно корректно определить длину дуги.

И такая кривая является фракталом. В 1977 г. сам Бенуа Мандельброт отметил, что ее топологическая (хаусдорфова) размерность H больше 1, т. е. размерности стандартной кривой на плоскости: это условие отвечает определению фрактальной кривой. Впрочем, это тонкая материя, потому что существует еще один стандартный способ измерения размерности, который обычно называется *грубой размерностью*, и он тоже характеризует кривую Вейерштрасса как фрактал. Ее размерность, подсчитанная таким способом, равна $D = 2 + \log a / \log b$, и значит, необходимые ограничения на параметры кривой – $0 < a < 1$ и $ab > 1$:

$$\log ab > 0 \Rightarrow \log a > -\log b \Rightarrow \frac{\log a}{\log b} > -1 \text{ а также } \frac{\log a}{\log b} < 0,$$

$$-1 < \frac{\log a}{\log b} < 0 \Rightarrow 1 < 2 + \frac{\log a}{\log b} < 2 \Rightarrow 1 < D < 2.$$

В нашем примере $a = 0.5$, $b = 5$, так что имеем

$$D = 2 + \frac{\log 0.5}{\log 5} = 1.569\ldots.$$

Но эта кривая таит еще одну загадку. Известно, что для фракталов $H \leqslant D$, причем равенство достигается очень часто, но не всегда, и многие убеждены, что для кривой Вейерштрасса $H = D$; беда в том, что никому не удалось это доказать – во всяком случае, не полностью.

Кривая Вейерштрасса продолжает жить главным образом во вводных курсах вещественного анализа, напоминая студентам, что переход от школьной математики к университетской сопровождается новыми требованиями к строгости: интуиции, которая, спору нет, играет важную роль в математических размышлениях, было сурово указано на ее место математиками девятнадцатого столетия. Таким образом, эта кривая является предостережением и фракталом, но ее современная роль этим не ограничивается (Kolwankar and Gangal 1996):

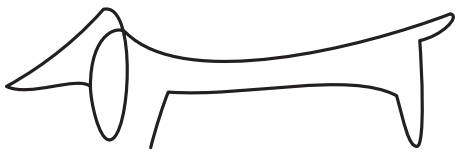
> «Функции Вейерштрасса – не просто математические курьезы, они естественно возникают в некоторых местах. Например, известно, что график этой функции является репеллером или аттрактором некоторых динамических систем. В функции такого вида также можно узнать характеристическую функцию полета Леви на одномерной решетке, а это значит, что полет Леви можно рассматривать как суперпозицию функций типа вейерштрассовой. Эта функция также использовалась для генерирования дробного броуновского сигнала путем умножения каждого члена на случайную амплитуду и рандомизации фаз каждого члена».

Что бы ни означали эти технические термины, ясно, что кривая Вейерштрасса – больше, чем фантастический экспонат из кунсткамеры Адамара.

И последнее: если привлечь к рассмотрению теорию дробных производных, то окажется, что она *всего лишь* не дифференцируема, потому что имеет производные всех порядков, меньших 1.

Глава 3

Кривые Безье



Почему именно эти кривые?

В силу их вездесущности и простоты в использовании. Они входят в любой пакет компьютерной графики, даже если так не называются, потому что любую разумную кривую можно достаточно точно, легко и быстро аппроксимировать ими. Их можно комбинировать для моделирования профилей промышленных изделий, будь то автомобили, самолеты, ботинки…, а их двумерные варианты играют роль самих поверхностей. Они присутствуют в любой книге, включая эту, поскольку с их помощью созданы все шрифты PostScript и Truetype – как, впрочем, и изображенная выше такса по кличке Лump.

3.1. Кривая кривых Безье

Самым знаменитым из никогда не живших математиков является Никола Бурбаки. Под этим псевдонимом выступала группа из девяти молодых (преимущественно) французских математиков в середине 1930-х гг. Но Бурбаки хотя бы существовал как личность, пусть и не математик: он был генералом, участвовавшим во франко-прусской войне. А вот профессор Онесим Дюран был блестящим математиком, который существовал только в воображении своего создателя, французского инженера и начальника отдела дизайна в автомобильной компании Рено: Пьера Безье. Разрабатывать дизайн современных автомобилей (и многого другого) помогают изощренные системы автоматизированного проектирования (CAD), а при изготовлении автомобилей используются системы автоматизированного производства (CAM). Те и другие нуждаются в мощных компьютерах, которых не было в 1960-е гг., когда Безье делал свою привычную работу. Тогдашнюю обстановку хорошо характеризует следующая цитата[1]:

[1] Alastair Townsend «On the spline: a brief history of the computational curve» (*http://www.alatown.com/spline-history-architecture/*).

> «Помимо чертежей в натуральную величину, компания “Ситроен” и другие производители автомобилей в процессе проектирования опирались в значительной степени на физические модели; они служили для концептуальных разработок, а также для сохранения, переноса и совместного использования геометрии автомобиля различными группами. Отдел художественного конструирования начинал с изготовления концептуальных моделей нового прототипа в уменьшенном масштабе. Криволинейные поверхности приходилось вручную увеличивать до натуральной величины, измеряя расстояния между точками на небольших глиняных макетах, а затем перенося их на чертежную доску в масштабе 1:1. Нанесенные на доску точки интерполировались с помощью лекал для получения наилучшего приближения. Из фанеры вырезались поперечные сечения, которые собирались вместе для изготовления опалубки. Каркас заполнялся глиной, после чего из него вылепляли детальную эталонную модель автомобиля. Сформованную эталонную модель дорабатывали до совершенства опытные лепщики, и наконец фиксировалась окончательная форма, готовая к производству.
>
> Дизайнеры, инженеры и механики могли обращаться к геометрии эталонной модели (или одной из нескольких ее копий) при проектировании деталей и оснастки, необходимой для их изготовления. Для извлечения нужной геометрии и обеспечения ее точной стыковки с соседними деталями требовалась столь же изнурительная методика переноса. Работа над каждой деталью состояла из нескольких итераций и сильно зависела от квалификации, изобретательности и субъективного суждения рабочих. Но интерпретация геометрии, которая часто производилась лишь качественно, оставляла много места несогласованности и человеческим ошибкам. Компании был необходим единый геометрический язык, чтобы можно было сохранять описание каждой детали в числовой форме, а не полагаться на трудоемкий и чреватый ошибками процесс ручного копирования».

По инициативе французского правительства промышленные предприятия начали внедрять первые компьютеры, а Безье применил свои аналитические способности, чтобы поставить новые (пусть и ограниченные) вычислительные мощности на службу путем проектирования кривых и поверхностей, которые в сочетании придавали элегантную и аэродинамически эффективную форму автомобилю. Его идея заключалась в том, что форму кривой можно охарактеризовать динамической ломаной, показанной пунктиром на рис. 3.1, и что, манипулируя этой описанной ломаной, можно манипулировать и самой кривой.

Именно здесь профессор Дюран сыграл важную роль, как впоследствии объяснял сам Безье (Bézier 1990):

> «В течение многих лет на презентациях своей работы в компании “Рено” и других местах я ссылался на исследования своего мифического учителя, которого называл Дюраном. Я приписал ему результаты своих собственных размышлений, поскольку это вселяло в слушателей уверенность. Ведь если бы я сказал, что изобрел полиномы сам, то, думаю, был бы воспринят как сквернавец в доме! Вот я и рассказывал о функциях Дюрана, а люди, глядя на кривые, были очень довольны. Я даже читал курс по этим функциям в Национальной консерватории искусств и ремесел. А три года назад мне задали вопрос о Дюране в компании “Дженерал Моторс”».

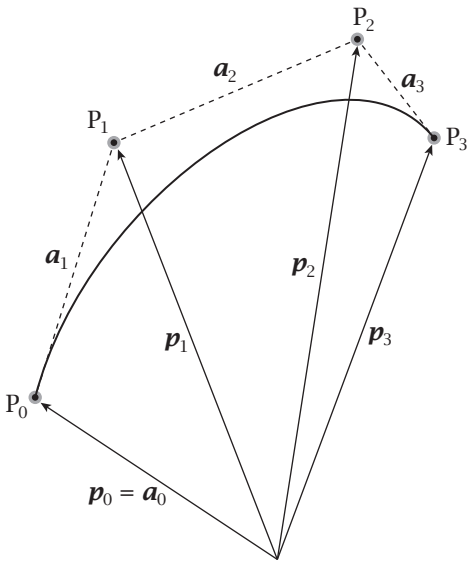


Рис. 3.1. Инициатива Безье

Мы объясним подход Безье, рассмотрев ломаную линию с четырьмя вершинами (и, стало быть, состоящую из трех отрезков). На рис. 3.1 показаны вершины $\{P_0, P_1, P_2, P_3\}$, радиус-векторы $\{\boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3\}$ и (предполагаемые независимыми) векторы, направленные вдоль отрезков:

$$\{\boldsymbol{a}_1 = \boldsymbol{p}_1 - \boldsymbol{p}_0, \boldsymbol{a}_2 = \boldsymbol{p}_2 - \boldsymbol{p}_1, \boldsymbol{a}_3 = \boldsymbol{p}_3 - \boldsymbol{p}_2\}.$$

В терминах отрезков векторное параметрическое уравнение кривой будет иметь вид

$$\boldsymbol{r}(t) = \boldsymbol{a}_0 + f_{3,1}(t)\boldsymbol{a}_1 + f_{3,2}(t)\boldsymbol{a}_2 + f_{3,3}(t)\boldsymbol{a}_3,$$

где многократно встречающееся число 3 служит напоминанием о существовании трех отрезков; диапазон изменения параметра предполагается равным $0 \leqslant t \leqslant 1$.

Функциональные коэффициенты определяются условиями, вытекающими из природы кривой. Первое из них – наличие концевых точек P_0 и P_3, т. е. $\boldsymbol{r}(0) = \boldsymbol{a}_0 = \boldsymbol{p}_0$ и $\boldsymbol{r}(1) = \boldsymbol{p}_3$, откуда следует, что

$$f_{3,1}(0) = f_{3,2}(0) = f_{3,3}(0) = 0 \text{ и } f_{3,1}(1) = f_{3,2}(1) = f_{3,3}(1) = 1,$$

где последнее условие – следствие использования сложения векторов.

Следующее условие состоит в том, что первый и последний отрезки являются касательными к кривой в точках P_0 и P_3 соответственно, т. е. касательные векторы в этих точках направлены вдоль $\boldsymbol{a}_1$ и $\boldsymbol{a}_3$. Для его вычисления продифференцируем уравнение:

$$\boldsymbol{r}'(t) = f'_{3,1}(t)\boldsymbol{a}_1 + f'_{3,2}(t)\boldsymbol{a}_2 + f'_{3,3}(t)\boldsymbol{a}_3,$$

и получим условия в виде:

$$\left.\begin{aligned}\boldsymbol{r}'(0) &= f'_{3,1}(0)\boldsymbol{a}_1 + f'_{3,2}(0)\boldsymbol{a}_2 + f'_{3,3}(0)\boldsymbol{a}_3 = k\boldsymbol{a}_1\\ \boldsymbol{r}'(1) &= f'_{3,1}(1)\boldsymbol{a}_1 + f'_{3,2}(1)\boldsymbol{a}_2 + f'_{3,3}(1)\boldsymbol{a}_3 = k\boldsymbol{a}_3\end{aligned}\right\}$$
$$\rightarrow f'_{3,2}(0) = f'_{3,3}(0) = f'_{3,1}(1) = f'_{3,2}(1) = 0.$$

Промежуточные участки ломаной используются для управления кривой путем более глубокого влияния на ее геометрию в двух концевых точках. Для этого используется вторая производная, которая определяется только векторами $\{\boldsymbol{a}_1, \boldsymbol{a}_2\}$ в P_0 и только векторами $\{\boldsymbol{a}_2, \boldsymbol{a}_3\}$ в P_3. Взятие вторых производных дает

$$\boldsymbol{r}''(t) = f''_{3,1}(t)\boldsymbol{a}_1 + f''_{3,2}(t)\boldsymbol{a}_2 + f''_{3,3}(t)\boldsymbol{a}_3,$$

а применение этих условий дает

$$\left.\begin{aligned}\boldsymbol{r}''(0) &= f''_{3,1}(0)\boldsymbol{a}_1 + f''_{3,2}(0)\boldsymbol{a}_2 + f''_{3,3}(0)\boldsymbol{a}_3 = k\boldsymbol{a}_1 + l\boldsymbol{a}_2\\ \boldsymbol{r}''(1) &= f''_{3,1}(1)\boldsymbol{a}_1 + f''_{3,2}(1)\boldsymbol{a}_2 + f''_{3,3}(1)\boldsymbol{a}_3 = k\boldsymbol{a}_2 + l\boldsymbol{a}_3\end{aligned}\right\}$$
$$\rightarrow f''_{3,3}(0) = f''_{3,1}(1) = 0.$$

Итак, мы имеем двенадцать независимых условий на функции, выступающие в роли коэффициентов. В итоге (но не с самого начала) Безье выбрал кубические полиномы, которых нужно три, каждый с четырьмя коэффициентами: $4 \times 3 = 12$ переменных. Для их нахождения достаточно элементарной алгебры. Мы проделаем это ниже, записав полиномы в виде

$$f_{3,i}(t) = \alpha_i t^3 + \beta_i t^2 + \gamma_i t + \delta_i,\ 1 \leqslant i \leqslant 3$$

и применив условия поочередно:

$$f_{3,1}(0) = f_{3,2}(0) = f_{3,3}(0) = 0 \rightarrow \delta_1 = \delta_2 = \delta_3 = 0,$$
$$f_{3,1}(1) = f_{3,2}(1) = f_{3,3}(1) = 1 \rightarrow \left.\begin{cases}\alpha_1 + \beta_1 + \gamma_1 + \delta_1 = 1,\\ \alpha_2 + \beta_2 + \gamma_2 + \delta_2 = 1,\\ \alpha_3 + \beta_3 + \gamma_3 + \delta_3 = 1,\end{cases}\right\} \rightarrow \begin{cases}\alpha_1 + \beta_1 + \gamma_1 = 1,\\ \alpha_2 + \beta_2 + \gamma_2 = 1,\\ \alpha_3 + \beta_3 + \gamma_3 = 1,\end{cases}$$

$$f'_{3,i}(t) = 3\alpha_i t^2 + 2\beta_i t + \gamma_i,$$

$$f'_{3,2}(0) = \gamma_2 = f'_{3,3}(0) = \gamma_3 = 0 \rightarrow \gamma_2 = \gamma_3 = 0,$$
$$f'_{3,1}(1) = f'_{3,2}(1) = 0 \rightarrow \left.\begin{cases}3\alpha_1 + 2\beta_1 + \gamma_1 = 0,\\ 3\alpha_2 + 2\beta_2 + \gamma_2 = 0,\end{cases}\right\} \rightarrow \begin{cases}\gamma_2 = \gamma_3 = 0,\\ \begin{cases}3\alpha_1 + 2\beta_1 + \gamma_1 = 0,\\ 3\alpha_2 + 2\beta_2 = 0,\end{cases}\end{cases}$$

$$f''_{3,3}(0) = f''_{3,1}(1) = 0 \rightarrow \begin{cases}2\beta_3 = 0 \rightarrow \beta_3 = 0,\\ 6\alpha_1 + 2\beta_1 = 0.\end{cases}$$

Эта система уравнений имеет единственное решение:

$$\begin{pmatrix}\alpha_1 & \alpha_2 & \alpha_3\\ \beta_1 & \beta_2 & \beta_3\\ \gamma_1 & \gamma_2 & \gamma_3\\ \delta_1 & \delta_2 & \delta_3\end{pmatrix} = \begin{pmatrix}1 & -2 & 1\\ -3 & 3 & 0\\ 3 & 0 & 0\\ 0 & 0 & 0\end{pmatrix},$$

а значит, векторное параметрическое уравнение кривой имеет вид:

$$\boldsymbol{r}(t) = \boldsymbol{a}_0 + (t^3 - 3t^2 + 3t)\boldsymbol{a}_1 + (-2t^3 + 3t^2)\boldsymbol{a}_2 + t^3\boldsymbol{a}_3,\ 0 \leqslant t \leqslant 1.$$

Выбор векторов, определяющих ломаную, определяет и кривую, а их динамическое изменение приводит к динамическому изменению формы кривой. И конечно же, процедура обобщается на n отрезков и $n + 1$ вершин $\{P_0, P_1, P_2, \ldots, P_n\}$, где $\{\boldsymbol{a}_1 = \boldsymbol{p}_1 - \boldsymbol{p}_0, \boldsymbol{a}_2 = \boldsymbol{p}_2 - \boldsymbol{p}_1, \boldsymbol{a}_3 = \boldsymbol{p}_3 - \boldsymbol{p}_2, \ldots, \boldsymbol{a}_n = \boldsymbol{p}_n - \boldsymbol{p}_{n-1}\}$, а уравнение кривой принимает вид:

$$\boldsymbol{r}(t) = \boldsymbol{a}_0 + \sum_{i=1}^{n} f_{n,i}(t)\boldsymbol{a}_i \quad \text{for } 0 \leqslant t \leqslant 1.$$

Тогда условия имеют вид:

$$\begin{aligned}
f_{n,i}(0) &= 0, \quad 1 \leqslant i \leqslant n,\\
f_{n,i}(1) &= 1, \quad 1 \leqslant i \leqslant n,\\
f_{n,i}^{(m)}(0) &= 0, \quad m \leqslant i \leqslant n,\ 1 \leqslant m \leqslant n-1,\\
f_{n,i}^{(m)}(1) &= 0, \quad 1 \leqslant i \leqslant n-m,\ 1 \leqslant m \leqslant n-1.
\end{aligned}$$

Элементарная алгебра должна уступить место более глубокому анализу, но Безье установил совершенно невероятный вид функциональных коэффициентов:

$$\begin{aligned}
f_{n,i}(t) &= \frac{(-1)^i}{(i-1)!} t^i \frac{\mathrm{d}^{i-1}}{\mathrm{d}t^{i-1}} \frac{(1-t)^n - 1}{t}\\
&= \sum_{j=1}^{n} (-1)^{i+j} \binom{n}{j} \binom{j-1}{i-1} t^j.
\end{aligned}$$

Их вывод вовсе не так страшен, как может показаться, но мы решили не приводить его, отчасти за недостатком места, а отчасти чтобы избежать усложненного повторения уже достигнутого. Но главным образом потому, что на практике кубических полиномов вполне достаточно для тех целей, для которых используется построение Безье. На самом деле даже одной управляющей точки и квадратичного приближения хватает для многих целей, и мы оставляем читателю в качестве упражнения провести похожее, но более простое рассуждение и получить такое уравнение кривой:

$$\boldsymbol{r}(t) = \boldsymbol{a}_0 + (-t^2 + 2t)\boldsymbol{a}_1 + t^2\boldsymbol{a}_2.$$

Наконец, мы вскользь упомянули, что Безье экспериментировал и с другими формами функциональных коэффициентов; он – по необходимости, а мы – из любопытства. Приглашаем интересующегося читателя вернуться к трехзвенной ломаной, но в качестве коэффициентов выбрать функции, порожденные базисом:

$$\{\sin \tfrac{1}{2}\pi t, \cos \tfrac{1}{2}\pi t, \sin^2 \tfrac{1}{2}\pi t, \cos^2 \tfrac{1}{2}\pi t\} \quad \text{for } 0 \leqslant t \leqslant 1,$$

которые, следовательно, имеют вид:

$$f_{3,i}(t) = \alpha_i \sin \tfrac{1}{2}\pi t + \beta_i \cos \tfrac{1}{2}\pi t + \gamma_i \sin^2 \tfrac{1}{2}\pi t + \delta_i \cos^2 \tfrac{1}{2}\pi t.$$

Соответствующие вычисления приведены в приложении С.

По счастью, компания «Рено» не стала защищать патентом работу своего служащего, а позволила Безье опубликовать свои (или профессора Дюрана) идеи, привлекшие внимание профессора Робина Форреста, который тогда работал на факультете вычислительной математики Кембриджского университета. Именно он заметил, что формулировка работы Безье с помощью вершин ломаной, а не самих отрезков была бы элегантнее и удобнее для вычислений (Forrest 1972). В терминах квадратичной и кубической полиномиальных форм поправка принимает вид:

$$\begin{aligned}\boldsymbol{r}(t) &= \boldsymbol{a}_0 + (-t^2 + 2t)\boldsymbol{a}_1 + t^2\boldsymbol{a}_2 \\ &= \boldsymbol{p}_0 + (-t^2 + 2t)(\boldsymbol{p}_1 - \boldsymbol{p}_0) + t^2(\boldsymbol{p}_2 - \boldsymbol{p}_1) \\ &= (1 + t^2 - 2t)\boldsymbol{p}_0 + (-2t^2 + 2t)\boldsymbol{p}_1 + t^2\boldsymbol{p}_2 \\ &= (1 - t)^2\boldsymbol{p}_0 + 2t(1 - t)\boldsymbol{p}_1 + t^2\boldsymbol{p}_2\end{aligned}$$

и

$$\begin{aligned}\boldsymbol{r}(t) &= \boldsymbol{a}_0 + (t^3 - 3t^2 + 3t)\boldsymbol{a}_1 + (-2t^3 + 3t^2)\boldsymbol{a}_2 + t^3\boldsymbol{a}_3 \\ &= \boldsymbol{p}_0 + (t^3 - 3t^2 + 3t)(\boldsymbol{p}_1 - \boldsymbol{p}_0) + (-2t^3 + 3t^2)(\boldsymbol{p}_2 - \boldsymbol{p}_1) \\ &\quad + t^3(\boldsymbol{p}_3 - \boldsymbol{p}_2) \\ &= (1 - t^3 + 3t^2 - 3t)\boldsymbol{p}_0 + (3t^3 - 6t^2 + 3t)\boldsymbol{p}_1 \\ &\quad + (-3t^3 + 3t^2)\boldsymbol{p}_2 + t^3\boldsymbol{p}_3 \\ &= (1 - t)^3\boldsymbol{p}_0 + 3t(1 - t)^2\boldsymbol{p}_1 + 3t^2(1 - t)\boldsymbol{p}_2 + t^3\boldsymbol{p}_3\end{aligned}$$

соответственно.

В таком виде мы можем записать кривую и работать с ней более естественным способом, выбрав две концевые и одну или две управляющие точки. Выражения очень удобны для построения графиков, и на самом деле все кривые в этой главе были так и нарисованы. На рис. 3.2 мы выбрали две произвольные точки на кривой (3, 4) и (27, 2) и произвольную управляющую точку (11, 18), которые дают векторное квадратное уравнение:

$$\begin{pmatrix} x \\ y \end{pmatrix} = (1 - t)^2 \begin{pmatrix} 3 \\ 4 \end{pmatrix} + 2t(1 - t) \begin{pmatrix} 11 \\ 18 \end{pmatrix} + t^2 \begin{pmatrix} 27 \\ 2 \end{pmatrix}$$

и две квадратичные параметризации кривой:

$$\begin{aligned} x &= 3(1 - t)^2 + 22t(1 - t) + 27t^2, \\ y &= 4(1 - t)^2 + 36t(1 - t) + 2t^2. \end{aligned}$$

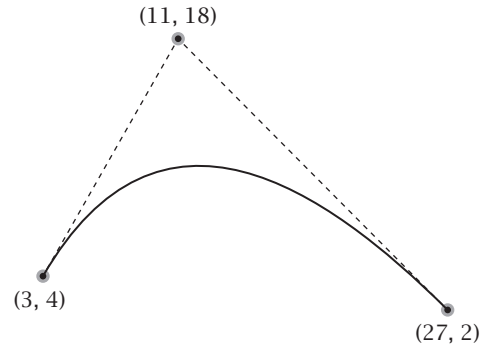


Рис. 3.2. Квадратичная поправка Форреста

В случае кубической формы на кривой выбраны точки (2, 1) и (12, 6), а также две управляющие точки (4, 7) и (10, 9). В векторной параметрической форме имеем

$$\begin{pmatrix} x \\ y \end{pmatrix} = (1-t)^3 \begin{pmatrix} 2 \\ 1 \end{pmatrix} + 3t(1-t)^2 \begin{pmatrix} 4 \\ 7 \end{pmatrix} + 3t^2(1-t) \begin{pmatrix} 10 \\ 9 \end{pmatrix} + t^3 \begin{pmatrix} 12 \\ 6 \end{pmatrix},$$

а в скалярном виде два кубических параметрических уравнения:

$$x = 2(1-t)^3 + 12t(1-t)^2 + 30t^2(1-t) + 12t^3,$$
$$y = (1-t)^3 + 21t(1-t)^2 + 27t^2(1-t) + 6t^3.$$

Сама кривая показана на рис. 3.3.

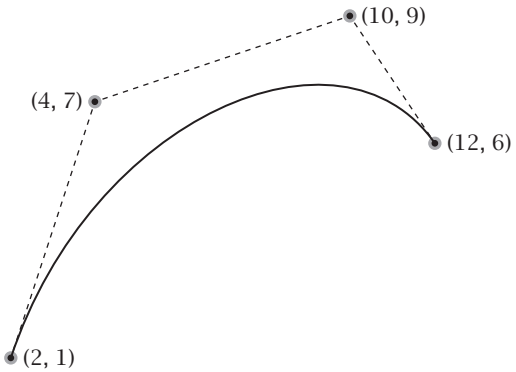


Рис. 3.3. Кубическая поправка Форреста

В такой формулировке очевидно, что параметризации описывают соответственно параболу и кубическую кривую; первые две (самые практически полезные) кривые Безье представляют собой эти самые элементарные полиномы; это имеет как положительные (быстрота и простота вычислений и манипулирования), так и отрицательные стороны (никакими манипуляциями из них нельзя получить, к примеру, трансцендентные кривые). Для полноты включим еще вырожденную форму:

$$\boldsymbol{r}(t) = (1 - t)\boldsymbol{p}_0 + t\boldsymbol{p}_1,\ 0 \leqslant t \leqslant 1,$$

которая вообще не содержит управляющих точек. Разумеется, это не что иное, как векторное параметрическое уравнение отрезка прямой с концевыми точками P_0 и P_1.

На рис. 3.4 показаны графики функциональных коэффициентов для линейной, квадратичной и кубической кривых Безье. Видно, как они изменяются, когда параметр t пробегает отрезок [0,1]. В случае кубической аппроксимации в первой четверти отрезка доминирует член $(1 - t)^3$, во второй четверти – член $3t(1 - t)^2$, в третьей четверти – член $3t^2(1 - t)$, а в последней – член t^3. Эти коэффициенты, которые обычно называют *стыковочными* функциями, можно рассматривать как переменные веса, присоединенные к соответственным вершинам, влияние которых изменяется, как описано выше.

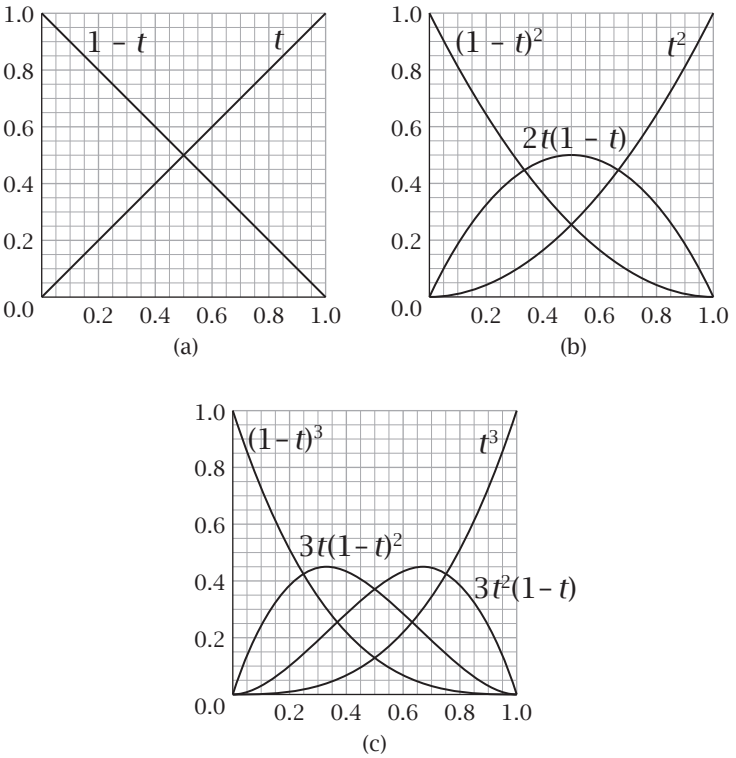


Рис. 3.4. Функциональные коэффициенты

Именно благодаря этим стыковочным функциям соединяются различные математические идеи. Во-первых, они являются членами биномиального разложения:

$$1 = 1^n = (t + (1 - t))^n = \sum_{i=0}^{n} \binom{n}{i} t^i (1 - t)^{n-i} = \sum_{i=0}^{n} B_i^n(t).$$

Нотация $B^n_i(t)$ отражает тот факт, что они еще известны как *полиномы Бернштейна:*

$$B_i^n(t) = \binom{n}{i} t^i (1-t)^{n-i}, \quad 1 \leqslant i \leqslant n,\ 0 \leqslant t \leqslant 1,$$

в знак признания заслуг Сергея Натановича Бернштейна (1880–1968), российского математика, который использовал их в качестве базиса пространства полиномов степени n с одной переменной t: $\{1, B_0^n, B_1^n, B_2^n, \ldots, B_n^n\}$ вместо обычного $\{1, t, t^2, t^3, \ldots, t^n\}$.

Интересно, что их роль в качестве компонентов кривых Безье можно интерпретировать как нечто прямо противоположное первоначальному назначению. Бернштейн поставил перед собой цель дать первое конструктивное доказательство *теоремы Вейерштрасса о приближении*. Этот знаменитый результат, относящийся к полезности полиномов в математике, словами можно выразить так: «Полиномами достаточно высокой степени можно равномерно аппроксимировать любую непрерывную функцию на любом конечном интервале», а в символьном виде – как

$$|f(x) - p_n(x)| < \varepsilon \text{ для всех } x \in [a, b].$$

n-й полином Бернштейна был одним из компонентов n-й полиномиальной функции Бернштейна:

$$B_n(f, t) = \sum_{i=0}^{n} f\left(\frac{i}{n}\right) \binom{n}{i} t^i (1-t)^{n-i},$$

для которой, как доказал Бернштейн,

$$\lim_{n\to\infty} B_n(f, t) = f(t)$$

на отрезке $t \in [0, 1]$. Из того, что преобразование

$$t \to \frac{t-a}{b-a}$$

переводит $[a, b] \to [0, 1]$, вытекает общий результат.

Мы решили проиллюстрировать этот результат на примере функции $f(x)$, описывающей ломаную, проходящую через точки (0, 0), (0.2, 0.6), (0.6, 0.8), (0.9, 0.7), (1, 0), а на рис. 3.5 показано ее сравнение с полиномиальными аппроксимациями (Davis 2014, стр. 116):

$$\begin{aligned} B_2(f, t) &= \tfrac{3}{2}(t - t^2), \\ B_4(f, t) &= \tfrac{5}{2}t - 3t^2 + \tfrac{3}{2}t^3 - t^4, \\ B_{10}(f, t) &= 3t - 30t^3 + 105t^4 - 189t^5 + 210t^6 \\ &\quad - 160t^7 + 90t^8 - 35t^9 + 6t^{10}. \end{aligned}$$

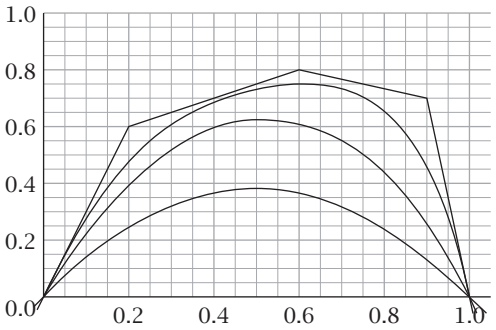


Рис. 3.5. Теорема Вейерштрасса о приближении

Итак, в этом примере теоремы Вейерштрасса кривые, определенные с помощью полиномов Бернштейна, аппроксимируют ломаную линию; в приложении же Безье те же самые полиномы используются для того, чтобы аппроксимировать кривую ломаной линией. Кроме того, кривые сходятся к ломаной очень медленно (Davis 2014, стр. 116), что не имеет большого теоретического значения, но налагает практические ограничения.

> «Похоже, этот факт препятствует применению полиномов Бернштейна в численных методах. Быть может, они найдут применение, когда свойства аппроксимирующих функций в целом окажутся важнее точности аппроксимации».

Это замечание было впервые опубликовано в 1963 г., когда публике еще не было известно, что функций $B_i^2(t)$ и $B_i^3(t)$ вполне достаточно для самой насущной практической задачи аппроксимации кривой.

И есть еще один важный аспект истории о кривых Безье, который не был широко известен до середины 1970-х гг. Им мы сейчас и займемся.

3.3. Безье и де Кастельжо

В основном благодаря работе молодого математика Поля де Фаже де Кастельжо другой французский автомобильный гигант, Ситроен, также решал проблемы устаревшего процесса проектирования. Но, в отличие от Рено, Ситроен присваивал плоды трудов своих служащих. Кастельжо начал работать в компании «Ситроен» в 1959 году, а в работе, опубликованной в 1963 г., – что характерно, в издании «Акционерного общества Ситроен», – он распространил свои идеи внутри организации. Название работы «Courbes et surfaces à pôles» (Кривые и поверхности при помощи отрезков) хранилось в секрете до 1975 г., когда Вольфганг Бём[1] открыл то, что сейчас известно под названием *алгоритм де Кастельжо*.

Отрезками в этом контексте называются линии, соединяющие некоторые точки согласно алгоритму. Оказывается, что это совершенно другая, но полностью эквивалентная формулировка кривых Безье, более пригодная для вычислений на компьютере.

[1] Лауреат премии Пьера Безье за 2017 год.

Берем две точки P_0, P_2, которые станут концами кривой, и третью управляющую точку P_1, которая не лежит на кривой, но определяет ее форму. То есть первоначально кривая определяется множеством точек $\{P_0, P_1, P_2\}$, а затем к нему добавляется по одной точке следующим образом.

Предположим, что точка Q движется вдоль отрезка P_0P_1, так что в момент времени t, $0 \leqslant t \leqslant 1$, ее радиус-вектор имеет вид $\boldsymbol{q} = t\boldsymbol{p}_0 + (1 - t)\boldsymbol{p}_1$. Аналогично вторая точка Q движется вдоль отрезка P_1P_2, так что в момент времени t ее радиус-вектор имеет вид $\boldsymbol{q} = t\boldsymbol{p}_1 + (1 - t)\boldsymbol{p}_2$. На рис. 3.6 изображена ситуация, когда точка на отрезке P_0P_1 достигла положения Q_0, а точка на отрезке P_1P_2 достигла положения Q_1. Тем самым определен отрезок Q_0Q_1, и мы воображаем, что одновременно с этими двумя движениями происходит третье движение, при котором точка Q перемещается из Q_0 в Q_1 и достигает положения $\boldsymbol{r} = t\boldsymbol{q}_0 + (1 - t)\boldsymbol{q}_1$. Точка R лежит на определяемой кривой.

В символьном виде все это записывается так:

$$\boldsymbol{q}_0 = t\boldsymbol{p}_0 + (1 - t)\boldsymbol{p}_1,$$
$$\boldsymbol{q}_1 = t\boldsymbol{p}_1 + (1 - t)\boldsymbol{p}_2,$$
$$\boldsymbol{r} = t\boldsymbol{q}_0 + (1 - t)\boldsymbol{q}_1 = t[t\boldsymbol{p}_0 + (1 - t)\boldsymbol{p}_1] + (1 - t)[t\boldsymbol{p}_1 + (1 - t)\boldsymbol{p}_2],$$
$$\boldsymbol{r} = t^2\boldsymbol{p}_0 + 2t(1 - t)\boldsymbol{p}_1 + (1 - t)^2\boldsymbol{p}_2.$$

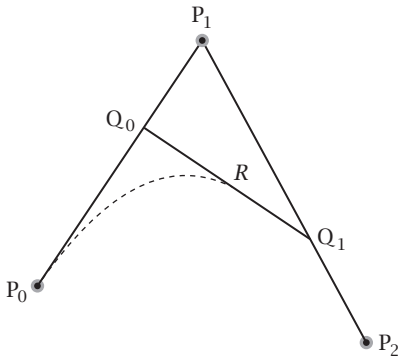


Рис. 3.6. Метод де Кастельжо

И мы имеем квадратичную кривую Безье, построенную способом, не имеющим ничего общего с использованием полиномов Бернштейна. Какой способ лучше? Зависит от обстоятельств; все кривые в этой главе были нарисованы с помощью полиномов Бернштейна, но описанная только что альтернатива может оказаться весьма привлекательной из-за одного свойства – и архитектуры арифметико-логического устройства компьютера. Чтобы разобраться в этом, продолжим рассматривать квадратичный вариант и будем ссылаться на рис. 3.6. Остановим динамический процесс при некотором фиксированном значении t и рассмотрим обе кривые, определенные тремя точками: $\{P_0, Q_0, R\}$ и $\{R, Q_1, P_2\}$. Для произвольной точки на первой кривой имеем

$$\begin{aligned}\boldsymbol{s} &= \lambda^2\boldsymbol{p}_0 + 2\lambda(1-\lambda)\boldsymbol{q}_0 + (1-\lambda)^2\boldsymbol{r}\\ &= \lambda^2\boldsymbol{p}_0 + 2\lambda(1-\lambda)[t\boldsymbol{p}_0 + (1-t)\boldsymbol{p}_1]\\ &\quad + (1-\lambda)^2[t^2\boldsymbol{p}_0 + 2t(1-t)\boldsymbol{p}_1 + (1-t)^2\boldsymbol{p}_2]\\ &= [\lambda^2 + 2\lambda(1-\lambda)t + (1-\lambda)^2t^2]\boldsymbol{p}_0\\ &\quad + [2\lambda(1-\lambda)(1-t) + 2t(1-t)(1-\lambda)^2]\boldsymbol{p}_1 + (1-\lambda)^2(1-t)^2\boldsymbol{p}_2.\end{aligned}$$

Теперь определим параметр T как $T = \lambda + t(1-\lambda)$. Очевидно, что при изменении λ для любого t имеем $T \geqslant 0$. Кроме того, поскольку $(1-\lambda)(t-1) \leqslant 0$, должно быть также $T \leqslant 1$. Итак, $0 \leqslant T \leqslant 1$, и, так как $1 - T = (1-\lambda)(1-t)$, мы можем записать

$$\boldsymbol{s} = T^2\boldsymbol{p}_0 + 2T(1-T)\boldsymbol{p}_1 + (1-T)^2\boldsymbol{p}_2,$$

и это точка на нашей первоначальной кривой – ее первой части, которую она будет пробегать. Эта процедура, называемая *разбиением кривой*, может быть реализована на кривых Безье любой степени и является одной из нескольких операций, которые к ним обычно применяются. Для наших целей возьмем $t = 0.5$, это означает, что Q_0, Q_1 и R – центры соответствующих отрезков. Теперь повторим все это, чтобы получить точку на другой половине кривой, а затем будем повторять процесс, деля кривую пополам и находя точку на первоначальной кривой, пока новая точка не окажется на расстоянии одного пикселя от предыдущей, так что крохотный участок первоначальной кривой является достаточно плоским, чтобы его можно изобразить в виде отрезка прямой. С точки зрения компьютерного экрана (или принтера) кривая нарисована. Человеку этот итеративный процесс вряд ли понравится, но для компьютера это прекрасная рекурсивная процедура, требующая только сложения координат точек и деления пополам, а такие операции выполняются очень быстро.

В последних разделах этой главы мы рассмотрим два применения кривых Безье: первое из них шуточное, но второе – что угодно, только не шутка.

3.4. История Лумпа

За два года до того, как Кастельжо поступил на работу в «Ситроен», Пабло Пикассо принимал друга, известного фотожурналиста Дэвида Дугласа Дункана, на своей вилле Калифорния в Каннах. Дункана сопровождала его такса Лумп (по-немецки *прохвост*), которая сразу же завладела сердцем Пикассо и осталась с ним на следующие шесть лет. Между Лумпом, боксером Пикассо Жаном и его козой Эсмеральдой сложились дружеские, а между Пикассо и Лумпом – прямо-таки сердечные отношения:

> «Это была настоящая любовь. Пикассо брал Лумпа на руки. Он кормил его из своих рук. Черт побери, этот песик стал хозяином. Он мог делать в доме все, что пожелает!»

Такова точка зрения Дункана, изложенная им самим в книге «Лумп: собака, которая съела Пикассо»[1]. Неудивительно, что Лумп принял на себя роль музы

[1] Издательство Ad Marginem, 2016.

художника и был запечатлен на многочисленных работах, включая интерпретацию Пикассо картины Диего Веласкеса «Менины», на которой мастиф был заменен этой таксой. Существует много рисунков собаки, и в начале этой главы приведена наша вариация на тему одного из них – ясно, что Пикассо нарисовал его одним росчерком пера, а нам потребовалось девять кубических кривых Безье. Мы полагаем, что Пикассо интересовала просто непрерывность линии, а не какая-то особая гладкость в точках соединения (которые на жаргоне называются *узлами*), и наш выбор кривых это отражает. Данные приведены ниже, а рисунок фрагментирован наиболее естественным образом и в том порядке, в каком, как нам кажется, его нарисовал бы художник-левша: начиная с точки соединения рта с ухом, далее по часовой стрелке и заканчивая передней лапой.

На рис. 3.7 показаны точки соединения участков, а сами участки определены как кубические кривые Безье со следующими координатами:

{(55, 60), (40, 67), (15, 53), (1, 60)},
{(1, 60), (57, 96), (37, 86), (64, 110)},
{(64, 110), (85, 120), (83, 25), (68, 25)},
{(68, 25), (50, 35), (50, 106), (79, 95)},
{(79, 95), (105, 60), (225, 80), (290, 107)},
{(290, 107), (295, 104), (285, 93), (268, 90)},
{(270, 90), (275, 0), (265, 10), (255, 54)},
{(255, 54), (215, 70), (205, 56), (106, 56)},
{(106, 56), (101, 44), (98, 32), (95, 20)}.

Разрывность соответствует небольшому перекрытию в том месте, где хвост соединяется с верхней частью задней лапы.

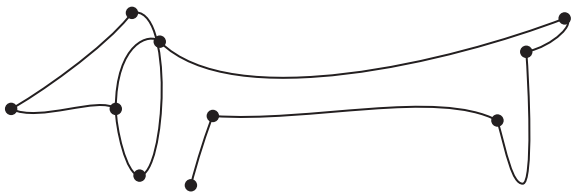


Рис. 3.7. Лump, состоящий из девяти кривых Безье

Лump умер 29 марта 1973 г., за десять дней до смерти самого Пикассо.

3.5. История буквы O

Второе применение кривых Безье куда более серьезно. На форзаце английского издания этой книги есть фраза «This book has been composed in LucidaBright»[1] (хотя мы-то набирали текст шрифтом Times Roman). Название Lucida происходит от слова «lucid»: четкий, легкий для чтения. Этот шрифт состоит из расширенного семейства взаимосвязанных гарнитур и был создан Чарльзом Бигелоу и Крисом Холмсом. Первая версия появилась в 1984 г., но с тех

[1] Эта книга набрана шрифтом LucidaBright. – *Прим. перев.*

пор он модифицировался, и текущая версия датируется 2012 г., группа TUG (TEX Users Group) выпустила шрифт Lucida OpenType Math. А что такое шрифт OpenType? Прежде чем ответить на этот вопрос, мы должны рассмотреть шрифты PostScript и TrueType.

Электронные шрифты – чрезвычайно сложные структуры, но, как и для большинства технических вопросов, связанных с компьютерами, знать (а тем более глубоко понимать) детали, к счастью, необязательно. Мы видим *глиф* на клавиатуре, нажимаем соответствующую клавишу, чтобы передать *литеру* процессору, – и на экране появляется этот глиф в выбранной нами гарнитуре из определенного *шрифта*. А если мы решим напечатать его на бумаге, то не ожидаем никаких сюрпризов. Что касается самого шрифта, программное обеспечение обычно предлагает много вариантов, скачать можно еще больше – иногда бесплатно, а иногда за деньги. Итак, мы нажимаем клавишу и ожидаем, что изображенный на ней глиф появится на экране в выбранном шрифте, размере и цвете, поскольку программа запросит необходимую для этого информацию у операционной системы. Но какую именно информацию? Иначе говоря, как сохранить символ (или, точнее, код команды), чтобы его можно было извлечь быстро, надежно и сколько угодно раз? Раньше шрифты хранились в виде растров, когда глифы описывались прямоугольными массивами черно-белых пикселей; чем выше разрешающая способность, тем больше требовалось пикселей. Дональд Кнут спроектировал систему METAFONT для сжатия растровых шрифтов и в 1978 г. включил ее в свою систему TEX, дальнейшее развитие которой остается основным способом работы с технической информацией для типографского набора.

Вторым решением, предложенным в 1985 г., мы обязаны Джону Уорноку, основателю компании Adobe. Он разработал язык программирования *PostScript*, позволяющий обращаться к странице с помощью определенных математических конструкций. В частности, язык PostScript включает формат шрифта, который по сей день остается самым распространенным в мире: *шрифты Type 1*, являющиеся примером *векторных шрифтов*.

Благодаря успеху PostScript и шрифтов Type 1 команда Adobe преуспевала, а Apple и Microsoft поставили задачей разрушить ее монополию. Их совместное детище, призванное составить конкуренцию шрифтам Type 1, получило название *TrueType*. Apple и Microsoft по отдельности приступили к работе по улучшению шрифтов TrueType, и в итоге Apple создала технологию *TrueType GX* (позже переименованную в AAT – Apple Advanced Typography). Microsoft объединилась со своим бывшим конкурентом Adobe, и совместно они представили соперника TrueType GX: *OpenType*. Последнее осложнение заключалось в том, что шрифты OpenType затем распались на две категории: *OpenType-TTF* (TrueType с несколькими дополнительными функциями) и *OpenType-CFF* (расширенные шрифты Type 1, объединенные со структурами TrueType).

Таким образом, язык PostScript посредством своего интерпретатора *Ghostscript* можно использовать для описания литер со всеми их частями: *основной штрих*, *верхний выносной элемент*, *нижний выносной элемент*, *овал*, *внутрибуквенный просвет*, *концевой элемент*, *перекладина*, *засечка и хвост*. С ними

ассоциированы команды *moveto*, *lineto*, *curveto* и *closepath*, каждая из которых выбирает свои параметры из стека данных. Именно в команду curveto встроена процедура рисования кубической кривой Безье, и мы можем проиллюстрировать ее использование в шрифте Ghostscript/Times Roman на примере кодирования буквы О.

Каждая строка набора данных

$$\left\{\begin{pmatrix}0.360998541\\0.673999\end{pmatrix},\begin{pmatrix}0.169997558\\0.673999\end{pmatrix},\begin{pmatrix}0.039997559\\0.530835\end{pmatrix},\begin{pmatrix}0.039997559\\0.33099854\end{pmatrix}\right\},$$

$$\left\{\begin{pmatrix}0.039997559\\0.33399854\end{pmatrix},\begin{pmatrix}0.039997559\\0.236999512\end{pmatrix},\begin{pmatrix}0.0697631836\\0.145998538\end{pmatrix},\begin{pmatrix}0.119995117\\0.0879980475\end{pmatrix}\right\},$$

$$\left\{\begin{pmatrix}0.119995117\\0.0879980475\end{pmatrix},\begin{pmatrix}0.177995607\\0.0249975584\end{pmatrix},\begin{pmatrix}0.266994625\\-0.0140014645\end{pmatrix},\begin{pmatrix}0.354995131\\-0.0140014645\end{pmatrix}\right\},$$

$$\left\{\begin{pmatrix}0.354995131\\-0.0140014645\end{pmatrix},\begin{pmatrix}0.551994622\\-0.0140014645\end{pmatrix},\begin{pmatrix}0.689997554\\0.125998542\end{pmatrix},\begin{pmatrix}0.689997554\\0.326999515\end{pmatrix}\right\},$$

$$\left\{\begin{pmatrix}0.689997554\\0.326999515\end{pmatrix},\begin{pmatrix}0.689997554\\0.425998539\end{pmatrix},\begin{pmatrix}0.660305202\\0.510998547\end{pmatrix},\begin{pmatrix}0.603996575\\0.570998549\end{pmatrix}\right\},$$

$$\left\{\begin{pmatrix}0.603996575\\0.570998549\end{pmatrix},\begin{pmatrix}0.540996075\\0.639997542\end{pmatrix},\begin{pmatrix}0.456999511\\0.673999\end{pmatrix},\begin{pmatrix}0.360998541\\0.673999\end{pmatrix}\right\}$$

содержит опорные точки и две промежуточные управляющие точки для внешней кривой буквы, а каждая строка набора данных

$$\left\{\begin{pmatrix}0.360998541\\0.63399905\end{pmatrix},\begin{pmatrix}0.406997085\\0.63399905\end{pmatrix},\begin{pmatrix}0.452998042\\0.618159175\end{pmatrix},\begin{pmatrix}0.488999\\0.58999753\end{pmatrix}\right\},$$

$$\left\{\begin{pmatrix}0.488999\\0.58999753\end{pmatrix},\begin{pmatrix}0.542998075\\0.540998518\end{pmatrix},\begin{pmatrix}0.58\\0.44699952\end{pmatrix},\begin{pmatrix}0.58\\0.328000486\end{pmatrix}\right\},$$

$$\left\{\begin{pmatrix}0.491999507\\0.0750024393\end{pmatrix},\begin{pmatrix}0.4569999511\\0.0400024429\end{pmatrix},\begin{pmatrix}0.411999524\\0.0260009766\end{pmatrix},\begin{pmatrix}0.358999\\0.0260009766\end{pmatrix}\right\},$$

$$\left\{\begin{pmatrix}0.491999507\\0.0750024393\end{pmatrix},\begin{pmatrix}0.456999511\\0.0400024429\end{pmatrix},\begin{pmatrix}0.411999524\\0.0260009766\end{pmatrix},\begin{pmatrix}0.358999\\0.0260009766\end{pmatrix}\right\},$$

$$\left\{\begin{pmatrix}0.358999\\0.0260009766\end{pmatrix},\begin{pmatrix}0.312998056\\0.0260009766\end{pmatrix},\begin{pmatrix}0.26799804\\0.04253418\end{pmatrix},\begin{pmatrix}0.232998043\\0.0710034147\end{pmatrix}\right\},$$

$$\left\{\begin{pmatrix}0.232998043\\0.0710034147\end{pmatrix},\begin{pmatrix}0.180998534\\0.117004395\end{pmatrix},\begin{pmatrix}0.15\\0.218005374\end{pmatrix},\begin{pmatrix}0.15\\0.329006344\end{pmatrix}\right\},$$

$$\left\{\begin{pmatrix}0.15\\0.329006344\end{pmatrix},\begin{pmatrix}0.15\\0.431005865\end{pmatrix},\begin{pmatrix}0.177197263\\0.528005362\end{pmatrix},\begin{pmatrix}0.217988043\\0.575004876\end{pmatrix}\right\},$$

$$\left\{\begin{pmatrix}0.217998043\\0.575004876\end{pmatrix},\begin{pmatrix}0.256997079\\0.618005395\end{pmatrix},\begin{pmatrix}0.30599609\\0.63399905\end{pmatrix},\begin{pmatrix}0.360996097\\0.63399905\end{pmatrix}\right\}$$

– то же самое для ее внутренней кривой. Получающаяся фигура показана на рис. 3.8, где внешняя кривая обходится против часовой стрелки, а внутренняя – по часовой стрелке, начиная с верхней точки; жирные точки обозначают переход от предыдущей кривой к следующей.

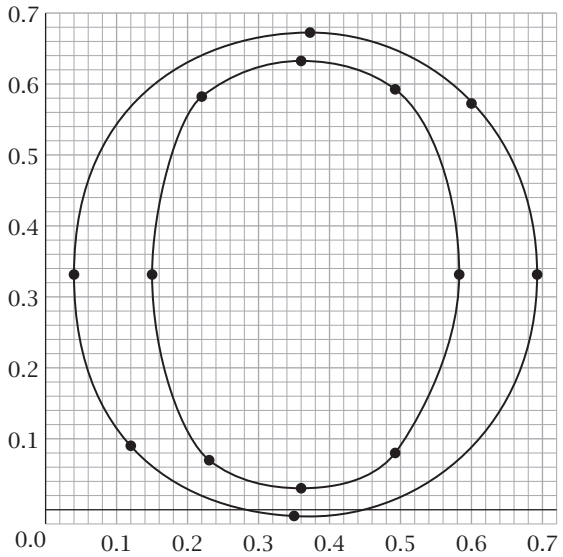


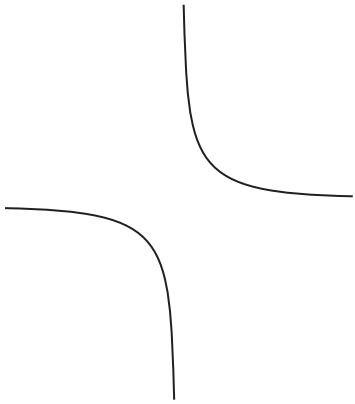
Рис. 3.8. Буква O шрифта Times Roman

Отсюда видно, что нарисовать букву «о» в слове «собака» куда труднее, чем саму собаку.

Наконец, мы упоминали о связи спирали Эйлера с переходными кривыми. Она использовалась также для рисования более общих кривых и прежде всего при проектировании шрифтов в качестве альтернативы кривым Безье: если кривые Безье основаны на использовании касательных, то в *спиро-кривых* применяется более мощное понятие кривизны (Levien 2009). К сожалению, у нас нет возможности проследить связь между главами 1 и 3, какой бы чарующей она ни казалась.

Глава 4

Равнобочная гипербола



Почему именно эта кривая?

Эта кривая привнесла в анализ логарифмы (т. е. таблицы чисел, призванных помочь в вычислениях). А с ними и экспоненциальную функцию. Повсеместное использование в дифференциальном и интегральном исчислении того, что мы теперь записываем в виде $\ln x$ и e^x, очевидно всякому, кто учился в старших классах средней школы и после того, а их роль в теоретическом изучении бесчисленных практических задач давно признана центральной. А началось-то все с равнобочной гиперболы.

4.1. Старые логарифмы

В своей первоначальной форме вычислительный инструмент, известный под названием *логарифмов*, служивший человечеству четыреста лет в качестве основного средства вычислений, был плодом двадцатилетнего единоличного труда шотландского дворянина Джона Непера, восьмого барона Мерчистона. Слово «логарифм» – второе название, придуманное Непером для своего изобретения, а первым было *искусственные числа*. Они были сведены в таблицу чисел, представляющих взаимосвязь между кинематикой двух связанных одномерных движений. Эти числа и инструкции по пользованию ими стали предметом книги «Mirifici logarithorum canonis descriptio» (Описание удивительной таблицы логарифмов), изданной в 1614 г. и переведенной на ан-

глийский в 1616 г., главным образом под влиянием английского математика и картографа Эдварда Райта. Полезность логарифмов была улучшена в последние годы жизни Непера благодаря его совместной работе с английским математиком Генри Бриггсом. Вышедшая в 1624 г. книга Бриггса «Arithmetica Logarithmica» (Арифметика логарифмов) содержала первую полную таблицу того, что впоследствии получило название *бригговых логарифмов*, – логарифмов целых чисел от 1 до 20 000 и от 90 000 до 100 000 с точностью до 14 знаков после запятой. К тому времени понятие о природе логарифмов подверглось развитию, и было дано точное определение логарифма (в его тогдашнем понимании) без привлечения кинематики. Первая строка главы 1 «Arithmetica Logarithmica» гласит: «Логарифмы – это числа, которые, будучи присоединены к числам в пропорции, сохраняют равные разности». Тут нет никакого упоминания кинематики, но нет и ни слова об основании.

Таблица 4.1. Примеры логарифмов от Бриггса

	A	B	C	D
1	1	5	5	35
2	2	6	8	32
4	3	7	11	29
8	4	8	14	26
16	5	9	17	23
32	6	10	20	30
64	7	11	23	17
128	8	12	26	14
Число в пропорции	**Log**	**Log**	**Log**	**Log**

По счастью, Бриггс дополнил свое определение таблицей, которая воспроизведена в нашей табл. 4.1.

Числа в *(непрерывной) пропорции* – это степени 2 в первом столбце, а четыре ассоциированных с ними набора чисел, которые *сохраняют равные разности*, приведены в остальных столбцах. Понятие непрерывной пропорции заключается в том, что последовательность чисел $\alpha, \beta, \gamma, \delta, \ldots$ такова, что $\alpha / \beta = \beta / \gamma = \gamma / \delta = \ldots$; т. е. $\beta = \sqrt{\alpha\gamma}$ – среднее геометрическое α и γ, $\gamma = \sqrt{\beta\delta}$ – среднее геометрическое β и δ и т. д. В современной терминологии числа образуют геометрическую прогрессию $A, Ar, Ar^2, \ldots$. Числа, присоединенные к ним, которые сохраняют равные разности, – это $\alpha', \beta', \gamma', \delta', \ldots$ такие, что $\beta' - \alpha' = \gamma' - \beta' = \delta' - \gamma' = \ldots$; в современной терминологии это арифметическая прогрессия $a, a + d, a + 2d, \ldots$. Тогда ассоциация – это соответствие $Ar^n \leftrightarrow a + nd$ между числами вида Ar^n и бесконечным количеством их логарифмов вида $a + nd$. В данном случае Бриггс выбрал $A = 1$, $r = 2$ и четыре пары $(a, d) = (1, 1), (5, 1), (5, 3), (35, -3)$. Очевидное следствие из этого определения – то, что для любых четырех чисел, взятых из последовательности в непрерывной пропорции, должно выполняться соотношение:

$$\frac{p}{q} = \frac{r}{s} \leftrightarrow L(p) - L(q) = L(r) - L(s),$$

где $L(\cdot)$ обозначает логарифм. В частности, если мы хотим, чтобы имело место знакомое мультипликативное свойство логарифмов

$$pq{:}q = p{:}1 \leftrightarrow L(pq) - L(q) = L(p) - L(1),$$

то должно быть $L(1) = 0$. Вдвоем Непер и Бриггс в конце концов пришли к этому выбору, и в первом предложении главы 3 «Arithmetica Logarithmica» определяют свои логарифмы следующим образом:

> «Выбрав значение Логарифма единицы, мы можем поискать другое число – то, которое будет использоваться чаще всего и, безусловно, является самым необходимым, и этому числу мы можем сопоставить какой-нибудь удобный логарифм, который будет легко запомнить и копировать так часто, как необходимо. Из всех чисел ни одно не кажется более подходящим для этой цели, чем 10, а его логарифм пусть будет равен 1,00000,00000,0000».

Так появились на свет бригговы логарифмы – то, что сегодня мы называем десятичными логарифмами и что оказывало неоценимую помощь в вычислениях до середины 1970-х гг. и даже позже. Читатель, незнакомый с вычислением с применением логарифмов, может поискать сведения об их использовании в книгах и в интернете, если не для чего другого, так хотя бы просто чтобы оценить, сколько для этого требовалось труда и какое чудесное изобретение – электронный калькулятор. А мы, в свою очередь, должны оценить, каким чудесным подспорьем должны были казаться логарифмы в те времена, когда чуть ли не единственной альтернативой было выполнение длинных, утомительных, повторяющихся и пересыпанных ошибками вычислений.

Логарифмы больше не являются инструментом вычислений, но попрежнему занимают центральное место в математическом анализе благодаря их неожиданной роли в решении задач, далеких от вычислений, но крайне важных для анализа.

4.2. Трудная проблема

Квадратура – исторический термин для неоднозначного процесса определения площади. Причины неоднозначности коренятся в ответах на вопросы, площадь чего определять и какими средствами. Согласно оригинальной древнегреческой интерпретации, разрешается использовать циркуль и линейку для построения квадрата той же площади, что и рассматриваемая фигура, или какой-то связанной с ней (отсюда и корень «квадрат» в слове «квадратура»): квадратура круга (построение квадрата, равновеликого кругу) – одна из самых долго не поддававшихся решению задач в истории математики (см. главу 5). Архимед значительно расширил арсенал средств, применив *метод исчерпывания Евдокса* для квадратуры параболы; по существу, он просуммировал бесконечную геометрическую прогрессию со знаменателем ¼ и установил, что площадь сегмента ABC на рис. 4.1 равна 1⅓ площади треугольника ABC, где C – точка касания прямой, параллельной хорде.

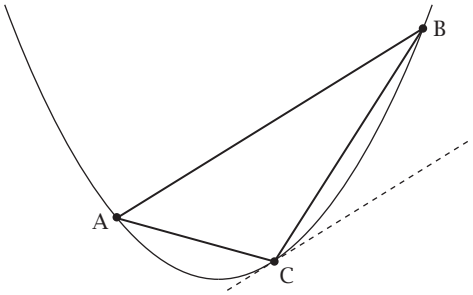


Рис. 4.1. Квадратура параболы Архимеда

Коль скоро известно, что площадь сегмента соизмерима с площадью треугольника, уже не составляет труда построить квадрат, равновеликий треугольнику, и тем самым выполнить квадратуру параболы. Но этот результат не то, что мы сегодня понимаем под «нахождением площади под параболой», и, чтобы очутиться на знакомой почве, нам следует перенестись на 1800 лет вперед. В 1640-х гг. Пьер де Ферма в работе, весьма далекой от его знаменитой «Великой теоремы», исследовал квадратуру кривых, получивших названия *высшие параболы* и *высшие гиперболы*, которые в современных обозначениях записываются в виде $y^m = x^n$ и $y^m = x^{-n}$ соответственно. Его результаты были опубликованы только в 1659 г. в труде под названием

«О преобразовании
и о преобразовании
и
упрощении уравнений геометрических мест точек
для сравнения всевозможных
криволинейных площадей с
прямолинейными площадями
и в то же время
об использовании геометрической прогрессии
для квадратуры парабол и гипербол до бесконечности».

Тогда краткие названия были редкостью.

В первом же предложении этой работы отдается дань Архимеду за использование бесконечных геометрических рядов при квадратуре параболы в античные времена. Сам же Ферма приспособил этот метод для формулирования в тогдашней терминологии гораздо более общих результатов, которые мы приведем в современной, привычной нам форме:

$$\int_0^a x^{n/m}\,\mathrm{d}x = \frac{a^{n/m+1}}{n/m+1} \quad \text{и} \quad \int_a^\infty x^{-n/m}\,\mathrm{d}x = \frac{a^{-n/m+1}}{-n/m+1}.$$

Правда, во втором случае, когда $m/n = 1$, имеется проблема, которую он сам признавал:

«…только для основной гиперболы, иначе говоря, простой гиперболы, или гиперболы Аполлония[1], этот метод не работает».

Что до самого метода (для высших гипербол с использованием современных символических обозначений, а не риторического изложения Ферма), то горизонтальная асимптота кривой – как мы понимаем, это бесконечная часть оси x – разбивается на интервалы, длины которых не одинаковы, а образуют бесконечную геометрическую прогрессию. На рис. 4.2 видно, что ось x от $x = a$ до бесконечности разбивается на отрезки с концами в точках a, ar, ar^2, ar^3, … , где $r > 1$, являющиеся основаниями прямоугольников с избытком, высоты которых для кривой с уравнением $y = 1/x^n$ равны соответственно

$$\frac{1}{a^n}, \quad \frac{1}{(ar)^n}, \quad \frac{1}{(ar^2)^n}, \quad \frac{1}{(ar^3)^n}, \quad \ldots.$$

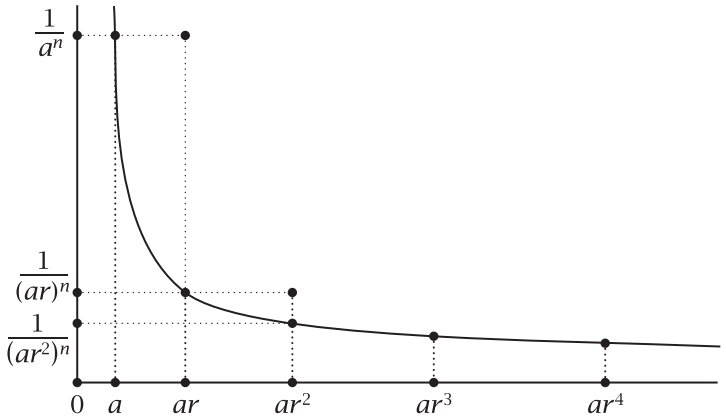


Рис. 4.2. Квадратура высшей гиперболы Ферма

Тогда сумма площадей этих прямоугольников равна

$$\begin{aligned}\sum_{i=1}^{\infty}(ar^i - ar^{i-1}) \times \frac{1}{(ar^{i-1})^n} &= \sum_{i=1}^{\infty}\frac{ar^{i-1}(r-1)}{a^n(r^{i-1})^n} \\ &= \frac{r-1}{a^{n-1}}\sum_{i=1}^{\infty}\frac{1}{(r^{i-1})^{n-1}} \\ &= \frac{r-1}{a^{n-1}}\sum_{i=1}^{\infty}\left(\frac{1}{r^{n-1}}\right)^{i-1} \\ &= \frac{r-1}{a^{n-1}} \times \frac{1}{1-1/r^{n-1}}\end{aligned}$$

[1] Который в своей восьмитомной работе «Конические сечения» и во многих других ввел в обиход хорошо теперь известные названия конических сечений. Ферма имеет в виду гиперболу, являющуюся коническим сечением, а не высшую гиперболу.

$$= \frac{r-1}{a^{n-1}} \times \frac{r^{n-1}}{r^{n-1}-1}$$
$$= \frac{r^{n-1}}{a^{n-1}} \times \frac{1}{r^{n-2}+r^{n-3}+\cdots+1},$$

где на последнем шаге упрощения мы воспользовались разложением двучлена $r^{n-1} - 1$ на множители.

Избавившись от члена $r - 1$, перейдем к пределу при r, стремящемся к 1 (и положим $r = 1$), и получим хорошо знакомый результат:

$$\int_a^\infty \frac{1}{x^n}\,\mathrm{d}x = \frac{1}{a^{n-1}} \times \frac{1}{n-1},$$

который, конечно, не имеет места при $n = 1$.

Снова процитируем Ферма:

> «Причина в том, что параллелограммы DH, EI, LK всегда равны. Члены, образующие прогрессию, принимая во внимание, что теперь они равны друг другу, не дают никакой разности, а именно разность – ключ ко всему делу. Я не привожу доказательства того, что для обычной гиперболы упомянутые параллелограммы всегда равны. Это видно сразу и легко выводится из свойства этой кривой: GD:HE = HA:GA. Тот же подход можно использовать для квадратуры всех (высших) парабол, за исключением одной, которая, как и гипербола, не поддается нашему методу».

Для нас это будут последние слова Ферма на интересующую нас тему, а буквы относятся к рис. 4.3, который позволит нам перейти к следующему выдающемуся персонажу: фламандскому иезуиту Грегуару де Сен-Венсану. В те годы, когда аргументы Ферма томились под спудом, проблему квадратуры равнобочной гиперболы изучали и другие люди, из которых самым примечательным был Сен-Венсан и его бывший ученик, а затем и коллега Альфонсо Антонио де Сараса.

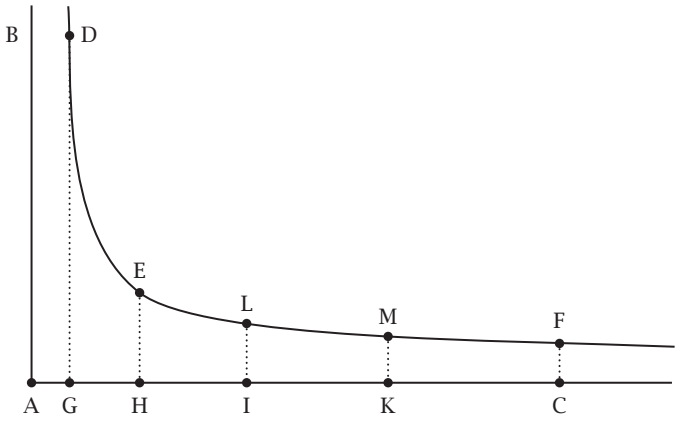


Рис. 4.3. Геометрические отрезки порождают арифметические площади

Принужденный покинуть Прагу в сумятице тридцатилетней войны, Сен-Венсан оставил свои математические работы, которые впоследствии были найдены, разобраны и составили его самое значительное математическое наследие: 1250-страничный том «Geometricum Quadraturae Circuli Sectionum Coni» (Геометрический труд о квадратуре круга и конических сечений), датированный 1647 г. Название говорит о том, что книга повествует о геометрии и конических сечениях, а также, к вящему ущербу для репутации автора, о якобы успешной квадратуре круга. Однако же внутри нее мы находим разрешение затруднения Ферма (и не только его): первую связь между гиперболой Аполлония и логарифмами Непера. Книга VI посвящена всестороннему изучению гиперболы, о глубине и полноте которого мы можем судить по предложению 109 и соседним с ним.

Предложение 109. *Пусть AB и AC – асимптоты гиперболы DEF. Разобьем отрезок AC так, что AG, AH, AI, AK, AC находятся в непрерывной пропорции. Пусть DG, EH, LI, MK, FC эквидистантны*[1] *AB. Я утверждаю, что сегменты HD, IE, KL, CM равновелики.*

Так, конечно, говорил и Ферма с характерным для него пренебрежением к обоснованию своих утверждений, но Сен-Венсан дал целых два доказательства, и идею одного из них мы приведем ниже.

Сначала переформулируем результат в современных терминах: пусть имеется равнобочная гипербола и точки A, G, H, I, K, C на оси x выбраны так, что расстояния между ними образуют геометрическую прогрессию:

$$\frac{AG}{AH} = \frac{AH}{AI} = \frac{AI}{AK} = \frac{AK}{AC},$$

тогда соответствующие площади под гиперболой EHGD, LIHE, MKIL, FCKM равны.

Взгляните на рис. 4.4. Эта формулировка означает, что мы берем последовательность координат на оси x: a, ar, ar^2, ar^3, … . Эти точки определяют участки под гиперболой $xy = 1$ с площадями A_1, A_2, A_3, A_4, … , и мы утверждаем, что $A = A_1 = A_2 = A_3 = A_4 = \ldots$.

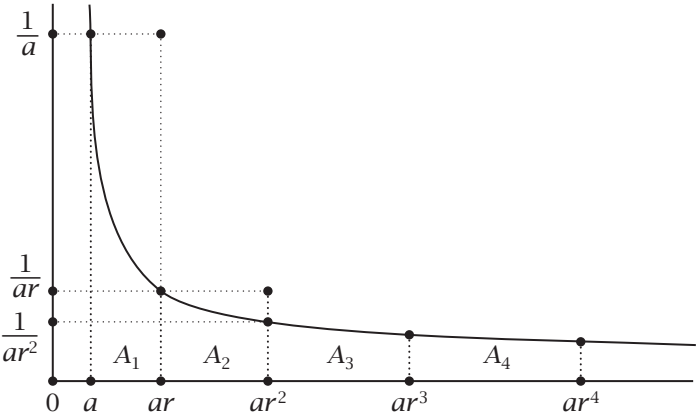


Рис. 4.4. Квадратура равнобочной гиперболы

[1] В этом контексте – параллельны.

Чтобы убедиться в этом, заметим, что каждая A_i заключена между площадями двух очевидных прямоугольников, так что

$$(ar - a)\frac{1}{ar} < A_1 < (ar - a)\frac{1}{a},$$
$$(ar^2 - ar)\frac{1}{ar^2} < A_2 < (ar^2 - ar)\frac{1}{ar},$$
$$\vdots$$

и потому

$$\frac{r-1}{r} < A_1 < r - 1, \quad \frac{r-1}{r} < A_2 < r - 1, \quad \ldots.$$

Поскольку r произвольно, площади должны быть равны. Таким образом, по крайней мере, на дискретном уровне, когда длины интервалов увеличиваются в геометрической прогрессии, площади под равнобочной гиперболой увеличиваются в арифметической прогрессии – а это и есть определяющее свойство логарифмов.

Имея в активе этот основополагающий результат, давайте измерим площадь от $a = 1$ и рассмотрим фиксированный интервал $[1, b]$, разбитый на части точками $x = 1, r, r^2, r^3, \ldots, r^n = b$. Из сказанного выше мы знаем, что каждая из равных частей площади удовлетворяет неравенству $(r - 1)/r < A < r - 1$, и значит, суммарная площадь удовлетворяет неравенству $n(r - 1)/r < \sum A < n(r - 1)$, где $r^n = b$, так что $r = b^{1/n}$. Это означает, что

$$n\frac{b^{1/n} - 1}{b^{1/n}} < \sum A < n(b^{1/n} - 1).$$

При $n \to \infty$ $b^{1/n} \to 1$, поэтому крайние части неравенства сходятся к одному и тому же пределу (при условии, что он существует), и мы получаем, что площадь равна

$$L(b) = \sum A = \lim_{n\to\infty} n(b^{1/n} - 1).$$

Конечно, тогда логарифмы еще не были функциями, этому результату предстояло дождаться Эйлера, к которому мы скоро вернемся.

Явное упоминание связи с логарифмами следует искать в последующей публикации другого автора, а еще до того у французского ученого-энциклопедиста отца Марена Мерсенна, игравшего роль главного организатора и репозитория тогдашних дебатов в значимых областях математики и других наук.

В публикации 1647 г. «Reflexiones Physico-mathematicae» (Размышления о физике и математике), которая, как явствует из названия, содержит его мысли о ряде вопросов физики и математики, имеется сдержанно критическая оценка квадратуры круга Сен-Венсана. Также в ней поставлена следующая задача:

> «Пусть даны три произвольные величины, рациональные или иррациональные, и известны логарифмы двух из них. Требуется геометрически найти логарифм третьей[1]».

[1] Здесь слово «геометрически» следует интерпретировать как «с математической точностью».

Проблема была актуальна на тот момент. В силу широкой известной связи между геометрической прогрессией чисел и арифметической прогрессией их логарифмов, знание логарифмов a и ar определяет систему в целом; следовательно, вопрос состоял в том, можно ли обнаружить третье число в этой геометрической прогрессии или какой-либо включающей ее. Здесь мы должны ясно понимать, что, хотя геометрическая прогрессия полностью определяется любыми своими двумя членами, ее можно рассматривать как составную часть бесконечного множества ассоциированных геометрических прогрессий; например, a, ar, ar^2, ar^3, … – главная прогрессия, но она также является частью более плотной прогрессии a, $a\sqrt{r}$, ar, $ar\sqrt{r}$, ar^2, $ar^2\sqrt{r}$, … . Искомое третье число может и не принадлежать исходной прогрессии, но являться членом прогрессии, включающей ее; это наблюдение было сделано еще одним интересным нам математиком-иезуитом, Альфонсо Антонио де Сараса. Сараса, фламандец испанского происхождения, был учеником Сен-Венсана, а в то время работал вместе с ним в Генте. В 1649 г. он опубликовал работу «Solutio Problematis a R P Marino Mersenno Minimo…», содержащую как защиту трудов своего старого наставника, так и ответ на вопрос Мерсенна. Оставляя в стороне его аргументы в защиту квадратуры круга, мы находим предложенное Сарасой решение проблемы логарифмов в следующем абзаце:

> «Из всего этого должно быть понятно, что если даны величины A и C и их логарифмы и также дана третья величина L, которая не может присутствовать ни в какой последовательности [непрерывной пропорции], содержащей величины A и C, как бы далеко ее ни продолжать и как бы ни делить или умножать ее (такое тоже возможно), то в этом случае невозможно найти логарифм величины L, и, следовательно, задача была некорректно сформулирована. Но, отвлекаясь от этого ограничения, мы можем найти то, что требуется в этой задаче, и свести задачу к геометрическому построению, применимому в допустимой ситуации».

А за доказательством, относящимся к этой задаче о логарифмах, он в первом предложении включенной в работу схолии отсылал к равнобочной гиперболе:

> «Но вы скажете, что мне не нужны эти отступления. И все же я приведу вас к логарифмам, какими бы далекими ни казались они от нашей цели. А затем кратко я объясню, как подходить к обучению логарифмам».

И в последнем:

> «По этой причине вы видите, что природа логарифмов с ее продолжением и большим числом членов находится в точном соответствии с гиперболой, поэтому вместо чисел вы можете брать части гипербол или заданное отношение отрезков».

И далее:

> «Откуда следует, что эти площади (под равнобочной гиперболой) могут занять место заданных логарифмов».

Логарифмы появились на свет в результате сравнения непрерывных движений двух точек, затем была установлена связь между дискретными ариф-

метическими и геометрическими прогрессиями, а с ее помощью был получен частичный ответ на трудную проблему квадратуры. После того как стало понятно, что логарифмы – это площади под равнобочной гиперболой, гиперболу стало возможно использовать для вычисления логарифмов.

4.3. Вычисление

Что бы ни представляли собой логарифмы, они нуждались в вычислении, и Бриггс при построении своей таблицы логарифмов применял несколько методов. Главным из них был отмеченный еще Непером факт: логарифм среднего геометрического двух чисел равен среднему арифметическому их логарифмов. То есть

$$L(\sqrt{ab}) = \tfrac{1}{2}(L(a) + L(b)).$$

Итак, используя уже установленные равенства $L(1) = 0$, $L(10) = 1$, можно показать, что $L(\sqrt{10}) = 0.5$, $L(\sqrt[4]{10}) = 0.25$, $L(\sqrt[8]{10}) = 0.125$, …, $L(\sqrt[2^{54}]{10}) = 2^{-54}$. Бриггс дошел до этого уровня разбиения, вычислив дробные степени 10 с поразительной точностью 32 знака после запятой, а логарифмы – с точностью до 14 знаков. Добавим к этому его наблюдение, что для малых α $L(1 + \alpha) \approx k\alpha$, где значение k он оценил как 0.434294481903251804 (это значение отличается от ln 10 только в двух последних знаках), – и вот мы имеем две вычислительные процедуры. Есть и другие, но мы надеемся, что читатель согласится с тем, что эту крайне необходимую таблицу чисел было трудно определить и не менее трудно построить. И тем не менее это удалось сделать в терминах квадратуры равнобочной гиперболы.

В десятом томе журнала «Transactions of the Royal Society» (1668) мы находим несколько важных статей, посвященных логарифмам и квадратуре гиперболы. Это подробный обзор в трактате Джеймса Грегори «Vera Circuli et Hyperbolae Quadratura» (Истинная квадратура круга и гиперболы), в котором он использовал метод исчерпывания для вычисления площади под равнобочной гиперболой, что привело к бесконечным рядам, сходящимся к логарифмам. Среди прочих упоминается статья, озаглавленная «Квадратура гиперболы с помощью бесконечных рядов рациональных чисел вкупе с ее демонстрацией именитым математиком достопочтенным лордом виконтом Браункером». Браункер, недавно вступивший в должность президента Королевского общества, наконец опубликовал результат, который, как напоминает нам преамбула, приписывался ему Джоном Валлисом несколькими годами ранее. Применив еще одно остроумное бесконечное исчерпывающее разбиение, он доказал, что площадь под стандартной равнобочной гиперболой на отрезке от 1 до 2 можно выразить в виде бесконечного ряда:

$$\begin{aligned}\tfrac{1}{1\times2} + \tfrac{1}{3\times4} + \tfrac{1}{5\times6} + \tfrac{1}{7\times8} + \cdots \\ &= (1 - \tfrac{1}{2}) + (\tfrac{1}{3} - \tfrac{1}{4}) + (\tfrac{1}{5} - \tfrac{1}{6}) + (\tfrac{1}{7} - \tfrac{1}{8}) + \cdots \\ &= 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \cdots .\end{aligned}$$

Поскольку площадь измеряется на отрезке от $a = 1$, вследствие чего логарифм 1 равен 0, этот результат говорит нам, что логарифм 2 равен 0.693147 ... . В обзоре Грегори упоминается также Николас Меркатор из Дании и его книга «Logarithmotechnia» (опубликованная на год раньше), первые две части которой были посвящены методу построения бригговых логарифмов. А в третьей части содержится обобщение результата Браункера, а его неуклюже сформулированное предложение 17 теперь выражено в виде знаменитого ряда:

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

Гипербола была перенесена на одну единицу влево, $1/(1+x)$, и это выражение было разложено в степенной ряд $1 - x + x^2 - x^3 + \ldots$; площадь под кривой для каждой степени x вычислялась методами Кавальери, результатом чего стала правая часть. Делая обзор этого результата, Джон Валлис улучшил обозначения и включил ограничение $x < 1$; он также установил, что (в современной нотации)

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots .$$

Наконец, в последней статье из того же номера «Transactions» Меркатор объединил оба результата и получил

$$\ln\frac{1+x}{1-x} = 2\left(x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots\right),$$

вычислив тем самым логарифмы 2, 3, 10 и 11. Кроме того, он умножил это на 0:43429 ($\sim 1 / \ln 10$), чтобы перейти от натуральных логарифмов к обыкновенным.

С помощью последнего ряда можно вычислить натуральный – а значит, и бриггов – логарифм любого числа, так как в его области сходимости, $-1 < x < 1$, величина $(1 + x)/(1 - x)$ принимает все положительные значения. Мы использовали функциональную нотацию, чтобы перенести аргументацию и результаты на почву, знакомую современному читателю, однако этот шаг еще только предстояло сделать, для чего нужно было осознать связь между логарифмом как степенью, в которую нужно возвести основание, и экспоненциальной функцией и обратной к ней.

4.4. Новые логарифмы

Человек, который осознал сущность логарифмов в их современном понимании, представляется неясной фигурой, а его имя осталось в истории только в связи с тривиальным и совершенно не связанным с логарифмами предметом: он ввел в обиход символ π для обозначения отношения длины окружности к диаметру. Этот эволюционный шаг был сделан в работе «Synopsis Palmariorum Mathesios», опубликованной в 1706 г. неким Уильямом Джонсом. Гораздо более важным и не столь широко известным является тот факт, что эта работа содержит обсуждение подхода Хэлли к логарифмам, а из более позднего трактата, который, вероятно, был написан спустя непродолжительное время (Джонс умер не позднее 1749 г.), мы можем сделать вывод, что автор был

одним из первых, кто понял связь между логарифмами и показателями степени, а также то, что за основание логарифмов можно принять любое число. Джонс был членом Королевского общества, поэтому спустя много лет после его смерти эта работа была подготовлена и зачитана обществу его библиотекарем, Джоном Робертсоном, 5 декабря 1771 г. Логарифмы вводились в этой статье без обиняков, и для наших целей достаточно двух вступительных абзацев:

> «1. Любое число можно выразить в виде некоторой однозначно определенной степени одного и того же радикала. Ибо всякое число находится где-то на шкале степеней радикала r с индексами $m - 1$, $m - 2$, $m - 3$, … , в которой можно выразить не только числа r^m, r^{m-1}, r^{m-2}, … , но и любое промежуточное число x представляется в виде r с подходящим показателем z. Показатель z называется Логарифмом числа x.
>
> 2. Поэтому, чтобы найти логарифм z любого числа x, нужно определить, какая степень радикала r на этой шкале равна числу x, или найти показатель степени z в уравнении $x = r^z$».

Мы пройдем мимо работ Ньютона (который, конечно же, медлил с их публикацией) и его великого соперника Лейбница. А также мимо работ Роджера Котса, нескольких членов семейства Бернулли и многих-многих других, оставивших вклады разной ценности, и сделаем последнюю остановку рядом с человеком, который оформил эти идеи в стройную теорию с хорошо нам знакомой системой обозначений.

Все дороги ведут в Рим, а все математические пути не минуют несравненного Леонарда Эйлера. Свободные концы, которые в совокупности формировали восприятие логарифмов, были связаны в тугой узел в первой книге его двухтомного учебника, вышедшего в 1748 г. под названием «Introductio in analysin infinitorum» (Введение в анализ бесконечных). Это именно учебник, а не исследовательская работа, и написан он примерно так же, как современные учебники (хотя и на латыни). И еще это шедевр проникновения в суть дела и мастерства изложения. Восемнадцать глав разбиты на 381 пронумерованный параграф, а к нашей теме имеют отношение главы VI и VII.

Во-первых, в параграфе 102 главы VI определяется точная природа логарифма[1]:

> «Как по любому значению z может быть найдено значение y, соответствующее данному числу a, так и обратно, можно найти значение переменного z, соответствующее любому заданному положительному значению переменного y так, чтобы $a^z = y$. Это значение переменного z, поскольку z рассматривается как функция y, обычно называется *логарифмом* переменного y. Итак, учение о логарифмах предполагает, что вместо a подставлено определенное постоянное число, которое поэтому носит название *основания* логарифмов; когда оно принято, то логарифмом любого числа y будет показатель степени a^z такой, что сама степень a^x будет равна числу y; логарифм числа y обычно обозначается через ly. Итак, если $a^z = y$, то $z = ly$. Отсюда понятно, что хотя основание логарифмов и зависит от нашего выбора, однако оно должно быть числом, большим, чем единица; отсюда можно получить в виде действительных чисел только логарифмы положительных чисел».

[1] Государственное издательство физико-математической литературы 1961. Перевод с лат. Е. Л. Пацановского, под ред. И. Б. Погребысского.

И далее параграф 103 начинается словами:

> «Какое бы число ни принять за основание a логарифмов, всегда будет $l1 = 0$; если в уравнении $a^z = y$, которое равносильно $z = ly$, положить $y = 1$, то будет $z = 0$».

Из этих наблюдений выводятся знакомые законы логарифмов, которые продемонстрированы на примере вычисления $l5 = \log_{10} 5$. И, как во всех хороших учебниках, приводятся возбуждающие интерес примеры, будь то вычисление $2^{7/12} = 1.498307\ldots$ или вычисление сложных процентов, что тогда было новшеством:

> «Пример 3
>
> После потопа род человеческий размножился от шести человек; положим, что 200 лет спустя число людей возросло до 1 000 000; требуется узнать, на какую свою часть число людей должно было увеличиваться ежегодно».

Предложенное им решение этой задачи выглядит так:

$$\begin{aligned} 6\left(1+\frac{1}{x}\right)^{200} = 10^6 &\to 1+\frac{1}{x} = \left(\frac{1000000}{6}\right)^{1/200} \\ &\to l\left(1+\frac{1}{x}\right) = \frac{1}{200} l\frac{1000000}{6} = 0.0261092 \\ &\to 1+\frac{1}{x} = 1.06196 \\ &\to x \approx 16. \end{aligned}$$

Эйлер продолжает:

> «Итак, для такого размножения людей достаточно было ежегодного увеличения на 1/16 часть; такое размножение не может считаться слишком большим ввиду того, что жизнь тогда была очень долголетней. Если бы увеличение числа людей продолжало идти далее в таком же отношении в течение промежутка в 400 лет, то тогда число людей должно было бы дойти до
>
> $$1000000 \times \frac{1000000}{6} = 166666666666,$$
>
> для прокормления такого числа не хватило бы всей земли».

Для такого вывода мы вычисляем:

$$\begin{aligned} \left(1+\frac{1}{x}\right)^{400} \times 6 &= \left[\left(1+\frac{1}{x}\right)^{200} \times 6\right] \times \left(1+\frac{1}{x}\right)^{200} \\ &= 1000000 \times \left(1+\frac{1}{x}\right)^{200} = 1000000 \times \frac{1000000}{6}. \end{aligned}$$

Перейдем к главе VII, которая начинается с параграфа 114 и приближения $a^x \sim 1 + kx$ экспоненциальной функции ее касательной в точке $x = 0$, где k, конечно, зависит от a. Отсюда в параграфе 116 Эйлер ткет математический гобелен и получает

$$a^z = 1 + \frac{kz}{1} + \frac{k^2z^2}{1\times 2} + \frac{k^3z^3}{1\times 2\times 3} + \frac{k^4z^4}{1\times 2\times 3\times 4} + \cdots$$

для любого значения z.

Переходя к параграфу 122, читаем:

> «Так как для построения системы логарифмов можно принять какое угодно основание a, то его подобрать так, чтобы было $k = 1$; тогда из найденного выше (§ 116) ряда
>
> $$1 + \frac{1}{1} + \frac{1}{1\times 2} + \frac{1}{1\times 2\times 3} + \frac{1}{1\times 2\times 3\times 4} + \cdots,$$
>
> если эти члены обратить в десятичные дроби и действительно сложить, то они дадут для a значение 2.71828182845904523536028, верное вплоть до последнего знака.
>
> Логарифмы, построенные при этом основании, обычно называются *натуральными* или *гиперболическими*, ибо квадратура гиперболы может быть выражена посредством логарифмов этого рода. Будем ради кратности вместо числа 2.718281828459... писать постоянно букву e, которая будет обозначать основание натуральных или гиперболических логарифмов, ему соответствует $k = 1$; эта буква e будет выражать также сумму ряда
>
> $$1 + \frac{1}{1} + \frac{1}{1\times 2} + \frac{1}{1\times 2\times 3} + \frac{1}{1\times 2\times 3\times 4} + \cdots».$$

Эйлер раньше уже использовал несколько букв вместо e, а сама e впервые появилась в письме Христиану Гольдбаху от 25 ноября 1731 г. Другие ученые тоже использовали альтернативные буквы, но влияние Эйлера было настолько сильным, что это обозначение вскоре стало стандартным. Как заметил немецкий математик Эдмунд Ландау (Landau 2001),

> «Букву e больше нельзя использовать ни для чего, кроме этой положительной универсальной постоянной».

Логарифмы стали функциями, а обратными к ним были экспоненциальные функции. Тем и другим были даны независимые определения:

$$\mathrm{e}^x = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n \quad \text{и} \quad \ln x = \lim_{n\to\infty} n(x^{1/n} - 1),$$

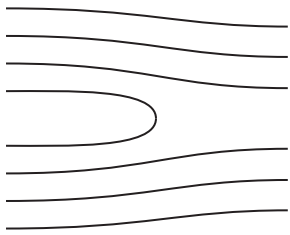
причем второй предел был предречен еще на стр. 66.

Бригговы логарифмы продолжали оказывать помощь в вычислениях вплоть до пришествия микросхем, а вслед за ними электронного калькулятора. Благодаря успешной квадратуре простой кривой гиперболический логарифм и его обращение заняли подобающее им почетное место в учебниках по основам анализа, где остаются с нами и по сей день. В последнем предложении главы Эйлер так подводит итог этой теме:

> «В остальном употребление гиперболических логарифмов будет более полно показано в интегральном исчислении».

Глава 5

Квадратриса Гиппия



Почему именно эта кривая?

Она «разрешила» две проблемы и породила новые. Изобретательное построение, выполненное умным человеком две с половиной тысячи лет назад, привело к кривой, которую тогда невозможно было оценить в полной мере, но которая послужила движущей силой для создания других похожих кривых и стала предшественницей тех, что мы теперь считаем сравнительно элементарными. Почти забытая кривая, напомнить о которой будет только справедливо.

5.1. Античные задачи

Древние вавилоняне и египтяне оставили грекам богатое наследие базовых геометрических идей. Но кладезь результатов в основном предназначался для практического применения землемерами, строителями и астрономами, а не для систематической дедуктивной науки, развитой греками. Действительно, когда мы, размышляя о математике, вспоминаем о Древней Греции, на ум, естественно, приходит чистая геометрия: геометрия Евклида, которой прежде всего посвящены его «Начала», хотя в них есть также важные результаты из теории чисел. Эти тринадцать книг являются самой знаменитой, самой влиятельной, самой переиздаваемой математической работой из когда-либо существовавших – за все время от их появления около 300 г. до н. э. и по сей день[1]. Но многие описанные в ней геометрические построения надлежало выполнять только циркулем и линейкой. Причем линейкой без делений, которые позволили бы сравнивать длины. Прямая и окружность были фундаменталь-

[1] Авраам Линкольн носил с собой и изучал экземпляр, содержавший первые шесть книг (Carpenter 1995).

ными геометрическими фигурами, а линейка и циркуль – их физическими воплощениями, с помощью которых – и только с их помощью – разрешалось создавать новые фигуры.

Несмотря на многочисленные успехи, достигнутые за много веков, три проблемы упрямо не поддавались решению – до такой степени, что в конце концов их разрешимость была поставлена под сомнение, хотя и это доказать также не удавалось.

Эти три великие античные задачи – перечисленные ниже – по праву прославились даже в новое время.

Трисекция угла. Поскольку любой угол можно разделить пополам с помощью циркуля и линейки (см. ниже), естественно поставить вопрос о делении его на три равные части (трисекции), тем более что для некоторых углов (90°, 72°, 27°…) возможность этого уже была установлена. Итак, вопрос: можно ли с помощью циркуля и линейки разделить произвольный угол на три равные части?

Квадратура круга. Представления древних греков о том, что такое площадь, по современным стандартам выглядят странно. Площадь – это не числовая величина, которая для некоторой фигуры измеряется в каких-то согласованных единицах, а связанное с ней свойство, которое следует сравнивать с самой фундаментальной фигурой: квадратом. Установить, что с некоторой фигурой связана такая же площадь, как с квадратом подходящего размера, значило выполнить квадратуру фигуры. Было хорошо известно, что квадратуру любой прямолинейной фигуры можно выполнить циркулем и линейкой; вопрос заключался в том, так ли это для самой фундаментальной из криволинейных фигур. Досократов философ Анаксагор из ионийского города Клазомены, живший примерно в 500–428 гг. до н. э., – первый, чье имя связывают с этой задачей. Он называл Солнце «раскаленным булыжником размером с весь Пелопоннес», а не «божественным Гелиосом», за что был заключен в тюрьму в 434 г. до н. э. по обвинению в нечестии. Сидя в тюрьме, он пытался построить квадратуру круга[1]. Много было последовавших его примеру. Не преуспел никто – и тому была основательнейшая из причин.

Удвоение куба. По преданию, Минос, критский царь и сын Зевса, был недоволен размером кубического надгробия жертвенника, воздвигнутого морскому божеству Главку[2]. Есть и другая, более известная легенда, согласно которой жители острова Делос, страдавшие от чумы, решили искать совета у дельфийского оракула. Оракул сообщил, что задобрить Аполлона можно, построив жертвенник святилища такой же формы, но вдвое большего объема, чем уже имевшийся, а тот жертвенник имел форму куба.

Кривая, являющаяся предметом настоящей главы, «решала» две из трех проблем, и мы скоро перейдем к ее определению, а затем и к ее роли в истории геометрии. Но сначала – отчасти в качестве предварительного шага, а отчасти

---

[1] Плутарх, «Об изгнании», глава 17.

[2] Бессмертному, хотя он был рожден человеком. Бессмертие он приобрел, поев волшебной травки.

ради чистого удовольствия – мы рассмотрим некоторые построения циркулем и линейкой. Отметим, что наш циркуль не «складывающийся», т. е., будучи раз раскрыт, он остается в таком положении, и угол между ножками не меняется, пока мы явно его не сложим.

5.2. Некоторые античные построения

Мы надеемся, что хотя бы некоторые читатели смахнут пыль с циркулей, хранящихся где-то в недрах ящика стола, и попробуют проделать описанное ниже, а быть может, даже разыщут и другие, более сложные построения. Все описанные построения имеют отношение к этой главе, а на некоторые из них мы будем ссылаться в последующих разделах (на те, что начинаются словами «построение *X*»).

Построение 1. Разделить заданный отрезок прямой пополам

Подготовка

Смотрите рис. 5.1. Проведите отрезок AB.

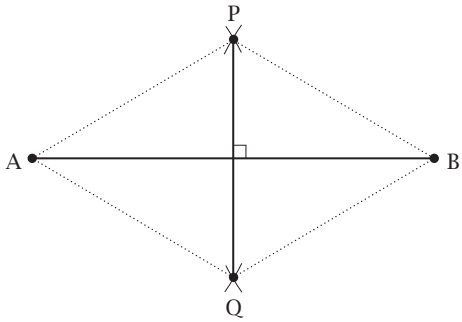


Рис. 5.1. Деление отрезка пополам

Метод

- Раскрыть циркуль на расстояние, большее половины AB.
- Поместить острие циркуля в точку A и описать две дуги, выше и ниже отрезка.
- Поместить острие циркуля в точку B и описать две дуги, выше и ниже отрезка, так что образуются две точки, P и Q.
- Соединить точки P и Q для получения искомого результата.

Объяснение

Фигура APBQ – ромб, одной из диагоналей которого является заданный отрезок, а другой – построенный отрезок. Диагонали ромба взаимно перпендикулярны.

Построение 2. Разделить заданный угол пополам

Подготовка

Смотрите рис. 5.2. Постройте подлежащий делению угол, проведя прямые AB и AC.

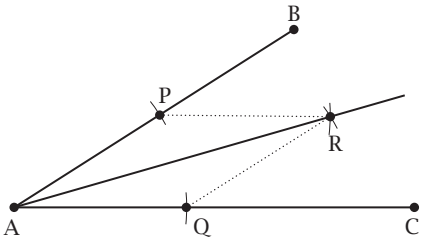


Рис. 5.2. Деление угла пополам

Метод

- Поместить острие циркуля в точку A, раскрыть его на удобное расстояние и провести дуги, определяющие точки P и Q.
- Оставить циркуль в этом положении или раскрыть его на удобное расстояние.
- Поместить острие циркуля в точку P и описать дугу.
- Поместить острие циркуля в точку Q и описать дугу. Пересечение двух дуг определяет точку R.
- Провести прямую AR, биссектрису угла A.

Объяснение

Треугольники APR и AQR конгруэнтны.

Построение 3. Опустить перпендикуляр к данной прямой из данной точки, не лежащей на ней

Подготовка

Смотрите рис. 5.3. Проведите прямую AB и выберите не принадлежащую ей точку P, из которой нужно будет опустить перпендикуляр.

Метод

- Раскрыть циркуль на удобное расстояние, поместить его острие в точку P и описать две дуги, пересекающие прямую в точках Q и R.
- Поместить острие циркуля в точку Q и описать дугу.
- Поместить острие циркуля в точку R и описать дугу. Пересечение двух дуг определяет точку S.
- Соединить точки S и P для получения искомой прямой.

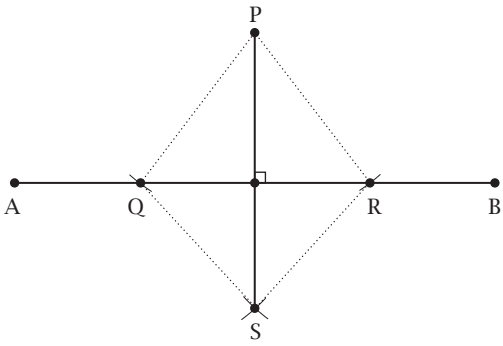


Рис. 5.3. Опущенный перпендикуляр

Объяснение

Четырехугольник PQRS – ромб, поэтому его диагонали взаимно перпендикулярны.

Построение 4. Восставить перпендикуляр к данной прямой из данной ее точки

Подготовка

Смотрите рис. 5.4. Проведите прямую AB и выберите на ней точку P.

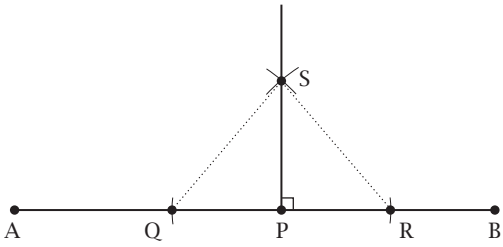


Рис. 5.4. Восставленный перпендикуляр

Метод

- Раскрыть циркуль на удобное расстояние, поместить его острие в точку P и описать дуги по обе стороны от точки P; они пересекают прямую в точках Q и R.
- Поместить острие циркуля в точку Q и описать дугу.
- Поместить острие циркуля в точку R и описать дугу. Пересечение двух дуг определяет точку S.
- Соединить точки S и P для получения искомой прямой.

Объяснение

Треугольники PQR и PRS конгруэнтны, а QPR – прямая линия.

Построение 5. Провести прямую через данную точку, параллельную данной прямой

Подготовка

Смотрите рис. 5.5. Проведите прямую AB и выберите точку P, через которую нужно провести параллельную прямую.

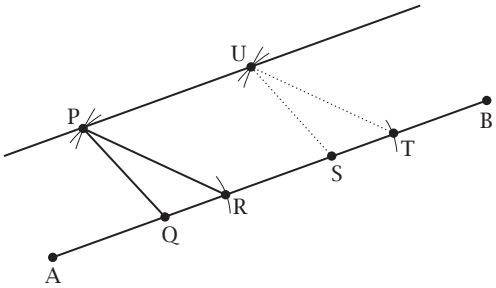


Рис. 5.5. Построение параллельной прямой

Метод

- Отметить на данной прямой две точки Q и R и соединить их с данной точкой P – получится треугольник PQR.
- Отметить произвольную точку S на данной прямой, отстоящую на удобное расстояние от R.
- Поместить острие циркуля в точку Q и раскрыть его ножки на расстояние QR.
- Поместить острие циркуля в точку S и отложить такое же расстояние, получив на прямой точку T.
- Поместить острие циркуля в точку Q и раскрыть его ножки на расстояние QP.
- Поместить острие циркуля в точку S и описать дугу.
- Поместить острие циркуля в точку R и раскрыть его ножки на расстояние RP.
- Поместить острие циркуля в точку T и отложить такое же расстояние, получив точку U.
- Провести прямую через U и P – это и будет параллельная прямая.

Объяснение

Два треугольника конгруэнтны.

Построение 6. Разделить данный отрезок на три равные части

Подготовка

Смотрите рис. 5.6. Проведите отрезок прямой AB.

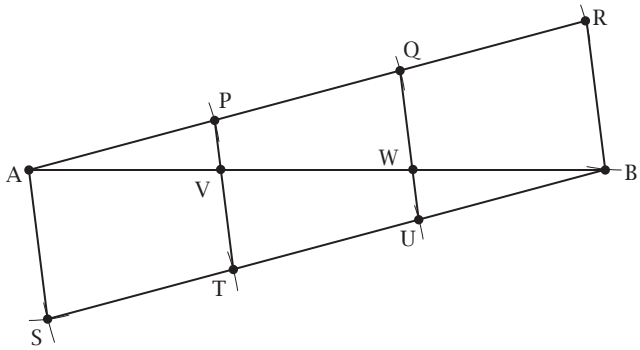


Рис. 5.6. Деление отрезка на три равные части

Метод

- Провести отрезок AR удобной длины под удобным углом к AB.
- Поместить острие циркуля в точку A, раскрыть его на удобное расстояние и описать дугу, которая пересечет AR в точке P.
- Поместить острие циркуля в точку P и, не меняя расстояние между ножками циркуля, описать дугу, которая пересечет AR в точке Q.
- Поместить острие циркуля в точку Q и, не меняя расстояние между ножками циркуля, описать дугу, которая пересечет AR в точке R.
- Поместить острие циркуля в точку R и раскрыть его на расстояние RB.
- Поместить острие циркуля в точку A и описать дугу ниже AB.
- Поместить острие циркуля в точку R и раскрыть его на расстояние RA.
- Поместить острие циркуля в точку B и описать дугу ниже AB для нахождения точки S, после чего провести прямую BS.
- Поместить острие циркуля в точку A и раскрыть его на расстояние AP.
- Поместить острие циркуля в точку S и описать дугу, пересекающую BS в точке T.
- Поместить острие циркуля в точку T и описать дугу, пересекающую BS в точке U.
- Провести отрезки PT и QU, они пересекут AB в точках V и W, делящих AB на три равные части.

Объяснение

- Прямые AR и SB параллельны.
- Прямые AS, PT, QU, RB параллельны.
- Треугольники AVP, AWQ, ABR подобны, а их стороны относятся как 1:2:3.

Это построение легко обобщить, чтобы разделить отрезок на любое число равных частей.

Построение 7. Разделить данный угол на три равные части

A-a-a... не получается.

Конечно, в общем случае это невозможно, но доказать этот факт удалось лишь в 1837 г. французу Пьеру Ванцелю.

Чтобы показать хрупкое равновесие частей, сошлемся на «Книгу лемм», в которой собрано 15 предложений, относящихся к окружностям. Ее оригинал известен на арабском языке, но в 1661 г. она переведена на латынь, затем сэр Томас Хит перевел ее на английский под названием «Труды Архимеда» (принято считать, что автором является Архимед, хотя это и не бесспорно). В предложении 8 показано, как произвольный угол можно разделить на три части с помощью линейки *с насечкой* – см. построение 8.

Построение 8. Архимедова трисекция угла

Подготовка

Смотрите рис. 5.7. Постройте угол, подлежащий трисекции, для чего проведите прямые AB и AC. Возьмите линейку, допускающую одну насечку.

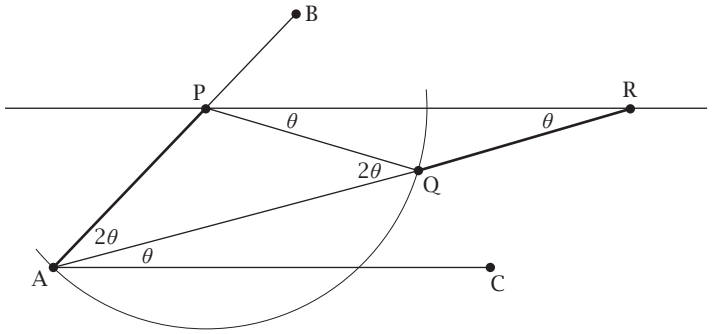


Рис. 5.7. Архимедова трисекция угла

Метод

- Выбрать какую-нибудь точку P на прямой AB.
- Построить прямую, проходящую через P параллельно AC (см. построение 5).
- Поместить острие циркуля в точку P и, раскрыв его на расстояние AP, описать большую дугу, как показано на рисунке.
- Расположить линейку вдоль AB, совместив один конец с точкой A, и отложить на ней расстояние AP, поставив насечку.
- Расположить линейку, так что она проходит через точку A, пересекает дугу окружности в точке Q, а параллельную прямую в точке R, где AP = QR (показаны жирными линиями на рисунке).
- Прямая AR высекает одну треть ∠BAC.

Объяснение

PQ = AP = QR, и две прямые параллельны. Равные углы показаны на рисунке.

Разумеется, разрешение насечки на линейке изменило правила, и теперь, подготовив почву, мы перейдем к следующему ослаблению правил с целью дать другое решение задачи – в котором участвует кривая, являющаяся предметом настоящей главы.

5.3. Квадратриса и трисекция

Вспоминая современные Олимпийские игры, мы, естественно, думаем о замечательных навыках, духе борьбы и атлетизме, демонстрируемом соревнующимися. А быть может, перед нашим мысленным взором встают церемонии открытия и закрытия Игр, которые соревнуются между собой так же ожесточенно, как атлеты. Предание возводит первые Игры к 776 г. до н. э., и в своей античной форме они проводились раз в четыре года до 393 г. н. э.; местом проведения была современная «периферийная единица» Элида на западном берегу полуострова Пелопоннес, в границах которого располагалась Олимпия. А Гиппий Элидский был, как следует из имени, уроженцем этой области. Большая часть известного о нем хоть сколько-нибудь достоверно взята из диалогов Платона, согласно которым он родился примерно в 460 г. до н. э. и умер примерно в 400 г. до н. э., т. е. жил в одно время с ним, и добрых чувств друг к другу они не испытывали: Платон характеризует Гиппия как тщеславного, надменного и не слишком умного человека[1]. Складывается впечатление, что Гиппий был богат, много путешествовал, обладал невероятной памятью и читал лекции по широкому кругу теоретических и практических предметов, включая математику. Что до его связи с Олимпийскими играми, то он регулярно присутствовал на них, первым составил список победителей и был замечен в публичном произнесении речей во время их проведения – эта практика в современные Игры не вошла. А в этой главе он появляется, потому что ему приписывается изобретение кривой, впоследствии названной квадратрисой, которая первоначально предназначалась для трисекции произвольного угла.

Путь Гиппия к успеху тоже пролегал через расширение доступного инструментария, только вместо линейки с насечкой он применил эту новую кривую: первую «механическую» кривую», которая подверглась изучению после линейки и циркуля и появилась лет на 60–70 раньше конических сечений.

Его определение было неочевидным – и противоречивым. Кривая является геометрическим местом точек пересечения двух движущихся прямых (см. рис. 5.8). Построим квадрат OXYZ со стороной 1, который заодно будет определять оси x и y будущей системы координат. В момент $t = 0$ сторона YZ начинает двигаться вертикально вниз с одной постоянной скоростью, а сторона OY начинает поворачиваться по часовой стрелке вокруг точки O с другой постоянной скоростью. Кривая является геометрическим местом всех точек пересечения P этих двух прямых. Очевидно, что она начинается в точке Y и идет вниз до пересечения с осью OX в некоторой (неизвестной) точке S.

Само построение было подвергнуто критике. Из комментария Паппа Александрийского мы знаем, что он и ученый Спор Никейский отмечали, что для синхронного движения, которое необходимо для построения кривой, нужно точно знать отношение стороны квадрата к длине вписанной в него части окружности, т. е. отношение радиуса окружности к четверти ее длины; это значит, что должно быть известно точное значение π и, стало быть, решение задачи о квадратуре круга. Тем не менее трисекция угла может быть произведена, как описано ниже.

---

[1] Смотрите, например, диалоги «Гиппий больший» и «Гиппий меньший».

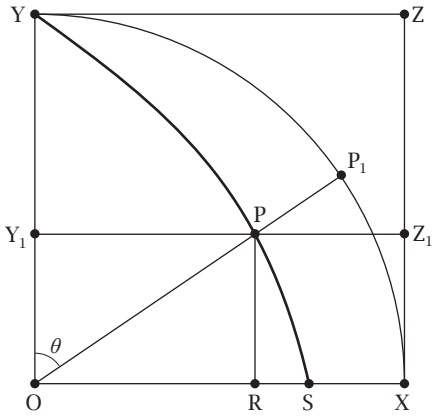


Рис. 5.8. Построение квадратрисы

Пусть требуется разделить на три равные части ∠POQ на рис. 5.9, где мы опустили четверть окружности. Пользуясь построениями из предыдущего раздела, поступим следующим образом:

- опустим перпендикуляр из точки P на прямую OQ, который пересечет ее в точке U (см. построение 3);
- разделим отрезок PU на три равные части, одной из которых является SU (см. построение 6);
- проведем горизонтальную прямую RST (см. построение 5);
- проведем отрезок OT;
- угол ∠TOQ = ⅓ ∠POQ.

Почему? Мы могли бы разобрать одно из ранних доказательств, но вместо этого перенесемся на много веков вперед, найдем уравнение кривой в декартовых координатах, а из него выведем этот результат и подготовим почву для следующего.

Вернемся к рис. 5.8, на котором изображен квадрат со стороной 1. Предположим, что вращающаяся прямая заняла положение OP_1, повернувшись на угол θ, а опускающаяся прямая в тот же момент заняла положение Y_1Z_1; тогда точка $P(x, y) = P(OP \sin\theta, OP \cos\theta)$ лежит на кривой. Так как оба движения равномерны, $\theta = \alpha t$ и $OY_1 = 1 - \beta t$, где α и β – некоторые постоянные. Параметрические уравнения кривой имеют вид:

$$\begin{aligned} y &= OY_1 = 1 - \beta t \\ x &= OP \sin\theta = OP \tan\theta \cos\theta = (OP \cos\theta)\tan\theta, \\ &= y \tan\theta = (1 - \beta t)\operatorname{tg} \alpha t. \end{aligned}$$

Таким образом,

$$\begin{aligned} x &= (1 - \beta t)\operatorname{tg}\alpha t, \\ y &= 1 - \beta t. \end{aligned}$$

И уравнение кривой может быть записано в виде $x = y \operatorname{tg} \alpha t$.

Чтобы исключить из уравнения t, заметим, что, поскольку обе прямые должны достичь основания квадрата одновременно в какой-то момент $t = T$, должна существовать связь между их скоростями, т. е. между α и β, и эту связь мы сейчас установим.

Имеем соотношения:

$$\theta = \alpha t \to \tfrac{1}{2}\pi = \alpha T,$$
$$\mathrm{OY}_1 = 0 \to 1 - \beta T = 0 \to \beta T = 1,$$

из которых следует связь между постоянными $\alpha = ½\ \pi\beta$.

Таким образом, $x = y \operatorname{tg} ½\ \pi\beta = y \operatorname{tg} ½\ \pi(1 - y) = y \operatorname{ctg} ½\ \pi y$ – уравнение нашей кривой в декартовых координатах. Оно включает простую тригонометрическую функцию и обобщается (почти) на все θ, а не только на значения $0 \leqslant \theta \leqslant \pi$, $\theta \neq ½\pi$, связанные с физическим построением, что и показано на рисунке в начале этой главы.

Зная это, мы легко можем понять, почему произведена трисекция. Ведь если $\angle\mathrm{POQ} = \gamma$ и $\angle\mathrm{TOQ} = \varphi$, а $\mathrm{P}(x, y)$, $\mathrm{T}(x_1, y_1)$ такие, как на рис. 5.9, то имеем

$$\operatorname{tg}\gamma = \frac{y}{x} = \operatorname{tg}\tfrac{1}{2}\pi y \quad \text{и} \quad \operatorname{tg}\varphi = \frac{y_1}{x_1} = \operatorname{tg}\tfrac{1}{2}\pi y_1.$$

Поэтому

$$\gamma = \tfrac{1}{2}\pi y \qquad \text{и} \qquad \varphi = \tfrac{1}{2}\pi y_1 = \tfrac{1}{2}\pi(\tfrac{1}{3}y) = \tfrac{1}{3}(\tfrac{1}{2}\pi y) = \tfrac{1}{3}\gamma.$$

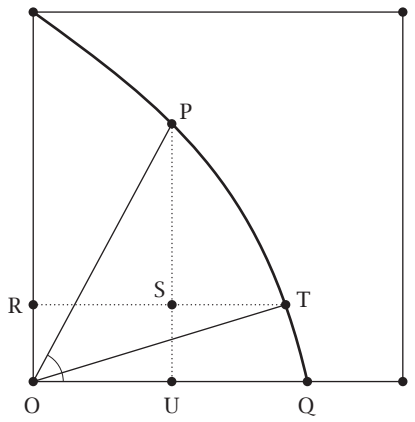


Рис. 5.9. Квадратриса и трисекция угла

Конечно, имея в виду обобщенное построение 6, мы можем разделить угол на любое число равных частей.

Квадратриса выполнила свое первоначальное предназначение, пусть и несколько спорно, которое состояло в ответе на вопрос, отличающийся по виду от двух других знаменитых античных проблем. Как мы уже упоминали, некоторые углы *можно* разделить на три равные части, например 90°, 72°, 27°..., а требовалось установить, что не существует никакого *общего* процесса, позволявшего сделать это только с помощью циркуля и линейки. С удвоением куба и ква-

дратурой круга утверждение звучит несколько иначе: никакой куб нельзя удвоить и никакой круг нельзя квадрировать. Но доказательства ни того, ни другого утверждения не было. И снова ослабление правил, как и в случае применения линейки с насечкой, в корне изменило ситуацию, а квадратриса опять оказалась в центре внимания, поскольку позволяла произвести квадратуру круга.

5.4. Квадратриса и квадратура круга

Наконец-то мы можем объяснить, почему кривая называется квадратрисой, хотя это название придумал не ее открыватель, Гиппий, а его современник, некий Гиппократ Хиосский, живший ок. 470–410 до н. э. Не следует путать его с гораздо более известным Гиппократом Косским, «отцом медицины», в честь которого названа «клятва Гиппократа». Наш Гиппократ родился на острове Хиос (по соседству с островом Самос, прославленным Пифагором), но, видимо, около двадцати лет жил в Афинах и за это время снискал репутацию отличного математика. Как мы уже отмечали, к тому времени было хорошо известно, что для любой прямолинейной фигуры можно построить равновеликий ей квадрат, но Гиппократ был первым, кому удалось построить квадратуру криволинейной фигуры – не круга, конечно, а *луночки*, точнее трех луночек.

Мы определим луночку как фигуру, образованную пересечением двух неравных окружностей; на рис. 5.10 она обведена полужирной линией.

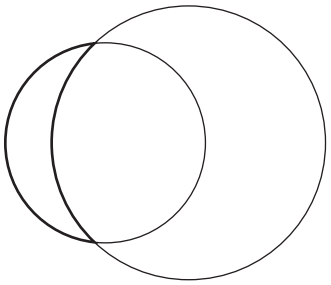


Рис. 5.10. Луночка

На рис. 5.11 показана оригинальная луночка, рассмотренная Гиппократом. Она образована пересечением четверти большей окружности радиуса 1 и половины меньшей окружности, опирающейся на хорду AC, как на диаметр. Эта фигура, называемая *прямоугольной равнобедренной треугольной луночкой*, обведена жирной линией и легко строится с помощью циркуля и линейки с помощью построений 1 и 4. Ее квадратура опирается на результат, который дошел до нас в виде предложения 2 из книги XII «Начал» Евклида:

> «Площади кругов относятся как квадраты их диаметров».

Зная это и обозначив D центр большей полуокружности, имеем:

- $AD = 1 - AB = 2, AC = \sqrt{2}$;
- $\dfrac{\text{площадь полукруга с диаметром AC}}{\text{площадь полукруга с диаметром AB}} = \left(\dfrac{\sqrt{2}}{2}\right)^2 = \dfrac{1}{2}$;

- площадь полукруга с диаметром AC = ½ площади полукруга с диаметром AB = площадь четверти круга, опирающегося на AD;
- площадь луночки + общая площадь = общая площадь + площадь треугольника ADC;
- площадь луночки = площадь треугольника ADC.

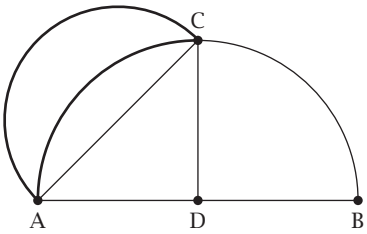


Рис. 5.11. Квадратура конкретной луночки

Эта конкретная луночка квадрируется. Гиппократу также удалось квадрировать еще две такие луночки: одна опирается на вписанную в полуокружность равнобедренную трапецию, у которой большая сторона в $\sqrt{3}$ раз длиннее равных меньших сторон, а другая – на неправильный, но симметричный пятиугольник. Подробности можно прочитать, например, в переводе Томаса Хита (Thomas Heath 1981, стр. 183–200).

Он сделал и еще несколько важных наблюдений, одно из которых дразняще близко к квадратуре круга и опирается на рис. 5.12, вариант которого – первая из двух упомянутых в предыдущем абзаце луночек. Здесь длина наибольшей стороны трапеции в два раза больше длины наименьшей стороны, а три равные луночки состоят из соответствующих полуокружностей и дуг окружностей. Левая часть рисунка – копия одной из меньших полуокружностей, а весь рисунок можно построить циркулем и линейкой. Гиппократ рассуждал так:

$$AD = 2AB \rightarrow AD^2 = 4AB^2 = AB^2 + AB^2 + BC^2 + CD^2.$$

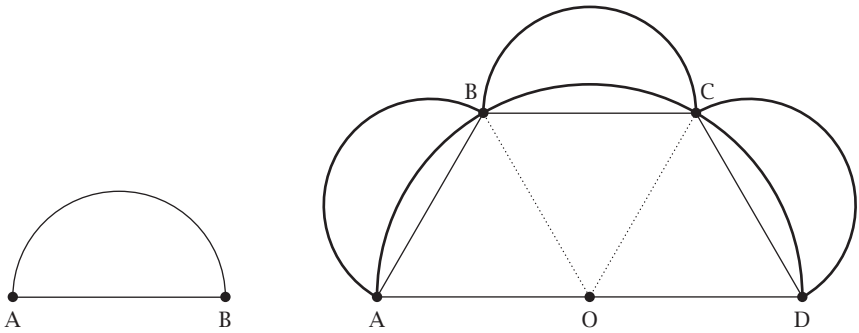


Рис. 5.12. Квадратура еще одной луночки

Поскольку площади кругов относятся как квадраты их диаметров, получаем, что

площадь большего полукруга

= площадь меньшего полукруга

+ сумма площадей меньших полукругов, опирающихся на стороны,

что можно переписать как

площадь трапеции + сумма площадей трех сегментов

= площадь меньшего полукруга

+ сумма площадей трех сегментов

+ сумма площадей трех луночек.

Отбрасывая одинаковые члены в обеих частях, получаем

площадь трапеции = площадь меньшего полукруга

+ сумма площадей трех луночек.

Если бы нам удалось квадрировать эти луночки, то мы смогли бы построить квадратуру полукруга, а значит, и круга. Но квадрировать их невозможно. Лишь в 1766 г. Мартин Юхан Валлениус открыл две другие квадрируемые луночки, а независимо от него в 1771 г.[1] их же открыл Эйлер. И только в 1934 г. русский математик Н. Г. Чеботарев подошел очень близко, а его ученик А. В. Дороднов (1947) довел до конца доказательство того, что три луночки Гиппократа и последующие две луночки – единственные допускающие квадратуру циркулем и линейкой, но ни одна не достигает конечной цели – квадрировать круг.

Какова же тогда роль квадратрисы в квадратуре круга? Чтобы ответить на этот вопрос, мы должны перенестись на сто лет после Гиппократа, когда греческий математик Динострат доказал теорему, носящую его имя. В обозначениях рис. 5.8 эта теорема утверждает, что сторона квадрата равна *среднему пропорциональному* (см. ниже) OS и дуги $\widehat{YX}$, т. е.

$$\frac{\widehat{YX}}{OX} = \frac{OX}{OS} \text{ и потому } \frac{\frac{1}{2}\pi}{1} = \frac{1}{OS},$$

откуда следует, что OS = 2/π. Доказательство, конечно, геометрическое и проводится от противного. Мы не станем здесь его повторять, а воспользуемся более современными методами, основанными на уравнении кривой. Нам нужно найти $\lim_{y\to 0} y \operatorname{ctg} \frac{1}{2}\pi y$, но выражение под знаком предела в точке $y = 0$, естественно, не определено, что не является серьезным препятствием для символьного калькулятора, поскольку можно применить правило Лопиталя:

$$\lim_{y\to 0} y \cot \tfrac{1}{2}\pi y = \lim_{y\to 0} \frac{y}{\tan \frac{1}{2}\pi y} = \lim_{y\to 0} \frac{1}{\frac{1}{2}\pi \sec^2 \frac{1}{2}\pi y} = \frac{2}{\pi}.$$

Квадратриса пересекает горизонтальную сторону в точке, отстоящей на расстояние 2/π от O, при условии, что она *вообще* пересекает эту сторону. Второе возражение Паппа заключалось в том, что из определения кривой вытекает,

[1] Доказательства – среди многих других вариантов – имеются в книге сэра Томаса Хита (Thomas Heath 1981, стр. 183–200).

что два отрезка прямой должны совпасть в конце движения, а это значит, что конечную точку S невозможно построить геометрически. Иначе говоря, если остановить движение в любой момент близко к окончанию, то невозможно точно экстраполировать остаток механической кривой, поскольку он еще не существует. Возражение справедливое, но мы его проигнорируем, потому что в итоге квадратриса приняла на себя основной груз квадратуры круга. «Построив» отрезок длины $2/\pi$, мы воспользуемся построением 1, чтобы разделить его пополам, и, получив $1/\pi$, применим построение 9 для построения обратной величины.

Построение 9. По заданному и единичному отрезку построить отрезок, длина которого обратна длине заданного

Подготовка

Смотрите рис. 5.13. Проведите известный отрезок AC и под любым удобным углом проведите отрезок AB единичной длины.

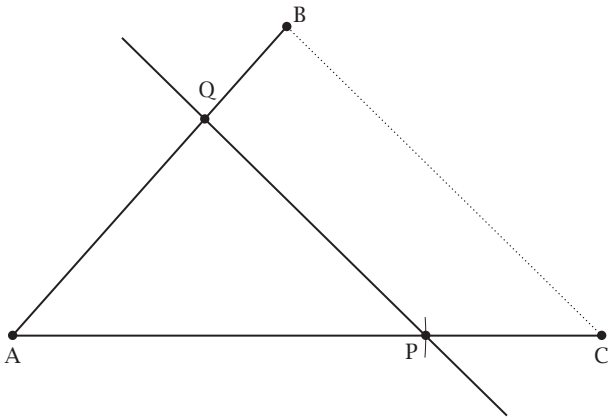


Рис. 5.13. Построение обратного отрезка

Метод

- Поместить острие циркуля в точку A и, раскрыв его на расстояние AB, описать дугу, пересекающую AC в точке P.
- Провести через P прямую, параллельную BC (построение 5), и пусть она пересекает AB в точке Q.
- Длина отрезка AQ обратна длине AC.

Объяснение

Треугольники AQP и ABC подобны, поэтому

$$\frac{AQ}{AB} = \frac{AP}{AC} \rightarrow \frac{AQ}{1} = \frac{1}{AC}.$$

Теперь мы построили π и для завершения работы воспользуемся построением 10.

Построение 10. По заданному и единичному отрезку построить отрезок, длина которого равна квадратному корню из длины заданного

Подготовка

Смотрите рис. 5.14. Проведите известный отрезок AP длины x и продолжите его вправо, так что длина PB равна 1.

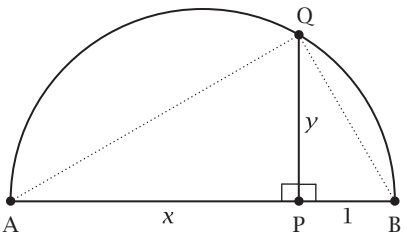


Рис. 5.14. Построение квадратного корня

Метод

- Поделив AB пополам, найти центр полуокружности, опирающейся на отрезок AB, как на диаметр (построение 1).
- Восставить перпендикуляр к AB из точки P, который пересекает полуокружность в точке Q (построение 4).
- Длина PQ равна $y = \sqrt{x}$.

Объяснение

- По теореме Пифагора, примененной к $\triangle$APQ, $AQ^2 = x^2 + y^2$.
- По теореме Пифагора, примененной к $\triangle$BPQ, $BQ^2 = y^2 + 1^2$.
- По теореме Пифагора, примененной к $\triangle$ABQ, $AB^2 = AQ^2 + BQ^2 \rightarrow (x + 1)^2 = (x^2 + y^2) + (y^2 + 1^2) \rightarrow y = \sqrt{x}$.

И таким образом мы построили сторону квадрата длиной $\sqrt{\pi}$, а значит, квадрировали круг.

Третья классическая задача на построение не решается с помощью квадратрисы и потому для нас не представляет особого интереса. Но Гиппократ показал, что куб можно удвоить, если удастся найти два *средних пропорциональных* числа и его же, умноженного на два. Эта величина уже упоминалась выше, но давным-давно была заменена средним геометрическим: чтобы поместить среднее пропорциональное x между числами a и b, нужно вычислить $x = \sqrt{ab}$. Метод построения среднего геометрического с помощью циркуля и линейки в точности совпадает с построением 10, где $AP = a$, $PB = b$, а y – искомое среднее геометрическое. Эту идею легко обобщить, если переписать среднее геометрическое в виде решения уравнения $a/x = x/b$, т. е. искать два средних пропор-

циональных, для которых $a/x = x/y = y/b$; здесь x – среднее геометрическое a и y, а y – среднее геометрическое x и b. Гиппократ пожелал заменить это соотношением $a/x = x/y = y/2a$, из которого следует $x^2 = ay$, $y^2 = 2ax$, так что $x = 2^{1/3}a$, и задача сводится к построению циркулем и линейкой отрезка длины $x = 2^{1/3}a$ по заданному отрезку длины a; тогда удалось бы удвоить куб. Как и в случае с трисекцией угла, мы должны обратить внимание на работу Пьера Ванцеля (тоже 1837 г.), который показал, что задача неразрешима, потому что отрезок длиной $\sqrt[3]{2}$ нельзя построить циркулем и линейкой.

Поэтому спустя некоторое время после квадратрисы появились конические сечения. Менехм, брат Динострата, который тоже пытался удвоить куб путем нахождения двух средних пропорциональных, пришел к первому представлению конических сечений в виде $a/x = x/y = y/b$, откуда следует, что $x^2 = ay$, $y^2 = bx$, $xy = ab$, и мы изучаем то, в чем узнаем пересечение параболы и гиперболы. Разумеется, сам он применял чисто геометрический подход.

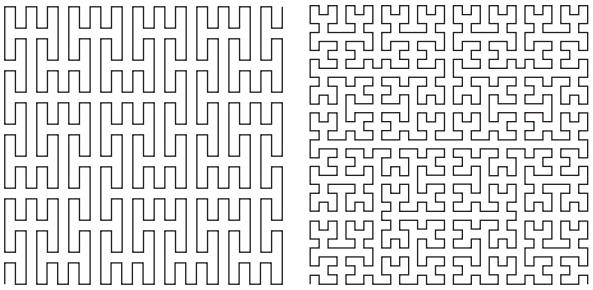
На этом наша история квадратрисы подошла к концу.

Мы упомянули конические сечения, а кроме них, есть еще много кривых, названия которых овеяны тайной: циссоида Диокла, конхоида де Слюза, конхоида Дюрера, конхоида Никомеда, декартов лист, кривая Евдокса, лемниската Бернулли, жемчужины Слюза, трисектриса Маклорена, трезубец Ньютона, локон Аньези и т. д.

Некоторые из них связаны с тремя классическими задачами на построение, существование других оправдано иными причинами. Но под какими бы именами они ни были известны, как бы ни использовались, свое происхождение математические кривые ведут от прямой, окружности и квадратрисы Гиппия.

Глава 6

Две кривые, заполняющие пространство



Почему именно эти кривые?

Первая кривая, описанная как функция, проложила путь второй – уже настоящей кривой. Кривая, заполняющая пространство, помогла привлечь внимание к опасностям, скрытым в первых двух определениях из книги I «Начал» Евклида, где объявляется, что «точка есть то, что не имеет частей», а «линия – длина без ширины». Срочному пересмотру были подвергнуты природа кривой и понятие размерности пространства, в котором кривая находится. Впоследствии это привело к возникновению новой обширной области топологии – теории топологической размерности.

6.1. Я вижу, но не верю этому

Конец XIX в. не был счастливой порой для математической интуиции. В главе 2 мы видели, что в 1872 г. Вейерштрасс предъявил всюду непрерывную, но нигде не дифференцируемую кривую. В 1873 г. немецкий математик, родившийся в России, Георг Кантор доказал, что рациональных и алгебраических чисел столько же, сколько натуральных, т. е. оба эти множества *счетны*; в том же году, но чуть позже он доказал, что множество вещественных чисел несчетно, а значит, почти все вещественные числа *трансцендентны*, с чем было особенно трудно смириться, потому что найти хотя бы одно трансцендентное число было отнюдь не легко[1]. Добавьте к этому противоречащую интуиции и страда-

[1] В 1844 г. Жозеф Лиувилль предъявил трансцендентное число, названное его именем, а в 1873 г. Шарль Эрмит доказал трансцендентность числа e. Трансцендентность π была доказана Фердинандом фон Линдеманом в 1882 г.

ющую внутренними противоречиями теорию трансфинитных чисел Кантора, опубликованную между 1879 и 1884 г., и вы получите точное представление о картине математических терзаний, сложившейся в то время. Нас здесь будет интересовать один конкретный инструмент из канторовой камеры математических пыток: тот, который ставит вопрос о том, что такое размерность.

> «Можно ли поставить некоторую поверхность (например, квадрат, включая границы) в однозначное отношение с кривой (например, с прямолинейным отрезком, содержащим концы) таким образом, что всякой точке поверхности соответствует точка кривой и, обратно, всякой точке кривой соответствует точка поверхности? Пока мне кажется, что ответ на этот вопрос представляет большие трудности, хотя и здесь настолько склонны давать отрицательный ответ, что доказательство считается почти излишним»[1].

Это цитата из письма Кантора своему другу Рихарду Дедекинду от 5 января 1874 г., а его тема направлена глубоко в сердце долго лелеемой интуиции – настолько глубоко, что сам Кантор чувствовал необходимость поставить под сомнение сформулированный им же вопрос. Негласное предположение о том, что наименьшее число координат, необходимых для определения положения точки в пространстве, определяет размерность пространства, казалось настолько очевидным, что даже обсуждать его не имеет смысла. Например, для задания точки на плоскости нужны две независимые координаты, поэтому плоскость двумерна, и представить все ее точки одним числом, принадлежащим числовой прямой, немыслимо. Дедекинд был для Кантора слушателем, на котором он опробовал свои идеи. Мнение Дедекинда, бывшего четырнадцатью годами старше и известнее, было для Кантора очень важно в силу уважения, которое он к нему испытывал, и он часто искал его одобрения в переписке, завязавшейся в 1872 г., когда Кантору было 27 лет, и продолжавшейся десять дет. В августе 1874 г. Кантор женился и свой медовый месяц провел в прелестном швейцарском городке Интерлакен. О многом говорит то, что Дедекинд проводил свой отпуск в то же время и в том же месте; известно, что они тогда проводили много времени вместе, обмениваясь математическими идеями, – хочется надеяться, что с одобрения новоявленной г-жи Кантор.

Несмотря на эти устные беседы, нам неизвестен письменный ответ Дедекинда на вопрос 1874 г., и мы перейдем к другому письму Кантора, датированному 20 июня 1877 г.:

> «Я хотел бы знать, считаете ли Вы арифметически строгим примененный мной метод доказательства? Речь идет о том, чтобы показать, что непрерывные поверхности, объемы и даже многообразия ρ измерений могут быть поставлены в однозначное соответствие с непрерывными кривыми, а значит и с многообразиями одного измерения, и что, следовательно, поверхности, объемы, многообразия ρ измерений имеют ту же самую мощность, что и кривые. Это мнение кажется противоположным общепринятому среди представителей новой геометрии, согласно которому говорят просто о дважды-, трижды- … ρ-кратно бесконечных многообразиях; иногда представляют вещи даже так, как если бы бесконечность точек у поверхности получалась в некотором роде возведением в квадрат, а у куба – в куб бесконечности точек у линии».

[1] Георг Кантор, «Труды по теории множеств». Под ред. А. Н. Колмогорова, А. П. Юшкевича. Москва. «Наука», 1985.

Кантор полагал, что ответил на свой вопрос. Процедуру вывода можно объяснить на примере применения к единичному квадрату и единичному интервалу, установив тем самым взаимно однозначное соответствие между ними; такое соответствие доказывает, что оба множества имеют одинаковую *мощность*, в терминологии Кантора. Сначала он избавился от неоднозначности, присущей конечному десятичному представлению рациональных чисел: например, он брал форму 0.1999 ... вместо 0.2[1]. И далее применил простое рассуждение: точке $(x, y) = (0.a_1a_2a_3 \ldots, 0.b_1b_2b3 \ldots)$ единичного квадрата ставилась в соответствие точка $z = 0.a_1b_1a_2b_2a_3b_3 \ldots$ единичного отрезка. Такое соответствие, очевидно, является взаимно однозначным, но, как сразу заметил Дедекинд, не является отображением *на*. Ответ, последовавший уже 22 июня 1877 г., содержал замеченную Дедекиндом трудность: в единичном отрезке существует бесконечно много чисел, недостижимых с помощью построения Кантора; в качестве примера он привел число $z = 0.120101010101\ldots$, которому должна была бы соответствовать точка с координатами $x = 0.100000 \ldots$ и $y = 0.21111 \ldots$, но такая координата x, по построению Кантора, недопустима. Дедекинд писал: «Не знаю, существенно ли мое возражение для Вашей идеи, однако я не захотел воздержаться от него», на что Кантор отвечал (почтовой открыткой на следующий же день): «К сожалению, Ваше возражение совершенно правильно; к счастью, оно относится только к доказательству, а не к самому факту. Действительно, я доказываю в некотором смысле больше, чем хотел доказать».

Отображение Кантора устанавливало взаимно однозначное соответствие между единичным квадратом и бесконечным подмножеством единичного отрезка, а не с самим единичным отрезком. Результат действительно выглядит более сильным, чем намеревался показать автор, – но такова природа бесконечного, что интуиция то и дело оказывается посрамленной. Однако Кантор все же хотел убедить Дедекинда в справедливости своего первоначального предположения о равномощности единичного квадрата и всего единичного отрезка, поэтому 25 июня 1877 г. он энергично продолжил переписку, приведя гораздо более сложное рассуждение, включавшее непрерывные дроби и бесконечные последовательности иррациональных чисел, сходящиеся к 1, а также графическую демонстрацию равномощности интервалов [0, 1] и [0, 1]. Доказательство достигло своей цели, но является удручающе громоздким и к тому же абсолютно излишним, потому что ранее описанное соответствие можно легко поправить следующим образом.

Если, например, $z = 0.12030045600007809 \ldots$, то разобьем число на «атомы» α_i – целые числа, не равные нулю, или последовательности, включающие целое число и все предшествующие ему нули (в данном случае $\alpha_1 = 1$, $\alpha_2 = 2$, $\alpha_3 = 03$, $\alpha_4 = 004$, $\alpha_5 = 5$, $\alpha_6 = 6$, $\alpha_7 = 00007$, $\alpha_8 = 8$, $\alpha_9 = 09$, ...), и поставим такому z в соответствие точку единичного квадрата с координатами $x = 0.\alpha_1\alpha_3\alpha_5 \ldots$ и $y = 0.\alpha_2\alpha_4\alpha_6 \ldots$. В данном случае $x = 0.10350000709 \ldots$ и $y = 0.200468 \ldots$. В примере, приведенном Дедекиндом, имеем $\alpha_1 = 1$, $\alpha_2 = 2$, $\alpha_3 = 01$, $\alpha_4 = 01$, $\alpha_5 = 01$, $\alpha_6 = 01$, $\alpha_7 = 01$, $\alpha_8 = 01$, $\alpha_9 = 01$. Этому числу будет поставлена в соответствие точка с координатами $x = 0.101010101 \ldots$ и $y = 0.201010101 \ldots$, и все хорошо. Обратное соответствие

[1] Изящное применение бесконечной геометрической прогрессии.

между единичным квадратом и единичным отрезком устанавливается очевидным образом и является не только однозначным, но и отображением *на*.

И это доказательство, и собственное доказательство Кантора без труда обобщаются на любую размерность, и в результате основания геометрии оказываются под угрозой. Дедекинд не ответил до 29 июня, поэтому Кантор отправил ему еще одно письмо, выдержка из которого приведена ниже.

> «Извините мое рвение в этом деле, если я слишком часто обращаюсь в Вашей доброте и Вашему труду. То, что я сообщил Вам совсем недавно, для самого меня столь неожиданно, столь ново, что я никак не могу успокоить мой ум, пока не получу, мой почитаемый друг, Вашего суждения об этом. Пока Вы не одобрите, я могу лишь сказать: *je le vois, mais je ne le crois pas*[1]. Именно поэтому прошу Вас прислать открытку, сообщив мне, когда Вы сможете закончить проверку, а если я могу рассчитывать на встречу с Вами, то внемлите моей мольбе, наверно, излишне настойчивой».

И здесь мы встречаем самое знаменитое из всех высказываний Кантора и, пожалуй, одно из самых знаменитых во всей математике: *je le vois, mais je ne le crois pas* – я вижу, но не верю этому[2]. Вся переписка, естественно, велась на немецком языке, но тут Кантор использовал цитату на французском по причинам, о которых можно только гадать. Он признал недостатки своего первого доказательства и не был полностью уверен в непогрешимости второго, поэтому нуждался в наметанном взгляде Дедекинда. Ответ Дедекинда от 2 июля начинается так: «Я еще раз проверил Ваше доказательство и не нашел в нем пробелов. Я убежден, что Ваша интересная теорема правильна, и поздравляю Вас с нею». Единичный квадрат действительно можно было сплющить в единичный отрезок, и не представлял больших трудностей следующий шаг – сплющить всю плоскость в прямую. Но последовавшая в результате ужасающая катавасия с размерностью (к счастью) не разрушила математическую структуру пространства; как и в случае многих парадоксов и противоречий, неудобное должно было привести к неизбежному, а в данном конкретном случае – к переоценке допущений и определений, лежащих в основе измерения пространства. Так появился целый новый раздел математики, который мы теперь называем *теорией топологической размерностии*. Примирение канторова соответствия между пространствами и идеей размерности лежит, как заметил сам Кантор и подчеркнул Дедекинд, в концепции непрерывности: по словам Дедекинда в ответном письме Кантору, «…ужасную, головокружительную размерность, вследствие которой все распалось на атомы, – такую разрывность, что всякая сколь угодно малая часть одной области оказалась разрывной, словно всюду разорванной в ее образе». Затем перед зрелищем кажущегося крушения идеи размерности Дедекинд предлагает решение:

---

[1] Я вижу, но не верю этому.

[2] Фраза «voir, c'est croire» переводится «видеть – значит верить». Это часть полной максимы «видеть – значит верить, но только чувства отражают истину» (seeing is believing, but feeling is the truth), принадлежащей английскому ученому и проповеднику XVII в. Томасу Фуллеру (которого не следует путать с Томасом Фуллером, мастером устного счета, жившим в XVIII в.).

> «Поэтому я пока убежден в справедливости следующей теоремы: “Если удалось установить взаимно однозначное и полное соответствие между точками непрерывного многообразия A, имеющего *a* измерений, с одной стороны и точками непрерывного многообразия B, имеющего *b* измерений, с другой, то, если *a* и *b* не равны, это соответствие необходимо *всюду разрывно*”».

Такое общее обратимое взаимно однозначное и полное (на) соответствие скоро было найдено Кантором (1878), когда он показал, что любые два конечномерных гладких многообразия любой размерности имеют одинаковое кардинальное число (или мощность). Тремя важными, существенными свойствами любого такого соответствия являются инъективность (взаимная однозначность), сюръективность (отображение на) и непрерывность, и, согласно, общему мнению Кантора и Дедекинда, оно не может обладать всеми тремя: соответствие Кантора и все ему подобные поэтому обязаны быть разрывными, но доказать это в общем случае никак не удавалось. В том же году четыре математика (с разной степенью убедительности и прибегая к очень сложным рассуждениям) показали, что для размерностей, не больших трех, непрерывное отображение между пространствами разной размерности не может быть взаимно однозначным, а в следующем году Кантор предложил свой вариант. Для больших размерностей орешек оказался покрепче; было представлено несколько ошибочных «доказательств», в том числе в 1899 г. самим Кантором – ошибку в нем удалось найти только двадцать лет спустя. И лишь в 1911 г. Л. Э. Я. Брауэр строго установил общий результат.

Итак, биективное отображение между единичным отрезком и единичным квадратом не может быть непрерывным, а значит, непрерывное соответствие не может быть биективным, и если оно сюръективно, то не может быть инъективным. Следующий очевидный шаг – предъявить пример сюръективного, но по необходимости не инъективного непрерывного соответствия между единичным отрезком и единичным квадратом. Правила игры определены, но сыграть в нее оказалось необычайно трудно – первое такое соответствие было найдено лишь через два десятилетия. В 1890 г. итальянский математик, логик и священник представил то, что требовалось, но весьма своеобразным способом.

6.2. Функция Пеано

Имя Джузеппе Пеано чаще всего связывают с *аксиомами Пеано*, лежащими в основе арифметики натуральных чисел (хотя он сам приписывал их Дедекинду). Но его вклад в математику шире и включает различные результаты из анализа, предъявление контрпримеров и обозначения (именно он ввел символ ∈ в теорию множеств и зеркальный ему символ ∋ для фразы «такой, что»). И он же впервые привел пример вожделенной непрерывной функции, которая сюръективно, но по необходимости не инъективно отображает единичный отрезок в единичный квадрат (Peano 1890). Этому примеру мы и посвятим следующие несколько страниц. Но, признавая его дар нахождения впечатляющих и зачастую неожиданных примеров и контрпримеров, нам очень трудно представить, как он додумался до того конкретного построения, детали которого рассматриваются ниже.

Построение Пеано

Определим оператор k как $kt = 2 - t$ для $t \in \{0, 1, 2\}$, а его n-ю итерацию обозначим k^n. В частности, $k^2t = k(kt) = k(2 - t) = 2 - (2 - t) = t$, т. е. $k^{2n}t = t$ и $k^{2n-1}t = kt = 2 - t$, а значит, если $k^nt = x$, то $t = k^nx$.

Теперь представим числа $t \in I$ в троичной системе счисления, так что $t = 0_{\dot 3}t_1t_2t_3t_4t_5t_6 \ldots$, где все $t_i \in \{0, 1, 2\}$.

Функцию $f : I \to S$ определим следующим образом:

$$\begin{aligned} f(t) &= (0_{\dot 3}t_1(k^{t_2}t_3)(k^{t_2+t_4}t_5)(k^{t_2+t_4+t_6}t_7)\ldots, \\ &\qquad 0_{\dot 3}(k^{t_1}t_2)(k^{t_1+t_3}t_4)(k^{t_1+t_3+t_5}t_6)\ldots) \\ &= (f_x(t), f_y(t)). \end{aligned}$$

Поясним это построение на примерах.

Во-первых, вычислим образ взятой наугад точки:

$$t = 0_{\dot 3}1102100212 \ldots = 0.4736405 \ldots,$$

где $t_1 = 1, t_2 = 1, t_3 = 0, t_4 = 2, t_5 = 1, t_6 = 0, t_7 = 0, t_8 = 2, t_9 = 1, t_{10} = 2, \ldots$, как

$$\begin{aligned} f_x(t) &= 0_{\dot 3}1(k^10)(k^31)(k^30)(k^51)\cdot\cdot\cdot = 0_{\dot 3}1(k0)(k1)(k0)(k1) \\ &= 0_{\dot 3}12121\cdot\cdot\cdot = 0.62139\ldots, \\ f_y(t) &= 0_{\dot 3}(k^11)(k^12)(k^22)(k^32) = 0_{\dot 3}(k1)(k2)(2)(k2) \\ &= 0_{\dot 3}1020\cdot\cdot\cdot = 0.4074\ldots. \end{aligned}$$

В десятичной системе это соответствие имеет вид:

$$0.4736405 \ldots \to (0.62139 \ldots, 0.4074 \ldots).$$

А во-вторых – образ вполне определенной точки $0_{\dot 3}1000 \ldots = ⅓$. Здесь мы имеем $t_1 = 1$ и $t_i = 0$ для прочих i, а значит,

$$\begin{aligned} f_x(t) &= 0_{\dot 3}1(k^00)(k^00)(k^00)(k^00)\cdot\cdot\cdot = 0_{\dot 3}10000\cdot\cdot\cdot = \tfrac{1}{3}, \\ f_y(t) &= 0_{\dot 3}(k^10)(k^10)(k^10)(k^10) = 0_{\dot 3}2222\cdot\cdot\cdot = 1. \end{aligned}$$

В десятичном виде соответствие имеет вид ⅓ → (⅓, 1). Но так ли мы уверены? Вспомним, что при построении оригинального соответствия Кантор брал бесконечное десятичное представление конечного десятичного числа; в данном случае мы можем записать $⅓ = 0_{\dot 3}0222 \ldots$ в троичном виде. По счастью, все хорошо, потому что применение функции к этой форме также дает ($0_{\dot 3}0222 \ldots$, $0_{\dot 3}2222 \ldots$) = (⅓, 1). Читатель может проверить, что так обстоит дело всегда, поэтому неоднозначность представления не влияет на определение функции, т. е. она определена корректно.

Итак, мы имеем корректно определенную функцию, причем в явной (пусть и несколько экзотической) аналитической форме. Теперь нужно показать, что она обладает обоими требуемыми свойствами – и по необходимости не обладает третьим.

Сюръективность

Для любой точки

$$s = (0_3 x_1 x_2 x_3 x_4 x_5 x_6 \ldots; 0_3 y_1 y_2 y_3 y_4 y_5 y_6 \ldots) \in S$$

мы должны предъявить такое $t = 0_3 t_1 t_2 t_3 t_4 t_5 t_6 \ldots \in I$, что $f(t) = s$. Для этого потребуем, чтобы

$$\begin{aligned}(0_3 t_1 (k^{t_2} t_3)(k^{t_2+t_4} t_5)(k^{t_2+t_4+t_6} t_7) \ldots,\\ 0_3 (k^{t_1} t_2)(k^{t_1+t_3} t_4)(k^{t_1+t_3+t_5} t_6) \ldots)\\ = (0_3 x_1 x_2 x_3 x_4 x_5 x_6 \ldots, 0_3 y_1 y_2 y_4 y_4 y_5 y_6 \ldots).\end{aligned}$$

Из сравнения координат получаем:

$$\begin{aligned} t_1 &= x_1,\\ k^{t_1} t_2 &= y_1 \to t_2 = k^{t_1} y,\\ k^{t_2} t_3 &= x_2 \to t_3 = k^{t_2} x_2,\\ k^{t_1+t_3} t_4 &= y_2 \to t_4 = k^{t_1+t_3} y_2,\\ k^{t_2+t_4+t_6} t_7 &= x_3 \to t_7 = k^{t_2+t_4+t_6} x_3,\\ k^{t_1+t_3+t_5} t_6 &= y_3 \to t_6 = y_3 k^{t_1+t_3+t_5},\\ &\vdots \end{aligned}$$

И таким образом, имеем троичное разложение t.

Непрерывность

Мы покажем, что функция Пеано непрерывна, доказав, что ее компонента f_x непрерывна в любой точке – и точно так же (с соответствующими изменениями) для компоненты f_y. Впрочем, результат для компоненты f_y можно было бы получить по-другому, заметив, что $f_y(t) = 3f_x(\frac{1}{3}t)$.

Итак, запишем произвольную точку в виде

$$a = 0_3 a_1 a_2 a_3 \ldots a_n a_{n+1} a_{n+2} a_{n+3} \ldots$$

и положим

$$b = 0_3 a_1 a_2 a_3 \ldots a_n \beta_{n+1} \beta_{n+2} \beta_{n+3} \ldots$$

– число, близкое к a, совпадающее с ним в первых n цифрах троичного представления, где n предполагается четным. Тогда ясно, что $|a - b| < 1/3^{n+1}$. Применим f_x к a и b и рассмотрим модуль разности между обоими образами, обращая внимание на первую позицию, в которой они различаются. Для этого преобразуем троичные представления образов в бесконечный ряд дробей.

Напомним, что

$$\begin{aligned} f_x(a) = 0_3 a_1 (k^{a_2} a_3)(k^{a_2+a_4} a_5)(k^{a_2+a_4+a_6} a_7) \cdots\\ \cdots (k^{a_2+a_4+a_6+\cdots+a_n} a_{n+1})(k^{a_2+a_4+a_6+\cdots+a_n+a_{n+2}} a_{n+3}) \cdots \end{aligned}$$

и что

$$f_x(b) = 0_{\dot{3}}a_1(k^{a_2}a_3)(k^{a_2+a_4}a_5)(k^{a_2+a_4+a_6}a_7)\cdots \\ \cdots(k^{a_2+a_4+a_6+\cdots+a_n}\beta_{n+1})(k^{a_2+a_4+a_6+\cdots+a_n+\beta_{n+2}}\beta_{n+3})\cdots.$$

Преобразуем эти выражения в соответствующие дроби и вычтем одно из другого:

$$\begin{aligned}
&|f_x(a) - f_x(b)| \\
&\quad = \left| \frac{k^{a_2+a_4+a_6+\cdots+a_n}a_{n+1} - k^{a_2+a_4+a_6+\cdots+a_n}\beta_{n+1}}{3^{(n/2)+1}} \right. \\
&\qquad \left. + \frac{k^{a_2+a_4+a_6+\cdots+a_n+a_{n+2}}a_{n+3} - k^{a_2+a_4+a_6+\cdots+a_n+\beta_{n+2}}\beta_{n+3}}{3^{(n/2)+2}} + \cdots \right| \\
&\quad \leqslant \left| \frac{k^{a_2+a_4+a_6+a_n}a_{n+1} - k^{a_2+a_4+a_6+\cdots+a_n}\beta_{n+1}}{3^{(n/2)+1}} \right| \\
&\qquad + \left| \frac{k^{a_2+a_4+a_6+\cdots+a_n+a_{n+2}}a_{n+3} - k^{a_2+a_4+a_6+\cdots+a_n+\beta_{n+2}}\beta_{n+3}}{3^{(n/2)+2}} \right| + \cdots \\
&\quad \leqslant \frac{2}{3^{(n/2)+1}} + \frac{2}{3^{(n/2)+2}} + \frac{2}{3^{(n/2)+3}} + \cdots \\
&\quad = \frac{2}{3^{(n/2)+1}}\left(1 + \frac{1}{3} + \frac{1}{3^2} + \cdots\right) \\
&\quad = \frac{2}{3^{(n/2)+1}} \times \frac{3}{2} = \frac{1}{3^{n/2}},
\end{aligned}$$

поскольку абсолютная величина разности между любой парой троичных цифр не превосходит 2. А так как $b \to a$, то $f_x(b) \to f_x(a)$, и функция f_x действительно непрерывна.

Таким образом, функция Пеано сюръективна и непрерывна. Значит, она не может быть инъективной, и это легко доказать.

Таблица 6.1. Возможные числители

a_i	0	1	2
β_i	1	0	1
$k^{\text{even}}a, k^{\text{even}}\beta$	0,1	1,0	2,1
$k^{\text{odd}}a, k^{\text{odd}}\beta$	2,1	1,2	0,1

Неинъективность

Предположим, что $f(s) = f(t)$, где $s = 0_{\dot{3}}s_1s_2s_3\ldots$, $t = 0_{\dot{3}}t_1t_2t_3\ldots$; тогда

$$0_{\dot{3}}s_1(k^{s_2}s_3)(k^{s_2+s_4}s_5)(k^{s_2+s_4+s_6}s_7)\cdots \\ = 0_{\dot{3}}t_1(k^{t_2}t_3)(k^{t_2+t_4}t_5)(k^{t_2+t_4+t_6}t_7)\cdots.$$

Следовательно, $s_1 = t_1$, но, например, $k^{s_2}s_3 = k^{t_2}t_3 \to s_3 = k^{s_2+t_2}t_3$, а это означает, что они могут быть различны, и, стало быть, одна и та же точка может являться образом более чем одной точки.

Недифференцируемость

Последняя строка работы Пеано содержит утверждение без доказательства: обе компонентные функции, будучи всюду непрерывными, нигде не дифференцируемы. Нам, наверное, никогда не узнать, как он убедился в этом факте, а математическому миру пришлось ждать еще десять лет, прежде чем в 1900 г. появилось первое доказательство Э. Г. Мура, который подошел к проблеме с новаторских позиций; но детали мы отложим до следующего раздела. А сейчас приведем более прямое, но столь же элегантное доказательство (Sagan 1994). Чтобы установить недифференцируемость, нам нужно только для любой точки $a \in I$ построить последовательность, которая стремится к a, но для которой производная не имеет предела (как мы делали для кривой Вейерштрасса в главе 2). Для этого возьмем фиксированную точку $a = 0_3 a_1 a_2 a_3 \ldots a_{2n} a_{2n+1} a_{2n+2} \ldots \in I$ и переменную точку $b = 0_3 a_1 a_2 a_3 \ldots a_{2n} \beta_{2n+1} a_{2n+2} \ldots \in I$, где $\beta_{2n+1} = (a_{2n+1} + 1) \bmod 2$. Эти две точки различаются только в $(2n + 1)$-й позиции, так что $|a - b| = 1/3^{2n+1}$ и $b \to a$ при $n \to \infty$. Снова рассмотрим модуль разности:

$$|f_x(a) - f_x(b)| = \frac{|k^{a_2+a_4+a_6+\cdots+a_{2n}} a_{2n+1} - k^{a_2+a_4+a_6+\cdots+a_{2n}} \beta_{2n+1}|}{3^{n+1}} = \frac{1}{3^{n+1}}.$$

Последнее равенство справедливо в силу табл. 6.1, из которой видно, что числитель всегда равен 1.

Следовательно,

$$\left| \frac{f_x(a) - f_x(b)}{a - b} \right| = \frac{1}{3^{n+1}} \times 3^{2n+1} = 3^n \xrightarrow[n\to\infty]{} \infty,$$

и мы получили желаемый результат.

6.3. Кривая Гильберта

Мы уже дошли до третьего раздела главы книги, посвященной кривым, а слово «кривая» так и не прозвучало. В статье Пеано нет и намека на кривые, она содержит чисто аналитические рассуждения, которые мы привели выше. Однако же работу прочитал немецкий принц математики, Давид Гильберт. Мы даже не пытаемся дать краткий очерк человека, математические способности которого Г. Х. Харди оценивал на 80 из 100 по своей условной шкале, отводя себе скромные 25 из 100[1]. Гильберт проник глубоко в детали работы

[1] Своему давнему коллеге Дж. Э. Литтлвуду он давал 30 из 100, а гениальному Раманудджану, с которым они так плодотворно работали, 100 из 100.

Пеано и в 1891 г. раскрыл ее геометрическую природу, предложив свой собственный вариант, определенный исключительно в геометрических терминах и получивший название *кривой Гильберта*; это первая из известных кривых, заполняющих пространство. На самом деле он начал двухстраничную статью (Hilbert 1891), в которой противопоставил свои методы методам Пеано, с построения этой кривой, а точнее трех ее первых итераций. На рис. 6.1 воспроизведен рисунок из этой статьи, а наша цель – переформулировать ее в современной форме.

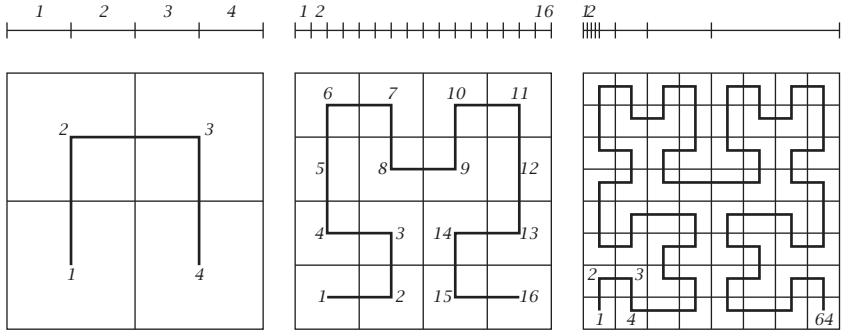


Fig. 1 Fig. 2 Fig. 3

Рис. 6.1. Оригинальные кривые Гильберта

Методику Гильберта можно описать следующим образом. Предположим, что единичный отрезок *I можно* отобразить на единичный квадрат *S*, тогда разбиение *I* на четыре конгруэнтных меньших отрезка и *S* на четыре конгруэнтных меньших квадрата индуцирует рекурсию такого отображения, и более того, соседние подотрезки будут отображаться на соседние подквадраты способом, гарантирующим непрерывность. Этот процесс можно повторять бесконечно, и в пределе непрерывная кривая заполнит весь квадрат, соединив точку (0, 0) с (1, 0). В процессе построения мы соединяем центры уменьшающихся подквадратов отрезками прямых уменьшающейся длины, поддерживая на каждом шаге непрерывность. Это требование непрерывности однозначно определяет начатый процесс. Мы последуем за Гильбертом и подробно рассмотрим первые три этапа распространения кривой, надеясь, что после этого принцип станет ясен.

Этап 1

Разобьем квадрат на четыре конгруэнтных подквадрата, найдем их центры и соединим в порядке, показанном на рис. 6.2a. Эта первая итерация кривой показана на рис. 6.2b.

Мы видим простой путь, соединяющий центры четырех подквадратов, который начинается и заканчивается, как нам нужно. Теперь усложним его, придвинув крайние точки ближе к нижним вершинам.

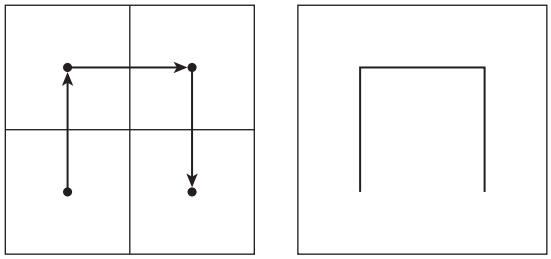


Рис. 6.2. Первый этап построения кривой Гильберта

Этап 2

Разобьем левый нижний подквадрат на четыре конгруэнтных квадрата, найдем их центры и, начиная с левого нижнего подквадрата (т. е. сместив центр ближе к точке (0, 0)), соединим эти центры единственно возможным способом, при котором каждый центр посещается только один раз и путь покидает подквадрат через верхнюю сторону, гарантируя непрерывность (рис. 6.3a). Повторим это построение для трех других подквадратов по очереди, начиная в единственно возможной точке и сохраняя непрерывность. Результат показан на рис. 6.3b, который освобожден от всего лишнего на рис. 6.3c. Это второй этап кривой Гильберта, а на рис. 6.3d показан этот этап и наложенный на него предыдущий (пунктиром).

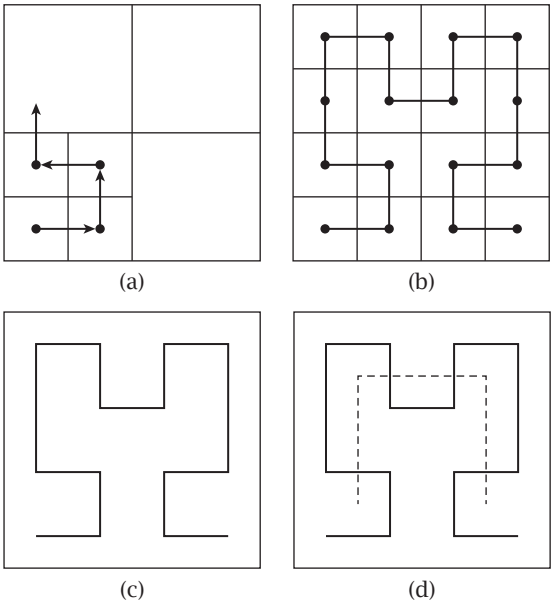


Рис. 6.3. Второй этап кривой Гильберта

Этап 3

Левый нижний подквадрат разбивается на четыре конгруэнтных подквадрата, находятся их центры, и путь, начинающийся в левом нижнем центре, идет в правый из новых подквадратов, затем вверх и влево, как показано на рис. 6.4a. Продолжая, мы получим третий этап кривой Гильберта на рис. 6.4b, а на рис. 6.4c на него наложен предыдущий этап.

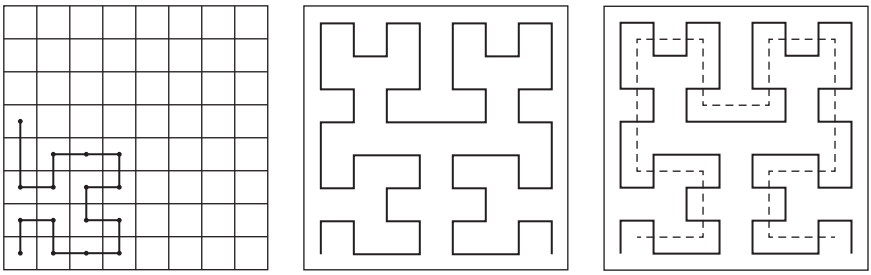


Рис. 6.4. Третий этап кривой Гильберта

Мы надеемся, что процесс теперь понятен, так что части a–c рис. 6.5, на которых изображены следующие три этапа построения кривой, не вызывают удивления.

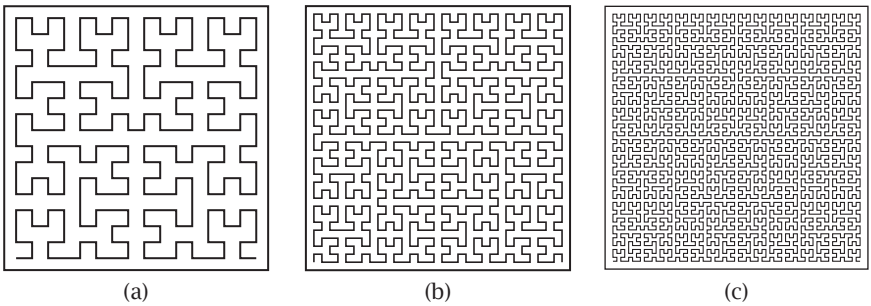


Рис. 6.5. Последующие этапы кривой Гильберта

Итак, мы имеем чисто геометрический процесс, порождающий все более и более плотно упакованную кривую, которая постепенно приближается к границам исходного квадрата. Но где же функция, которая отображает единичный интервал и этот единичный квадрат? Ниже приведено объяснение самого Гильберта (в переводе и с отсылками к его оригинальным рисункам):

> «Возьмем прямолинейный отрезок длины 1 и разобьем его на 4 части равной длины. Затем возьмем единичный квадрат и разобьем его двумя перпендикулярными прямыми на 4 равных квадранта, обозначенных 1, 2, 3, 4 (рис. 1). Далее разобьем каждую часть отрезка на четыре равные части, что даст нам 16 частей; одновременно разобьем каждый квадрант на четыре равных квадранта и запишем в 16 получившихся квадрантов числа 1, 2, 3, … , 16, где порядок квадрантов выбран так, что каждый имеет

общую сторону со своим предшественником (рис. 2). Если продолжить этот процесс дальше, – на рис. 3 изображен следующий шаг, – то, как легко видеть, каждой точке отрезка будет сопоставлена единственная точка в квадранте. Все что нужно – это отмечать часть отрезка, содержащую эту точку. Квадранты с одинаковыми номерами по необходимости вложены друг в друга и в пределе сходятся к точке единичного квадрата».

Фразеология Гильберта требует от читателя больше веры, чем понимания, и за разъяснениями мы обратимся к работе Э. Г. Мура (Moore 1900), представленной им Американскому математическому обществу 25 августа 1899 г. Для большей понятности разделим его подход на две части.

Построение

Пусть $I = \{t : 0 \leqslant t \leqslant 1\}$ – единичный отрезок, а $S = \{(x, y): 0 \leqslant x, y \leqslant 1\}$ – единичный квадрат. Для каждого целого положительного n разбиваем I на 4^n равных подотрезков длины 4^{-n} каждый, а S – на 4^n равных подквадратов со стороной 2^{-n} каждый. Сначала для каждого n мы строим взаимно однозначное соответствие между подотрезками I и подквадратами S, удовлетворяющее двум условиям.

Соседство. Соседним подотрезкам соответствуют соседние подквадраты (т. е. квадраты, имеющие общую сторону).

Вложенность. Если после n-го разбиения отрезку I_n соответствует квадрат S_n, то после $(n + 1)$-го разбиения I четырем подотрезкам I_n соответствуют четыре подквадрата S_n.

Заключение

Описанное выше соответствие определяет единственную функцию $f : I \to S$, что мы сейчас и докажем.

Если $t \in I$ не является концевой точкой ни одного подотрезка, то t принадлежит последовательности вложенных замкнутых подотрезков $\{J_1 \supset J_2 \supset J_3 \dots\}$, длины которых стремятся к 0. Для соответствующей последовательности замкнутых вложенных квадратов $\{T_1 \supset T_2 \supset T_3 \dots\}$ диаметры также стремятся к 0, и потому они сходятся к однозначно определенной точке $s \in S$: определим $f(t) = s$.

Концевые точки I, где $t = 0$ и $t = 1$, ассоциированы с левым нижним и правым верхним углом S соответственно.

Если $t \in I$ – внутренняя точка, общая для двух соседних подотрезков J_n и J_n^1 для некоторого n, то она является общей для двух соседних подотрезков J_m и J_m^1 для всех $m > n$ и потому должна принадлежать двум последовательностям вложеннных подотрезков $\{J_n \supset J_{n+1} \supset J_{n+2} \supset \dots\}$ и $\{J_n^1 \supset J_{n+1}^1 \supset J_{n+2}^1 \supset \dots\}$, которым соответствуют две последовательности вложенных квадратов, каждая из которых определяет одну и ту же точку $x \in S$ (квадраты соседние, и их диагонали стремятся к 0): определим $f(t) = s$.

Это и есть кривая Гильберта, определенная как функция из единичного отрезка в единичный квадрат. Далее мы изучим ее свойства.

Сюръективность

Мы построили однозначную функцию $f : I \to S$. Теперь нужно показать, что это «функция на». Каждая точка $s \in S$ принадлежит (по меньшей мере) одной последовательности замкнутых вложенных квадратов, которая сходится к точке s. Соответствующая последовательность замкнутых вложенных квадратов сходится к точке $t \in I$, для которой $f(t) = s$.

Непрерывность

Возьмем две сколь угодно близкие точки I и формализуем это условие, потребовав, чтобы $|t_1 - t_2| < 4^{-n}$. t_1 и t_2 должны лежать в одном и том же или в соседних подотрезках n-го разбиения. Соответствующие образы в худшем случае будут лежать в двух соседних квадратах, образующих прямоугольник со сторонами 2^{-n} и 2×2^{-n}, т. е. $|f(t_1) - f(t_2)| < \sqrt{5} \times 2^{-n}$; соответствующие точки-образы поэтому сколь угодно близки, и функция непрерывна.

Неинъективность

Предположим, что подквадрат S_n с центром C в n-м разбиении соответствует подотрезку I_n. Теперь перейдем к $(n + 1)$-му разбиению этого подквадрата и соответствующего ему подотрезка (рис. 6.6), где числа указывают порядок соответствия между подквадратами и подотрезками. Существует последовательность вложенных подквадратов, лежащих в подквадрате 1, сходящаяся к P, и соответствующая ей последовательность вложенных подотрезков сходится к точке t в подотрезке 1 отрезка I_n. Также существуют последовательности вложенных подквадратов, лежащих в подквадратах 3 и 4, которые сходятся к C, а соответствующие им последовательности подотрезков сходятся к точкам t_3 и t_4, ни одна из которых не может совпадать с t. Это означает, что в I существуют по меньшей мере две разные точки, которые отображаются в C, а поскольку это рассуждение применимо к каждому квадрату в каждом разбиении, кривая имеет бесконечно много кратных точек.

С недифференцируемостью также не составит труда разобраться.

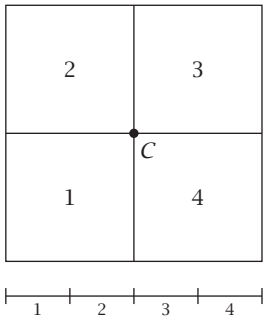


Рис. 6.6. Неинъективность

Недифференцируемость

Доказательство снова проведем для компонентной функции $f_x(t)$, понимая, что для f_y оно строится точно так же. Для произвольного $t \in I$ построим последовательность $t_n \in I$ такую, что $|t - t_n| \to 0$ при $n \to \infty$, но отношение разностей расходится. Для этого рассмотрим последовательность из шестнадцати соседних подотрезков I (каждый длины 4^{-n}), которая содержит t и отображается на подквадрат 4×4, состоящий из шестнадцати подквадратов со стороной 2^{-n} каждый. Где бы в I ни находилась t и в каком бы подквадрате ни находилась $f(t)$, t_n можно выбрать так, чтобы $f(t_n)$ лежала в подквадрате, отделенном от $f(t)$ по меньшей мере одним подквадратом со стороной 2^{-n}; следовательно,

$$|f_x(t) - f_x(t_n)| > 2^{-n}$$

и

$$\left| \frac{f_x(t) - f_x(t_n)}{t - t_n} \right| > \frac{2^{-n}}{16 \times 4^{-n}} = \frac{2^n}{16} \xrightarrow[n \to \infty]{} \infty.$$

Мы завершили рассмотрение кривой Гильберта, заполняющей пространство, и порождающей ее функции. Но это книга о кривых, а у нас до сих пор нет геометрической формы заполняющей пространство функции Пеано – того, что Гильберт извлек из абстрактного построения для описания своей собственной кривой. Именно к этой геометрической форме мы хотим напоследок привлечь внимание читателя.

6.4. Кривая Пеано

Мы имеем дело с довольно сложным примером задачи на изображение кривой. Возвращаясь к построению Пеано, наш первый шаг состоит в определении концевых точек кривой: поскольку $f(0) = (0, 0)$ и $f(1) = f(0_{\dot{3}}2222 \ldots, 0_{\dot{3}}2222 \ldots) = (1, 1)$, то кривая начинается в левом нижнем и заканчивается в правом верхнем углу единичного квадрата.

Теперь заметим, что для произвольных значений букв

$$f(0_{\dot{3}}00t_3t_4t_5\cdots) = \begin{pmatrix} 0_{\dot{3}}0\alpha_2\alpha_3\alpha_4\cdots \\ 0_{\dot{3}}0\beta_2\beta_3\beta_4\cdots \end{pmatrix},$$

т. е. функция отображает подотрезок $[0, \frac{1}{9}]$ в подквадрат $[0, \frac{1}{3}] \times [0, \frac{1}{3}]$.

Аналогично, так как

$$f(0_{\dot{3}}01t_3t_4t_5\cdots) = \begin{pmatrix} 0_{\dot{3}}0\alpha_2\alpha_3\alpha_4\cdots \\ 0_{\dot{3}}1\beta_2\beta_3\beta_4\cdots \end{pmatrix},$$

подотрезок $[\frac{1}{9}, \frac{2}{9}]$ отображается в подквадрат $[0, \frac{1}{3}] \times [\frac{1}{3}, \frac{2}{3}]$, а подотрезок $[\frac{2}{9}, \frac{3}{9}]$ – в подквадрат $[0, \frac{1}{3}] \times [\frac{2}{3}, 1]$ и т. д. Вообще, подотрезок $[\frac{1}{9}(n-1), \frac{1}{9}n]$ отображается в подквадрат n на рис. 6.7a. Поскольку кривая непрерывна, проходит через каждый подквадрат и начинается и заканчивается в противоположных углах квадрата, путь должен быть таким, как показано на рис. 6.7b, а первая итерация выглядит, как на рис. 6.7c.

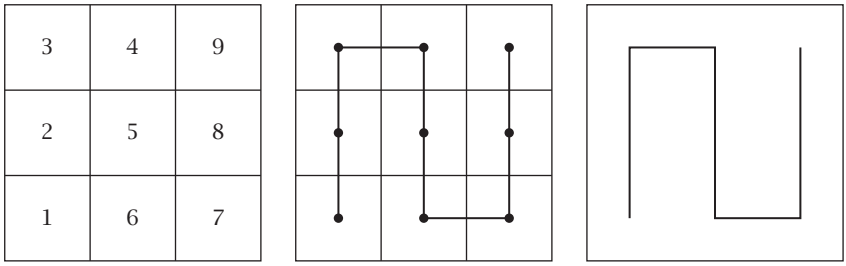


Рис. 6.7. Построение кривой Пеано

Неизбежное дальнейшее разбиение подотрезков ведет далее, как и в построении Гильберта, к рис. 6.8a, затем к рис. 6.8b и т. д. Таким образом, кривая строится рекурсивно.

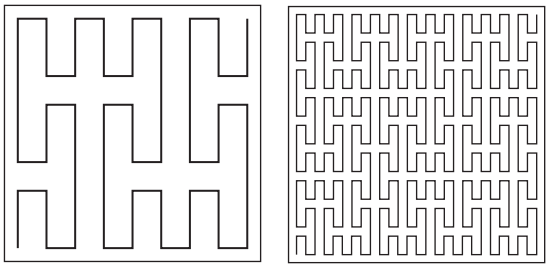


Рис. 6.8. Продолжение построения кривой Пеано

Первоначальное смятение, которое многие математики испытали, когда концепция размерности была поставлена под сомнение результатами Кантора и были предъявлены эти кривые, сменилось, как мы уже отмечали, строгой переоценкой некоторых наивно понимаемых идей, а в итоге математическая копилка пополнилась новыми топологическими инструментами. Виленкин (Vilenkin 1995) очень изящно подвел итог:

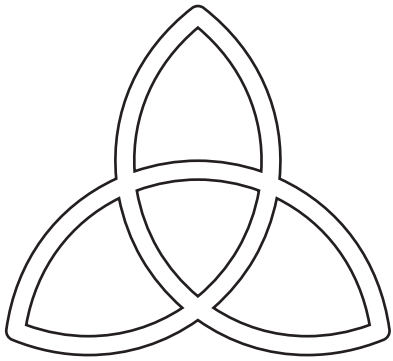
> «Трудно передать словами впечатление, произведенное на математический мир результатом Пеано. Казалось, что все рухнуло, что самые основные математические определения потеряли всякий смысл».

Своим существованием эти кривые сослужили математике хорошую службу. Но в современном мире им нашлись новые применения, поскольку кривая, заполняющая пространство, неизбежно имеет бесконечную длину. В случае кривой Гильберта легко видеть, что после n-й итерации кривая имеет длину $H_n = 2^n - 1/2^n$, а в случае кривой Пеано $P_n = 3^n - 1/3^n$. Поскольку каждая целиком заключена в единичный квадрат, они дают весьма эффективный способ систематической упаковки очень длинной кривой в очень малое пространство: при $n = 10$ длина кривой Гильберта в 1000 с лишним раз больше длины содержащего ее квадрата, а длина кривой Пеано – почти в 60 000 раз. Природа их непрерывности также особая. Сущность непрерывности функции заключается в том,

что образы сколь угодно близких точек будут сколь угодно близкими; природа кривых Гильберта и Пеано (да и других тоже) такова, что образы точек, *очень* близких на единичном отрезке, также *очень* близки в квадрате. На техническом языке это означает, что сохраняется *пространственная локальность*; т. е. кривая «кластеризующая». Если представить себе, что видимый спектр света закодирован числами в единичном отрезке с сохранением порядка, то любую итерацию соответствующего образа можно рассматривать как хребет, на который нанизываются квадратики подходящего цвета; соседние квадратики, скорее всего, будут иметь похожие цвета. У этого свойства есть многочисленные применения: построение индексов для пространственных баз данных, обработка изображений, балансировка нагрузки при параллельной обработке, численное решение дифференциальных уравнений в частных производных, картирование хромосом и т. д. и т. п. Все это мы оставляем интересующемуся читателю для самостоятельного изучения.

Глава 7

Кривые постоянной ширины



Почему именно эти кривые?

На первый взгляд, удивительно, что они вообще существуют. Окружность кажется единственной в этом роде, но на самом деле это лишь один образчик кривых постоянной ширины, а вообще-то их существует бесконечно много, причем принципиально разных. Они бывают неожиданно полезными, а иногда еще и очень красивыми; на рисунке выше изображен вариант *трикетры* (орнамента из трех переплетенных петель). К тому же изучение кривых постоянной ширины естественно включает, казалось бы, разрозненные математические идеи: теорию выпуклых кривых (включая их сложение), огибающие кривых и ряды Фурье. И они естественно представляются новой формой параметрических кривых. Кроме того, они иллюстрируют удивительное различие между роликом и колесом. Для нас единственными аспектами их истории, не вызывающими никакого удивления, является то, что первым их начал изучать Леонард Эйлер в XVIII в. и что вниманию публики их представил Мартин Гарднер (Gardner 1991, глава 18) двумя веками позже.

7.1. Треугольник Рёло...

28 января 1986 г. космический челнок НАСА «Челленджер» взорвался на 73-й секунде полета, в результате чего все семь членов экипажа погибли. В созданную президентскую комиссию по расследованию катастрофы вошел ныне покойный лауреат Нобелевской премии по физике Ричард Фейнман, который впоследствии представил мировым СМИ знаменитую демонстрацию

катастрофического воздействия холода на уплотнительное кольцо – небольшую, но жизненно важную резиновую деталь космического корабля. В свою вторую автобиографическую книгу (Feynman 1988) Фейнман включил много сведений о работе комиссии и о своей роли в ней, и ниже мы приводим его мысли по поводу еще одной возможной причины взрыва:

> «Затем я исследовал нечто такое, что, как мы считали, могло стать обстоятельством, способствующим аварии: когда ракетные ускорители падают в океан, от столкновения с поверхностью их круглая форма (в поперечном сечении) несколько искажается. В Космическом центре им. Кеннеди ускорители разбирают на секции – четыре для каждой ракеты – и по железной дороге отправляют в "Тиокол", расположенный в штате Юта, где заправляют новым топливом. Затем их отправляют поездом обратно во Флориду. Во время транспортировки секции (которые везут на боку) несколько сплющиваются, потому что так называемое твердое топливо (на самом деле довольно "мягкое") очень тяжелое. Общая величина этой деформации составляет лишь доли дюйма, но когда секции ракеты опять собирают вместе, то и малюсенького зазора достаточно, чтобы через него просочились горячие газы: толщина уплотнительных колец всего-то четверть дюйма, а сжимаются они только на две сотых дюйма!
>
> Я подумал о том, чтобы произвести кое-какие расчеты. НАСА предоставило мне все цифры относительно того, каких значений может достигать отклонение от круглого сечения ракеты, и я попробовал вычислить результирующее сдавливание и его локализацию – может быть, минимальное сжатие было именно там, где произошла утечка. Данные НАСА представляли собой измерения, проводившиеся по трем диаметрам через каждые 60°. Но сопоставление диаметров не гарантирует, что все подогнано, – не важно, будь их три, шесть или любое другое количество.
>
> Возьмем, к примеру, фигуру – что-то вроде треугольника со скругленными углами, – в которой три диаметра, измеренные через 60°, имеют равную длину.
>
> Я вспомнил, что, когда был маленьким, видел подобный трюк в музее. Там была зубчатая рейка, которая двигалась назад и вперед абсолютно ровно, и при этом под ней находились некруглые, смешные зубчатые шестерни причудливой формы, поворачивающиеся на валах, которые качались. Это выглядело невозможным, но работало, потому что шестеренки имели такую форму, при которой диаметры были всегда одни и те же.
>
> Таким образом, цифры, предоставленные мне НАСА, не пригодились».

Юный Фейнман, очевидно, видел фигуру, являющуюся предметом настоящей главы, и, возможно, ее обобщения. Мартин Гарднер (Gardner 1991) особенно подчеркнул этот опасный аспект необоснованного допущения на своем собственном примере:

> «Представим себе, что на кораблестроительном заводе собирают корпус подводной лодки, проверяя его цилиндричность промерами максимального диаметра по всем направлениям. Как мы вскоре узнаем, корпус мог бы быть чудовищно деформированным и тем не менее благополучно пройти подобные испытания. Именно поэтому цилиндричность корпуса подводной лодки проверяется специальными шаблонами».

Наличие таких шаблонов позволяет предположить, что измерение равных диаметров замкнутой кривой не гарантирует цилиндричности, – и это действительно

так: окружность и вправду является витриной всех кривых постоянной ширины, но *треугольник Рёло* тоже относится к этому классу, хотя и далек от окружности.

Французская фамилия Франца Рёло выдает его бельгийские корни, а немецкое имя он получил, потому что родился в немецкоговорящей местности поблизости от города, который немцы называют Ахеном, а французы и англичане – Экс-ля-Шапель. Он воплощение нового типа ученого, появившегося в Викторианскую эпоху, *ученого-инженера*, и в этом качестве добился широкой известности и влияния, присоединив к своему имени эпитет «отец современной кинематики». И все же его имя носит только специальная криволинейная фигура: треугольник Рёло. Эта конструкция появилась в главе 3 вышедшей в 1875 г. книги «Кинематика механизмов», в которой автор рассматривает ограничения на вращение, предварительно обсудив ограничения на линейное движение. С помощью простого рисунка, не сильно отличающегося от рис. 7.1, он доказал, что две пары параллельных прямых, каждая из которых касается кривой, определяют ее ширину в перпендикулярных этим прямым направлениях, и если эта ширина не постоянна, то прямые ограничивают вращательное движение. Превратите случайную кривую в окружность, а параллелограмм в ромб или квадрат, и ограничения будут сняты – именно потому, что ширина окружности постоянна. Но далее он замечает, что «фигуры постоянной ширины можно легко построить из дуг окружности», и тут же принимается за дело.

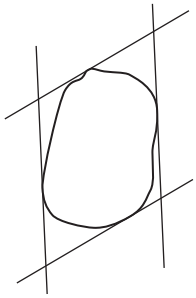


Рис. 7.1. Ограниченное вращательное движение

Результатом построения был, конечно, ныне широко известный треугольник Рёло, образованный равносторонним треугольником, на стороны которого опираются дуги окружности; центр каждой дуги находится в одной из вершин треугольника, и она проходит через две другие вершины, как показано на рис. 7.2. Эту кривую он назвал *равносторонним криволинейным треугольником*.

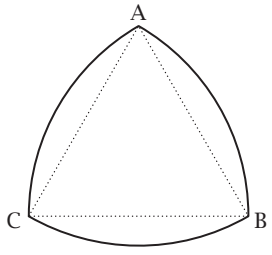


Рис. 7.2. Простой треугольник Рёло

То, что эта фигура имеет постоянную ширину, можно счесть очевидным или же проверить с помощью приведенного ниже рассуждения.

Зафиксируем положение треугольника Рёло. Представим себе, что жесткая Т-образная часть фигуры на рис. 7.3, образованная прямой, проходящей через вершину C и пунктирным перпендикуляром к ней, поворачивается вокруг C на угол 30° против часовой стрелки, так что точка P переходит в точку A; затем, начав вращение из исходного положения, она может повернуться на угол 30° по часовой стрелке, так что точка P перейдет в B. Для каждого промежуточного положения P касательная к дуге окружности AB в точке P будет перпендикулярна радиусу CP и, значит, параллельна прямой, проходящей через C. Эти касательная и прямая образуют пару прямых, между которыми заключена кривая, а расстояние между ними постоянно и равно радиусу окружности. Когда такой криволинейный треугольник катится по плоскости, одна из его вершин постоянно находится в контакте с одной из двух ограничивающих его параллельных прямых.

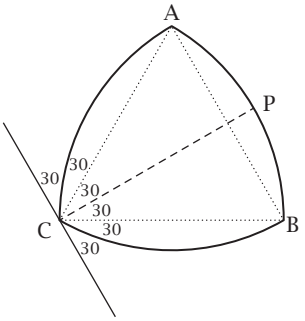


Рис. 7.3. Постоянство ширины треугольника Рёло

Элегантная форма этой кривой приписывалась Леонардо да Винчи, потому что в его бумагах[1] была найдена карта мира, состоящая из восьми октантов глобуса, каждый из которых, будучи распрямлен, принимает форму треугольника Рёло. Эта форма часто встречается в архитектуре, античной и современной, и как каменная деталь, и как часть окна или двери. В некоторых американских городах задвижки пожарных гидрантов закреплены гайками, имеющими такую форму; так делается, чтобы предотвратить несанкционированный доступ, поскольку обычный гаечный ключ просто соскальзывает с такой гайки. Такую же форму имеют гитарные медиаторы, а равно и принадлежащий жене автора древний портновский мелок фирмы Hancock. Все это, конечно, интересно, однако настоящую важность сия элегантная и полезная форма приобретает, когда что-то приводит ее в движение.

Переместить гигантские валуны весом по 25 тонн каждый с известковых холмов графства Марльборо в Стоунхендж (а это около 20 миль) 2500 лет назад можно было, только используя хлысты деревьев в качестве катков, но да Винчи

[1] Хранится в Королевской библиотеке в Виндзорском замке. Вопрос о том, принадлежит ли эта работа самому Леонардо да Винчи или одному из его учеников, вызывает ожесточенные споры.

Бронзового века мог додуматься придать поперечному сечению стволов криволинейную форму, и тогда Стоунхендж был бы построен, потому что по природе своей такой каток сохраняет постоянное расстояние между уровнем земли и лежащей на нем глыбой.

Чтобы разобраться в кинематике движения, проследим за вершиной треугольника Рёло, который катится без проскальзывания вдоль горизонтальной прямой (см. рис. 7.4). Точка, движущаяся по траектории $P \to P_1 \to P_2 \to P_3$ при вращении криволинейного колеса, на участке $P \to P_1$ описывает прямую, потому что это центр дуги окружности QR; на участке $P_1 \to P_2$ она описывает дугу окружности, потому что треугольник вращается вокруг $R = R_1$; наконец, на участке $P_2 \to P_3$ она описывает дугу циклоиды, потому что теперь это точка на окружности, катящейся вдоль прямой. Центр тяжести и все остальные точки описывают дуги окружности, когда треугольник поворачивается вокруг одной из своих вершин, и дуги циклоиды, когда он катится на одной из своих круговых сторон.

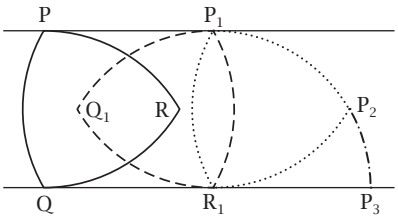


Рис. 7.4. Ролик Рёло

Итак, движение, с одной стороны, простое, как у окружности, а с другой стороны, более сложное и демонстрирует различие между роликом и колесом: каменные глыбы вполне можно было бы перемещать на треугольных катках Рёло, но если бы все четыре колеса безбортовой грузовой платформы имели форму треугольника Рёло, то ее оси поднимались бы и опускались во время движения, а вместе с ними и гигантские валуны.

На этом мы оставим линейное движение треугольника. Вернемся к замечанию Рёло об ограниченном вращательном движении и рассмотрим рис. 7.5a, где треугольник Рёло заключен в квадрат со стороной, равной ширине треугольника, внутри которого он может поворачиваться, так что в любой момент две вершины будут соприкасаться со сторонами квадрата. Траектория вершины теперь состоит из почти полных сторон квадрата, соединенных чем-то вроде эллиптических дуг в окрестности вершин, как показано на рис. 7.5b. Во всех четырех вершинах имеется недостижимая область; для квадрата со стороной 1 можно показать, что ее площадь равна $1 - \frac{\sqrt{3}}{2} - \frac{1}{24}\pi$, так что площадь области, заметаемой вращающимся треугольником Рёло, равна

$$1 - 4\left(1 - \frac{\sqrt{3}}{2} - \frac{\pi}{24}\right) = 2\sqrt{3} + \frac{\pi}{6} - 3 = 0.9877\ldots,$$

т. е. больше 98 % всей площади квадрата. Сравните с площадью, заметаемой вращающейся окружностью, которая равна ¼π = 0.7853... = 79 % площади

квадрата. Зная это, читатель поймет, почему компания Panasonic так рекламировала дизайн своего робота-пылесоса *Rulo*.

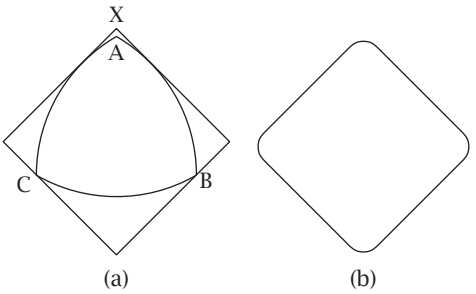


Рис. 7.5. Вращение внутри квадрата

Но поистине самое удивительное применение вращающегося треугольника Рёло мы находим в следующем рекламном листке:

> «Все мы слышали о разводных ключах для левшей[1], ваннах, обитых изнутри мехом, чугунных бананах. Мы называли их дурацкими шутками и отказывались верить в их существование, и тут вдруг появляется инструмент, который умеет сверлить квадратные дырки».

Это объявление разместила компания Watts Brothers Tool Works, когда в 1917 г. Гарри Уоттс, английский инженер из городка Тертл-Крик в Пенсильвании, запатентовал[2] биту для дрели, которую изобрел в 1914 г.: она просверливала квадратные отверстия, поскольку имела в сечении форму модифицированного треугольника Рёло. Чтобы справиться с вышеупомянутым эллиптическим дефектом, который обычно измеряется как AX = AX = $h(\sqrt{2} - \frac{1}{2} - \frac{\sqrt{3}}{2})$ (см. рис. 7.5a), понадобилась инженерная смекалка; кроме того, нужно было решить проблему блуждающего центра тяжести, который описывал траекторию, состоящую из четырех одинаковых эллиптических дуг. Но эти проблемы были успешно разрешены, и квадратные дырки стали реальностью – каковой остаются и сейчас.

Это воплощение кривых постоянной ширины можно считать крайним случаем: угол при ее вершине равен 120°, и ни у какой кривой постоянной ширины угол не может быть меньше. Кроме того, из всех кривых заданной постоянной ширины окружность имеет наибольшую площадь, а треугольник Рёло – наименьшую (равную $\frac{1}{2}w^2(\pi - \sqrt{3})$ для ширины w).

Прежде чем расстаться с этой главной кривой постоянной ширины, восхитительной в своей простоте, элегантности и полезности, кратко упомянем еще некоторые сферы ее применения. В качестве одного из применений часто называют крышки люков, потому что такие крышки, как и круговые, не могут провалиться в колодец, который закрывают. Динамики фирмы Cambridge Soundworks

---

[1] Несуществующий инструмент, за которым часто посылают новичков, когда хотят над ними подшутить. –*Прим. перев.*

[2] Патент США «Рабочий элемент для сверления или бурения», номера 1 241 675/6/7 от 25 сентября 1917 г.

OontZ, поражающие невероятным звучанием, также имеют такую форму; как и индейские амулеты «ловцы снов». Их вращательные свойства используются в механизмах протяжки пленки. Они встречаются в фантастике[1] и в исследованиях по химии (Ng and Fan 2014). И ручка на письменном столе автора тоже имеет форму треугольника Рёло, чтобы не скатывалась, а ее центр масс расположен выше, чем у ручек с шестиугольным сечением, на случай, если она покатится. И закончим мы развенчанием заблуждений: как настоящий треугольник Рёло не может сверлить квадратные отверстия, так и в конструкции роторного двигателя Ванкеля он применяется не в идеальной, а в слегка уплощенной форме.

7.2. … И ЕГО ОБОБЩЕНИЯ

Коль скоро идея построения круговых дуг на сторонах равностороннего треугольника оказалась такой плодотворной, на ум приходит очевидное обобщение (на которое указал еще Рёло) – повторить то же самое для других правильных многоугольников, но обязательно с нечетным числом сторон. На рис. 7.6a показан результат для пятиугольника, а на рис. 7.6b – для семиугольника соответственно. По той же причине, что и выше, обе кривые имеют постоянную ширину.

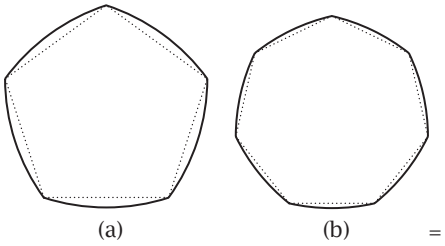


Рис. 7.6. Еще две кривые постоянной ширины

Британские монеты по 20 и 50 пенсов, ирландский 50-пенсовик, ямайский доллар и монета ОАЭ в 50 филсов – все они имеют форму семиугольника Рёло, а канадский доллар – его 11-угольный вариант.

Более того, мы можем модифицировать базовое построение, например с использованием равностороннего треугольника, как изображено на рис. 7.7.

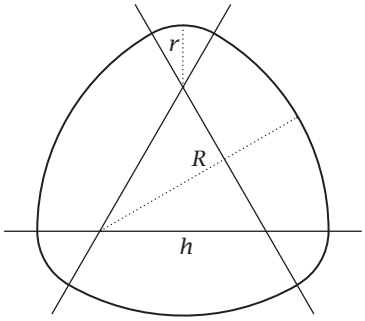


Рис. 7.7. Продолженный равносторонний треугольник Рёло

[1] Смотрите, например, рассказ Пола Андерсона «Треугольное колесо».

Снова начнем с равностороннего треугольника со стороной h и продолжим его стороны, так чтобы образовались внешние углы. Из каждой вершины как из центра опишем большую дугу некоторого фиксированного радиуса R, пересекающую соответствующие продолжения сторон. Также из каждой вершины опишем малую дугу радиуса $r = R - h$, пересекающую те же прямые с противоположной от вершины стороны. Дуги должны соединиться, и мы получаем кривую постоянной ширины $R + r$, поскольку если протащить касательную в любой точке сквозь фигуру к противоположной дуге, то она пройдет расстояние $R + r$.

Если мы готовы пожертвовать симметрией, то можем начать с разностороннего треугольника, изображенного на рис. 7.8, и описать дуги окружности с центрами в каждой вершине, пересекающие продолжения сторон, так чтобы их радиусы удовлетворяли простым соотношениям, необходимым, чтобы кривая имела постоянную ширину $a + c - b + 2x$, где x – радиус первой дуги (мы решили назвать так дугу с центром в вершине B):

$$\begin{aligned} &\text{A: } r = a + x - b, && R = c + x,\\ &\text{B: } r = x, && R = a + c - b + x,\\ &\text{C: } r = c + x - b, && R = a + x. \end{aligned}$$

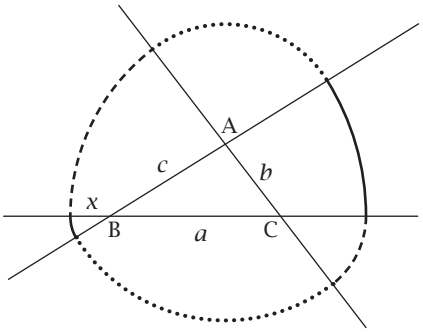


Рис. 7.8. Разносторонний треугольник Рёло

Это не что иное, как частный случай нескольких пересекающихся прямых. На рис. 7.9 мы видим четыре прямые, пересекающиеся в пяти точках; четыре из них пронумерованы, а пятая нам не понадобится. Две пересекающиеся прямые, порождающие каждую из пронумерованных точек, ограничивают две области, помеченные тем же числом. Мы начинаем с точки 1 и описываем дугу окружности произвольного радиуса, соединяющую обе прямые. Продолжаем так для точек 2, 3, 4 по порядку, описывая дуги, так чтобы получающаяся кривая оставалась непрерывной. Затем повторяем эту процедуру, описывая дуги в противоположных областях.

Если обозначить x, y, z длины трех отрезков прямой L, то с помощью выписывания несложных соотношений между длинами отрезков (с введением временных переменных для обозначения длин других отрезков) можно показать, что все четыре отрезка прямых, соединяющие противоположные дуги, имеют длину $x + y + z$, так что ширина кривой постоянна.

Наконец, обсудим их почти наверняка первое появление в качестве математического объекта в самой короткой из всех публикаций Эйлера: анонимном

письме, состоящем из восьми строчек, в журнал «Nova Acta Eruditorum» под названием «Геометрическая задача, публично предложенная анонимным геометром» (Euler 1745, стр. 523, E79)[1]. Ниже приведен перевод латинского оригинала с отсылкой к рис. 7.10:

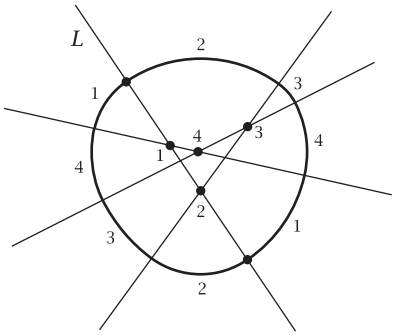


Рис. 7.9. Общий случай

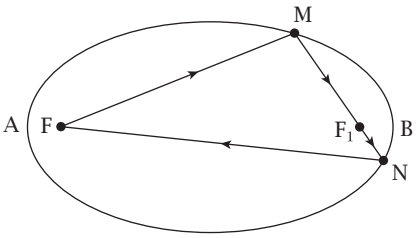


Рис. 7.10. Катоптриса

> «Дана точка *F*. Найти все кривые *AMBN* такие, что любой луч, исходящий из *F*, после двух отражений в точках *M* и *N* возвращается в *F*. Первые среди таких кривых – эллипсы с фокусом *F*. Однако существует бесконечно много других кривых, обладающих этим свойством, изучение которых выходит за пределы анализа».

Из оптического свойства эллипса действительно вытекает, что любой эллипс – пример такой кривой: пусть F – один фокус, а M – произвольная точка эллипса, тогда луч света, исходящий из F и приходящий в M, отразится во второй фокус, F_1, а его продолжение достигнет эллипса в точке N и обязательно отразится обратно в F. Эйлер назвал кривую, обладающую этим свойством, *катоптрисой* (от греческого слова, обозначающего зеркало).

Какой бы интерес эта задача ни вызвала у других ученых, уступила она анализу самого Эйлера в серии статей[2], третья из которых пересекается с нашим предметом. Почти за сто лет до публикации Рёло своей книги, содержащей изучение его треугольника (и многое сверх того), 12 мая 1774 г. Эйлер доложил Санкт-Петербургской Академии наук свою работу «De curvis triangularibus («О треугольных кривых»), которую позже опубликовал в виде третьей из вы-

[1] Возможно, поставленная Эйлеру Христианом Гольдбахом.

[2] E85, 106, 513.

шеупомянутых статей. Треугольные кривые, о которых писал Эйлер, похожи на изображенную на рис. 7.11a и называются *трехрогими астроидами*, или просто *дельтоидами*. Он использовал их для построения кривых постоянной ширины, но не в качестве конечной цели, а как один из этапов поиска катоптрис. И для этой цели прибег к *эвольвенте* кривой.

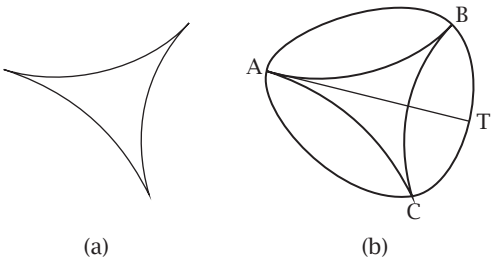


Рис. 7.11. Дельтоида и ее построение

Стандартный способ наглядно изобразить эвольвенту кривой – представить, что к рогу A на рис. 7.11b прикреплен кусок нити длины, равной длине дуги, и расположен вдоль этой дуги. Держа нить туго натянутой, разматывайте ее и следите за траекторией, которую описывает ее свободный конец; эта траектория и называется эвольвентой. На рисунке прослежена дуга BT эвольвенты, которая завершится дугой BC, если мы обратим процесс и будем наматывать нить вдоль дуги AC. Повторив ту же процедуру для двух других рогов, мы завершим построение эвольвенты; эту конкретную эвольвенту Эйлер назвал *орбиформой*, поскольку понял, что у нее, как и у окружности, постоянная ширина. Обоснование этого факта достаточно понятно. Взгляните на рис. 7.12: когда нить разматывается с двух вершин, каждая точка P на дельтоиде определяет две точки Q и R на орбиформе, причем расстояние по прямой равно расстоянию вдоль кривой, так что PQ = PC и PR = PB. Поскольку в любой момент времени прямая часть разматывающейся нити является касательной к дельтоиде, линия QPR прямая и касается дельтоиды в точке P, а следовательно, QR = PQ + PR – длина нити и, стало быть, постоянна. Поскольку PQ и PR – мгновенные радиусы вращения вокруг центра P, прямая QPR перпендикулярна обеим ограничивающим касательным к орбиформе и, значит, QR – ширина кривой.

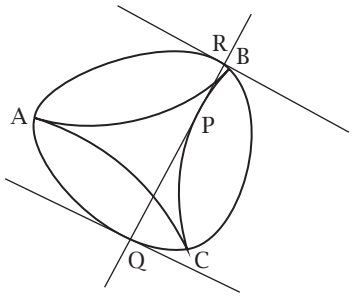


Рис. 7.12. Орбиформа

Разобравшись с орбиформой, Эйлер предложил первое известное нам применение кривой постоянной ширины: для построения катоптрис. Его краткий комментарий занимает всего несколько строчек в разделе 5 статьи и сопровождается чертежом, который мы с некоторыми изменениями воспроизвели на рис. 7.13, сохранив его обозначения.

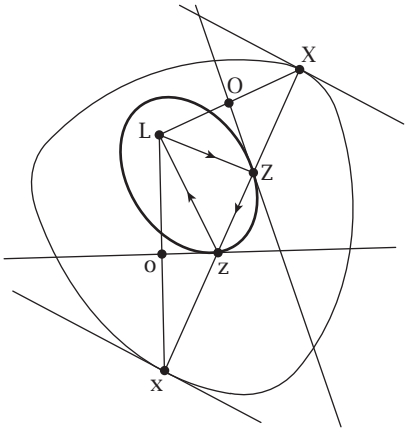


Рис. 7.13. Построенная орбиформа

Из этого довольно запутанного на первый взгляд набора линий можно выделить орбиформу с двумя противолежащими касательными и диаметр вкупе с произвольной точкой L, находящейся внутри: это фокальная точка, из которой луч света исходит и после двух отражений от построенной катоптрисы возвращается назад. Что касается деталей построения, то Эйлер приложил все силы, чтобы добиться ясности, и ниже мы приводим разбитый на пункты перевод:

- соединим L с двумя противоположными точками X и x на орбиформе;
- построим срединный перпендикуляр к LX, проходящий через O, и аналогичный перпендикуляр к Lx, проходящий через o;
- обозначим Z и z точки, в которых эти перпендикуляры пересекают Xx;
- эти точки лежат на катоптрисе, которая на нашем рисунке обведена жирной линией;
- построим треугольник LZz, определяющий искомый путь луча.

Что до обоснования, то Эйлер настроен весьма оптимистично, замечая, что доказательство нетрудное, но он не включил его, не желая излишне затягивать рассмотрение предмета, и надеется, что его за это простят. Однако тут *очень даже* есть что доказывать. Во-первых, что действительно порождается замкнутая кривая, а во-вторых – что прямые OZ и oz являются касательными к ней; все это мы оставляем интересующемуся читателю.

Мы закончили с чисто геометрическими способами построения кривых постоянной ширины, но прежде чем завершить этот раздел, заметим, что уже построенные нами кривые можно использовать для построения других подобных кривых с помощью операции «сложения» выпуклых областей, изобретение которой приписывается немецкому математику российского происхождения Герману Минковскому. *Сумма Минковского* строится следующим образом.

Пусть имеются две выпуклые области на плоскости, R_1 и R_2, включающие граничные кривые C_1 и C_2. Выберем произвольное начало координат O и рассмотрим радиус-вектор $\boldsymbol{r}_1$ произвольной точки R_1 и радиус-вектор $\boldsymbol{r}_2$ произвольной точки R_2. Образуем вектор-сумму $\boldsymbol{r}_1 + \boldsymbol{r}_2$ и определим

$$R_1 + R_2 = \{\text{множество точек плоскости с радиус-векторами вида } \boldsymbol{r}_1 + \boldsymbol{r}_2\}.$$

Иными словами, сумма двух областей образуется сложением радиус-векторов точек, принадлежащих R_1 и R_2. То есть сумма – это множество образов параллельных переносов одной области на радиус-векторы каждой точки другой области. Определим сумму двух выпуклых кривых $C_1 + C_2$ как граничную кривую $R_1 + R_2$. Отметим, что это не то же самое, что сложение двух граничных кривых. Например, сложение двух отрезков, исходящих из начала координат, дает параллелограмм, а не кривую. Из общей теории такого сложения нас будут интересовать три результата:

- сумма двух выпуклых фигур выпукла;
- сумма инвариантна относительно изменения начала координат и параллельного переноса слагаемых; это приводит просто к параллельному переносу результирующей фигуры (вращение слагаемых – совсем другое дело);
- ширина суммы двух кривых равна сумме ширин отдельных кривых в том же направлении.

Таким образом, если две кривые имеют постоянную ширину, то ширина их суммы также постоянна, и мы можем воспользоваться этим фактом для построения новых кривых постоянной ширины из уже известных. На самом деле пример на рис. 7.7 – это сумма треугольника Рёло с окружностью.

Итогом нашего обсуждения является построение бесконечного числа кривых постоянной ширины геометрическими способами. Но можно построить и другие, подойдя к вопросу аналитически.

7.3. И их обобщение...

Начнем этот раздел с задачи из элементарной аналитической геометрии: как выглядит уравнение прямой, перпендикулярной данному отрезку прямой и проходящей через его конец? На рис. 7.14 изображена постановка задачи, когда один конец отрезка совпадает с началом координат, а другой находится в точке P($p\cos\theta$, $p\sin\theta$); его длина, следовательно, равна p, а угол с положительным направлением оси x равен θ.

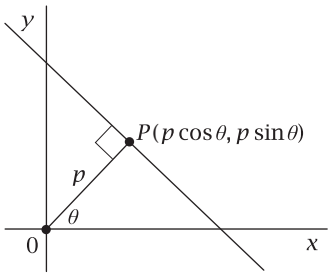


Рис. 7.14. Элементарная геометрия

Коэффициент наклона отрезка равен tg θ, а интересующей нас прямой – -ctg θ, поэтому ее уравнение имеет вид $y - p \sin\theta = -\text{ctg}\,\theta(x - p\cos\theta)$, или после упрощения $-x\cos\theta + y\sin\theta = p$.

Если мы теперь позволим p изменяться вместе с θ, то получим функциональную связь $x\cos\theta = y\sin\theta = p(\theta)$ и семейство прямых, параметризованное θ.

Далее перейдем к кривым, получившим общее название «овалы»; для их описания полезно мысленно представлять себе яйцо, хотя, как мы увидим, это не совсем точно. На рис. 7.15 показан общий вид такой кривой, обязательно замкнутой и выпуклой; это значит, что для любых двух точек, находящихся внутри (или на самой) кривой, отрезок, соединяющий их, целиком находится внутри кривой. Кривая также достаточно гладкая, так что в каждой ее точке существует однозначно определенная касательная. Без ограничения общности можно считать, что начало координат находится внутри кривой.

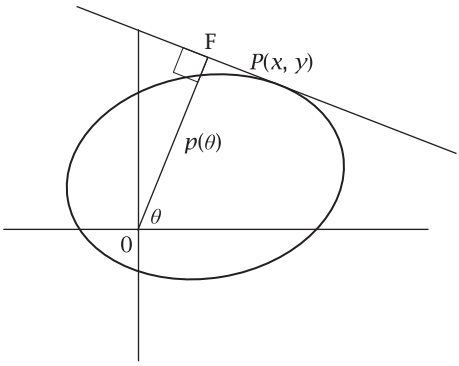


Рис. 7.15. Овал

Роль рассмотренной выше прямой заключается в том, что она является касательной к кривой, а длина перпендикулярного ей отрезка $p(\theta)$ называется *опорной функцией* кривой; эта функция θ периодическая с периодом 2π. В терминах опорной функции мы и выпишем новую форму параметрических уравнений кривой.

Кривая является огибающей своих касательных (см. приложение D), и для вывода ее уравнения мы рассмотрим функцию

$$F(\theta, x, y) = x\cos\theta + y\sin\theta - p(\theta)$$

и продифференцируем ее по θ:

$$\frac{\partial F}{\partial \theta} = -x\sin\theta + y\cos\theta - p'(\theta) = 0.$$

Тогда уравнения огибающей имеют вид

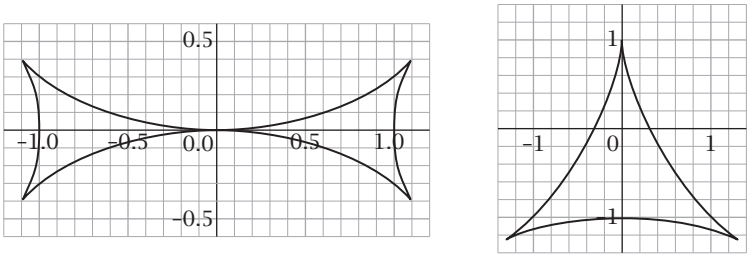
$$-x\sin\theta + y\cos\theta = p'(\theta),$$

$$x\cos\theta + y\sin\theta = p(\theta)$$

и легко преобразуются в параметрическую форму:

$$x = p(\theta)\cos\theta - p'(\theta)\sin\theta,$$
$$y = p(\theta)\sin\theta + p'(\theta)\cos\theta.$$

Любой выбор дифференцируемой функции $p(\theta)$ дает некоторую кривую, которая в общем случае вряд ли будет выпуклой, а если функция непериодическая, то и не замкнутой. Например, $p(\theta) = \cos^2\theta$ дает галстук-бабочку, изображенный на рис. 7.16a, а $p(\theta) = \sin\theta\cos 2\theta$ – криволинейный треугольник на рис. 7.16b, который известен нам как генератор кривой постоянной ширины.



(a) (b)

Рис. 7.16. Галстук-бабочка и астероид

Разумеется, опорная функция связана с шириной овала, определенной (как всегда) как расстояние по перпендикуляру между двумя параллельными касательными и, стало быть, функцией

$$w(\theta) = p(\theta) + p(\theta + \pi),$$

а если эта ширина постоянна, то мы ищем опорные функции, для которых

$$p(\theta) + p(\theta + \pi) = \alpha$$

для некоторой положительной постоянной ширины α.

Пустяковые усилия и результаты приносят пустяковые. Опорная функция должна быть дифференцируемой и иметь период 2π, так что на ум попрежнему приходят sin и cos, а при небольшом напряжении памяти – их связь в самом знаменитом тригонометрическом тождестве: $\cos^2\theta + \sin^2\theta = 1$. Руководствуясь этими соображениями, возьмем $p(\theta) = \alpha\cos^2(k\theta)$ для некоторой постоянной k. Требуется, чтобы $p(\theta + \pi) = \alpha\cos^2(k\theta + k\pi) = \alpha\sin^2(k\theta)$, и потому

$$\cos(k\theta + k\pi) = \pm\sin(k\theta) \to k = \tfrac{1}{2}(2n + 1);\ n = 0, 1, 2, \ldots.$$

Начав с $n = 0$, так что $p(\theta) = \alpha\cos^2\tfrac{1}{2}\theta$, получим фигуру, очень похожую на наш галстук-бабочку, но множитель ½ оказывает заметное влияние, потому что

$$\begin{aligned} x &= \alpha\cos^2\tfrac{1}{2}\theta\cos\theta + \alpha\sin\tfrac{1}{2}\theta\cos\tfrac{1}{2}\theta\sin\theta \\ &= \tfrac{1}{2}\alpha(1 + \cos\theta)\cos\theta + \tfrac{1}{2}\alpha\sin^2\theta = \tfrac{1}{2}\alpha(1 + \cos\theta), \\ y &= \alpha\cos^2\tfrac{1}{2}\theta\sin\theta - \alpha\sin\tfrac{1}{2}\theta\cos\tfrac{1}{2}\theta\cos\theta \\ &= \tfrac{1}{2}\alpha(1 + \cos\theta)\sin\theta - \tfrac{1}{2}\alpha\sin\theta\cos\theta = \tfrac{1}{2}\alpha\sin\theta. \end{aligned}$$

С точностью до масштабного коэффициента α это декартово уравнение окружности:

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

Не бог весть какой прогресс, но, не смущаясь неудачей, попробуем следующий случай, $n = 1$ и, следовательно, $p(\theta) = \alpha \cos^2 \tfrac{1}{2}\theta$. Получается отличная кривая, изображенная на рис. 7.17, которая бросает вызов самой идее постоянной ширины и которую вряд ли можно назвать выпуклой.

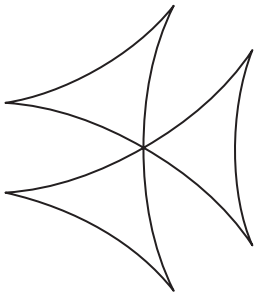
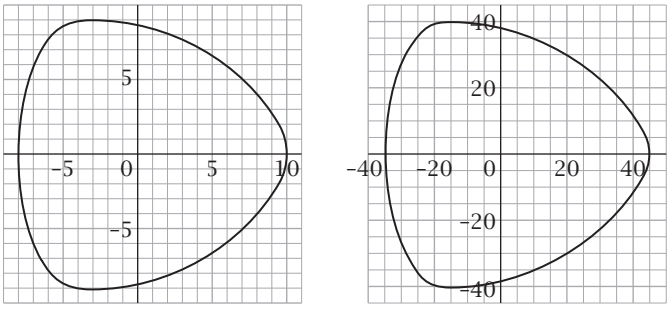


Рис. 7.17. Любопытная кривая

Ключ к прогрессу лежит в снятии ограничений на форму опорной функции, которая вполне могла бы быть и более гибкой, например $p(\theta) = \alpha \cos^2((n + \tfrac{1}{2})\theta + \beta$, и, экспериментируя со значениями параметров, мы действительно получаем то, что нужно: $n = 1$ с $\alpha = 2$, $\beta = 8$ или $\alpha = 10$, $\beta = 35$ и т. д. дает искомое; эти два случая показаны на рис. 7.18a и b соответственно.



(a) (b)

Рис. 7.18. (a) $n = 1, \alpha = 2, \beta = 8$; (b) $n = 1, \alpha = 10, \beta = 35$

При изменении n изменяется число вершин: $n = 2$, $\alpha = 2$, $\beta = 25$ дает рис. 7.19a, а $n = 7$, $\alpha = 1$, $\beta = 25$ – рис. 7.19b. Все они похожи на многоугольники Рёло и все являются самозванцами.

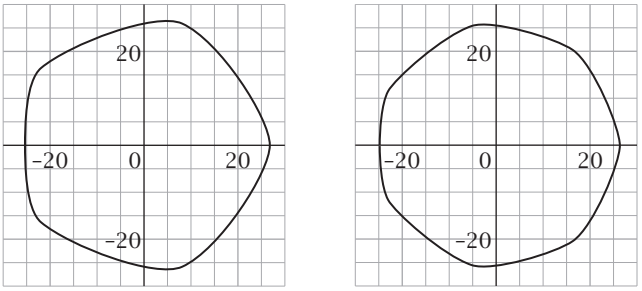


Рис. 7.19. (a) $n = 2, \alpha = 2, \beta = 25$; (b) $n = 7, \alpha = 1, \beta = 25$

К общему случаю мы подойдем, воспользовавшись формулой косинуса двойного угла $\cos 2\theta = 2\cos^2\theta - 1$, чтобы исключить дробь, а заодно преобразовать квадрат функции в функцию кратного угла:

$$\begin{aligned} p(\theta) &= \alpha \cos^2((n + \tfrac{1}{2})\theta) + \beta \\ &= \tfrac{1}{2}\alpha[1 + \cos((2n + 1)\theta)] + \beta \\ &= \tfrac{1}{2}a\cos((2n + 1)\theta) + \tfrac{1}{2}a + \beta. \end{aligned}$$

В этом случае параметрические уравнения принимают вид:

$$\begin{aligned} x &= [\tfrac{1}{2}a\cos((2n + 1)\theta) + \tfrac{1}{2}a + b]\cos\theta \\ &\quad + \tfrac{1}{2}(2n + 1)a\sin((2n + 1)\theta)\sin\theta, \\ y &= [\tfrac{1}{2}a\cos((2n + 1)\theta) + \tfrac{1}{2}a + b]\sin\theta \\ &\quad - \tfrac{1}{2}(2n + 1)a\sin((2n + 1)\theta)\cos\theta, \end{aligned}$$

что после несложных, хотя и утомительных, преобразований сводится к

$$\begin{aligned} x &= \tfrac{1}{2}(n + 1)a\cos 2n\theta - \tfrac{1}{2}na\cos 2(n + 1)\theta + (\tfrac{1}{2}a + b)\cos\theta, \\ y &= -\tfrac{1}{2}(n + 1)a\sin 2n\theta - \tfrac{1}{2}na\sin 2(n + 1)\theta + (\tfrac{1}{2}a + b)\sin\theta. \end{aligned}$$

И мы имеем общую форму класса кривых постоянной ширины (при подходящих значениях параметров).

А следующая опорная функция уникальна в своем роде. Рассмотрим

$$p(\theta) = \sum_r a_r \cos((2r + 1)\theta) + b.$$

Тогда

$$
\begin{aligned}
&p(\theta) + p(\theta + \pi) \\
&\quad = \sum_r a_r \cos((2r+1)\theta) + b + \sum_r a_r \cos(2r+1)(\theta + \pi) + b \\
&\quad = \sum_r a_r \cos((2r+1)\theta) + b + \sum_r a_r \cos[(2r+1)\theta + (2r+1)\pi] + b \\
&\quad = \sum_r a_r \cos((2r+1)\theta) + b - \sum_r a_r \cos((2r+1)\theta) + b = 2b.
\end{aligned}
$$

Это кривая постоянной ширины (хотя и необязательно выпуклая). Можно было бы взять вместо этого синус, для которого приведенное выше рассуждение по-прежнему справедливо, или комбинацию обеих форм (рассуждение снова справедливо), получив тем самым опорную функцию вида

$$p(\theta) = \sum_r a_r \cos((2r+1)\theta) + \sum_r b_r \sin((2r+1)\theta) + b. \tag{7.1}$$

При этом мы прошлись по верхам глубокой теории рядов Фурье: опорная функция периодическая и непрерывная, а если мы еще потребуем, чтобы ее производная также была непрерывной, то, согласно теории рядов Фурье, ее можно будет представить в виде конечной или бесконечной суммы синусов и косинусов – точно такой, как показана выше.

И еще одно последнее наблюдение, ради которого мы вернемся к одному из рассмотренных выше случаев ($n = 1$, $\alpha = 2$, $\beta = 8$). Имеем

$$x = 9\cos\theta + 2\cos 2\theta + \cos 4\theta = -3 + 9\cos\theta + 12\cos^2\theta - 8\cos^4\theta,$$

$$y = 9\sin\theta - 2\sin 2\theta - \sin 4\theta = \sin\theta\,(9 - 8\cos^3\theta).$$

Последняя форма обоих уравнений вытекает из стандартных тригонометрических тождеств. Возведение уравнения для y в квадрат привносит $\sin^2\theta$ и, стало быть, $\cos^2\theta$, и мы имеем параметризацию относительно переменной $\cos\theta$, которую в принципе (а на практике с помощью компьютера) можно исключить, получив уравнение в декартовых координатах; в данном случае оно имеет вид:

$$
\begin{aligned}
&y^8 + 4x^2y^6 - 48xy^6 - 45y^6 + 6x^4y^4 - 80x^3y^4 + 441x^2y^4 \\
&\quad + 16632xy^4 - 41283y^4 + 4x^6y^2 - 16x^5y^2 - 519x^4y^2 \\
&\quad + 11088x^3y^2 - 82566x^2y^2 - 799146xy^2 + 7950960y^2 \\
&\quad + x^8 + 16x^7 + 19x^6 - 5544x^5 - 41283x^4 + 266382x^3 \\
&\quad + 7950960x^2 - 373248000 = 0,
\end{aligned}
$$

и было сгенерировано в мгновение ока; скептический, но усидчивый читатель может ввести это уравнение в графопостроитель – и снова получить рис. 7.18а. Конечно, более сложные (в терминах количества тригонометрических членов в выражении (7.1)) опорные функции дают еще более сложные полиномиальные уравнения. С точки зрения степени простейшим является приведенное выше, степени 8: известно (Bardet and Bayen 2018), что если последний ненулевой член содержит в синусе или косинусе угол $(2N+1)\theta$, то степень многочлена будет равна $4N + 4$.

7.4. Окружность во всем, кроме названия?

Этот последний раздел посвящен одному свойству, которым обладают кривые постоянной ширины и которое в то же время является фундаментальной характеристикой окружности: длина окружности равна π, умноженному на диаметр (или ширину). Результат очевиден в частном случае фигур Рёло: если такая фигура построена на основе правильного многоугольника с n сторонами длины r, то ее периметр будет равен $n \times r \times 2\pi/n = 2\pi r$. Завораживает общая природа результата, и, чтобы ее установить, нам понадобится дополнительный математический аппарат.

Нужный нам результат связан с трагической фигурой француза Жозефа-Эмиля Барбье, человека, который оставил многообещающую должность в Парижской обсерватории, разорвал отношения с друзьями и коллегами и после 15-летнего отсутствия был обнаружен в лечебнице для душевнобольных. Свой вклад в кривые постоянной ширины он внес в 1860 г., задолго до появления первых признаков душевного расстройства, в тот год, когда из выдающегося студента Парижского лицея он стал профессором другого лицея в Ницце, но на этом поприще не добился успеха. Его результат, *теорема Барбье*, формулируется следующим образом: периметр замкнутой выпуклой кривой постоянной ширины $w > 0$ равен πw.

Сейчас известно несколько ее доказательств, и мы выберем то, что близко к оригинальному. Для этого нужна параметризация опорной функции и немного анализа, выходящего за рамки средней школы. Напомним читателю вид параметризации:

$$x = p(\theta)\cos\theta - p'(\theta)\sin\theta,$$
$$y = p(\theta)\sin\theta + p'(\theta)\cos\theta$$

и стандартную формулу длины дуги кривой в параметрической форме:

$$s = \int_{\theta_1}^{\theta_2} \sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}\theta}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}\theta}\right)^2}\,\mathrm{d}\theta.$$

Зная это, выполним естественные преобразования:

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}\theta} &= p'(\theta)\cos\theta - p(\theta)\sin\theta - p''(\theta)\sin\theta - p'(\theta)\cos\theta \\
&= -(p(\theta) + p''(\theta))\sin\theta, \\
\frac{\mathrm{d}y}{\mathrm{d}\theta} &= p'(\theta)\sin\theta + p(\theta)\cos\theta + p''(\theta)\cos\theta - p'(\theta)\sin\theta \\
&= (p(\theta) + p''(\theta))\cos\theta, \\
\left(\frac{\mathrm{d}x}{\mathrm{d}\theta}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}\theta}\right)^2 &= (p(\theta) + p''(\theta))^2\sin^2\theta + (p(\theta) + p''(\theta))^2\cos^2\theta \\
&= (p(\theta) + p''(\theta))^2.
\end{aligned}$$

В нашем случае пределы интегрирования равны 0 и 2π, и мы замечаем, что

$$w = p(\theta) + p(\theta + \pi) \to p(\theta + \pi) = w - p(\theta),$$
$$0 = p'(\theta) + p'(\theta + \pi),$$
$$0 = p''(\theta) + p''(\theta + \pi) \to p''(\theta + \pi) = -p''(\theta),$$

а это означает, что

$$\begin{aligned} s &= \int_0^{2\pi} p(\theta) + p''(\theta)\,\mathrm{d}\theta \\ &= \int_0^{\pi} p(\theta) + p''(\theta)\,\mathrm{d}\theta + \int_{\pi}^{2\pi} p(\theta) + p''(\theta)\,\mathrm{d}\theta, \end{aligned}$$

и, полагая во втором интеграле $u = \theta - \pi$, получаем

$$\begin{aligned} s &= \int_0^{\pi} p(\theta) + p''(\theta)\,\mathrm{d}\theta + \int_0^{\pi} p(u + \pi) + p''(u + \pi)\,\mathrm{d}u \\ &= \int_0^{\pi} p(\theta) + p''(\theta)\,\mathrm{d}\theta + \int_0^{\pi} w - p(u) - p''(u)\,\mathrm{d}u \\ &= \int_0^{\pi} p(\theta) + p''(\theta) + w - p(\theta) - p''(\theta)\,\mathrm{d}\theta = \int_0^{\pi} w\,\mathrm{d}\theta = \pi w. \end{aligned}$$

Мы доказали теорему Барбье. Уместно заметить, что в литературе ее можно встретить под названием *теоремы Меллиша* в честь молодого канадского математика Артура Престона Меллиша, жизнь короткого оборвалась слишком рано: он умер в 1930 г. в возрасте 25 лет. После его смерти коллеги по университету Брауна разобрали его заметки и подготовили на их основе статью, которая была опубликована в журнале «Annals of Mathematics» в 1931 г. (Mellish 1931). Там содержится не только другое доказательство результата, но и установлена эквивалентность следующих утверждений об овале:

- это кривая постоянной ширины;
- она имеет постоянный диаметр[1];
- все нормали двойные[2];
- сумма радиусов кривизны в противоположных точках постоянна.

Мы закончили рассказ об очень красивой теореме, доказать которую удалось с помощью математики, изучаемой в средней школе, но если вспомнить о сложении овалов по Минковскому, то она становится совершенно очевидной: нетрудно показать, что сумма кривой постоянной кривизны с ней самой, повернутой на 180°, является окружностью этой ширины. Отсюда сразу вытекает искомый результат.

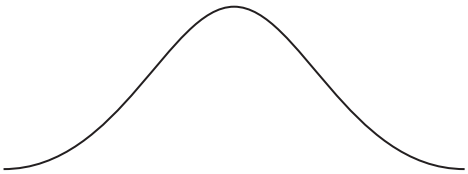
---

[1] *Шириной* называется наименьшее расстояние между параллельными касательными, а *диаметром* – наибольшее.

[2] То есть в тех случаях, когда это имеет смысл, нормаль к кривой в одной точке пересекает ее в противоположной точке, где тоже является нормалью к кривой.

Глава 8

Нормальная кривая



Почему именно эта кривая?

Хотя вопроса «Какое самое важное статистическое распределение?» лучше бы избегать, но если уж он задан, то самый разумный ответ – *нормальное распределение*. Оно имеет фундаментальное значение само по себе и является главным средством аппроксимации других распределений, как непрерывных, так и дискретных, – в силу *центральной предельной теоремы*. Неважно, напоминает оно читателю поперечное сечение колокола, как артиллерийскому офицеру, статистику страхового общества и математику Эспри Паскалю Жуффре в 1872 г., или треуголку, головной убор французской жандармерии в XIX в., как статистику Фрэнсису Исидору Эджуорту в 1888 г., – эта форма играет важнейшую роль в обширном и зачастую весьма запутанном мире статистического анализа. В ее истории естественным образом выделяются три части, и именно это наблюдение мы положили в основу своего подхода.

8.1. Полезный вопрос

Шотландец, сэр Александр Камминг, второй баронет Кутера (1690–1775), одно время был одновременно членом Шотландской коллегии авдокатов и капитаном русской армии, потом самопровозглашенным и признанным королем племени чероки, получившим в сопровождении вождей чероки аудиенцию у короля Георга II в Виндзорском замке, потом мошенником, укравшим у американских поселенцев крупные суммы денег, потом алхимиком, надеявшимся превратить основные металлы в золото, потом заключенным в Лондонской долговой тюрьме на Флит-стрит и, наконец, нищим братом на попечительстве благотворительного фонда в Чертерхаусе, где и умер в возрасте 85 лет. А еще он был членом Королевского общества. Понятное дело, до тех пор пока его не исключили за неуплату членских взносов – в то время для аристократов это было единственным обязательным условием членства в престижном Королевском обществе. Труды общества умалчивают о каком-либо вкладе Камминга

в науку, но ясно, что его контакты с некоторыми его членами были не просто мимолетными: в частности, в 1721 г. – через год после своего избрания – он поставил перед одним из них следующий вопрос из теории вероятностей:

> «*A* и *B*, играя между собой и имея равные Шансы на выигрыш в одной Игре, гарантируют Наблюдателю *S*, что после четного числа *n* сыгранных Игр Победитель передаст ему количество Фишек, равное сумме числа выигранных им игр и половины от общего числа сыгранных игр. Спрашивается, какова Ожидаемая величина *S*».

При путешествии назад сквозь века часто приходится жертвовать ясностью изложения, но эту формулировку можно распутать, рассмотрев многократное подбрасывание симметричной монеты, когда игрок A постоянно ставит на выпадение орла, а игрок B – на выпадение решки. Отклонение от теоретического среднего – половина орлов и половина решек – после четного числа подбрасываний само является случайной величиной, математическое ожидание которой и требуется оценить. Таким образом, мы имеем малознакомый вопрос о хорошо знакомом биномиальном распределении вероятностей. Благодаря таким светилам, как Паскаль, Ферма, Гюйгенс и Якоб I Бернулли, изучение азартных игр дало средства для исследования вопросов только нарождающейся теории вероятностей и статистики, а то, что мы теперь называем испытаниями Бернулли, или биномиальными испытаниями, было сравнительно неплохо изучено и достаточно легко поддавалось анализу. Сегодня эта тема входит в школьный курс теории вероятностей, а самый главный относящийся к ней вопрос – определение вероятности заданного числа успешных исходов при фиксированном числе испытаний. В современной нотации если случайная величина X – число успешных исходов в n повторениях эксперимента, то мы пишем $X \sim B(n, p)$, где p – постоянная вероятность успеха в одном испытании. Хорошо известно, что

$$P(X = r) = \binom{n}{r} p^r (1-p)^{n-r} \quad \text{для } r = 0, 1, 2, \ldots, n,$$

где

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

– элементы знаменитого треугольника Паскаля.

В этом варианте победителем может быть A или B (с одинаковой вероятностью), поэтому ожидаемое обогащение S вычисляется следующим образом:

вероятность, что выиграет A × ожидаемое обогащение, если выиграет A,

+ вероятность, что выиграет B × ожидаемое обогащение, если выиграет B.

Все вероятности равны (и каждая равна ½), и ожидаемые обогащения также равны: если выиграет A, то мы должны вычесть число выигранных им игр из $\frac{1}{2}n$, игнорируя проигрыш B, и наоборот. Поставленный вопрос можно записать одной формулой, воспользовавшись обозначением модуля:

Если $X \sim \text{B}(n, \frac{1}{2})$, то чему равно значение $E[S] = E[|X - \frac{1}{2}n|]$?

Итак, разобравшись с автором вопроса и самим вопросом, который мы переформулировали в современных терминах, неплохо бы узнать, кому был адресован этот вопрос и какой элегантный ответ на него был получен.

8.2. Ответ, но не решение

Членом Королевского общества, которому был адресован вопрос, являлся Абрахам де Муавр, и Камминг не мог бы выбрать более подходящего эксперта. Хотя в современных учебниках его имя чаще всего связывают с важным результатом из повышенного курса школьной математики – *теоремой де Муавра*, согласно которой $(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta$, он также был одним из пионеров теории вероятностей. Его книга «Доктрина случайностей», в особенности ее последнее, третье (посмертное) издание является основополагающей для ранней теории вероятностей и уже сама по себе оправдывает хорошую репутацию де Муавра в ранних исследованиях на эту тему. По мнению авторитетного Исаака Тодхантера (Todhunter 1865, стр. 193),

> «без сомнения, Теория Вероятностей обязана ему больше, чем любому другому математику, за исключением разве что Лапласа».

Хотя подходу де Муавра к задаче Камминга еще только предстояло быть опубликованным во втором и третьем изданиях «Доктрины случайностей», первое письменное упоминание о нем встречается в вышедшем в 1730 г. обзоре его исследований за предыдущие десять лет – 250-страничном труде под названием «Miscellanea Analytica de Seriebus et Quadraturis». В начале главы II книги V мы узнаем о том, что Камминг поставил этот вопрос де Муавру в 1721 г., а в задаче под номером 1 де Муавр дал ее решение, которое мы воспроизводим ниже в более полном и современном виде.

Его результат состоял в том, что

$$E[S] = E[|X - \tfrac{1}{2}n|] = \frac{1}{2^n}\frac{n}{2}\binom{n}{\frac{1}{2}n},$$

и получен он был путем переформулирования вопроса и последующего использования того, что мы теперь называем суммированием последовательностей *методом разностей*. Начнем с того, что

$$\begin{aligned} E[|X - \tfrac{1}{2}n|] &= \sum_{r=0}^{n} |r - \tfrac{1}{2}n| \times P(X = r) \\ &= \sum_{r=0}^{n} |r - \tfrac{1}{2}n| \times \frac{1}{2^n}\binom{n}{r} \\ &= \frac{1}{2^n} \times \sum_{r=0}^{n} |r - \tfrac{1}{2}n| \times \binom{n}{r}. \end{aligned}$$

Для переформулирования вспомним, что n четное и что, в силу знака модуля и симметрии биномиальных коэффициентов, сумму можно представить в виде отдельного среднего члена и удвоенной суммы первой половины членов:

$$E[|X - \tfrac{1}{2}n|] = 0 \times \binom{n}{\frac{1}{2}n} + 2 \times \frac{1}{2^n} \times \sum_{r=0}^{(n/2)-1} (\tfrac{1}{2}n - r) \times \binom{n}{r}$$
$$= 2 \times \frac{1}{2^n} \times \sum_{r=0}^{(n/2)-1} (\tfrac{1}{2}n - r) \times \binom{n}{r}.$$

Результат выглядит не слишком многообещающе, но сумму можно вычислить с помощью многократного применения тождества:

$$(n - r + 1)\binom{n}{r-1} - r\binom{n}{r} = 0$$

для $r = ½n, ½n - 1, ½n - 2, \ldots, 1$, так что в итоге получается

$$(n+2)\binom{n}{\frac{1}{2}n-1} - n\binom{n}{\frac{1}{2}n} = 0,$$
$$(n+4)\binom{n}{\frac{1}{2}n-2} - (n-2)\binom{n}{\frac{1}{2}n-1} = 0,$$
$$(n+6)\binom{n}{\frac{1}{2}n-3} - (n-4)\binom{n}{\frac{1}{2}n-2} = 0,$$
$$\vdots$$
$$(2n-2)\binom{n}{1} - 4\binom{n}{2} = 0,$$
$$2n\binom{n}{0} - 2\binom{n}{1} = 0.$$

Складывая члены, расположенные по диагонали (левый верхний с правым нижним), получаем

$$2\left\{2\binom{n}{\frac{1}{2}n-1} + 4\binom{n}{\frac{1}{2}n-2} + 6\binom{n}{\frac{1}{2}n-3} + \cdots \right.$$
$$\left. + (n-2)\binom{n}{1} + n\binom{n}{0}\right\} - n\binom{n}{\frac{1}{2}n} = 0$$

или

$$\frac{1}{2}n\binom{n}{\frac{1}{2}n} = 2\binom{n}{\frac{1}{2}n-1} + 4\binom{n}{\frac{1}{2}n-2} + 6\binom{n}{\frac{1}{2}n-3}$$
$$+ \cdots + (n-2)\binom{n}{1} + n\binom{n}{0}$$
$$= 2 \times \sum_{r=0}^{(n/2)-1} (\tfrac{1}{2}n - r) \times \binom{n}{r}.$$

Вычисленную таким образом сумму можно подставить в выражение для математического ожидания.

Как бы изящно ни выглядел результат, его применение было ограничено малыми значениями n, потому что для больших n выражение включает произведение очень малого числа $½^n$ на очень большое число

$$\binom{n}{\frac{1}{2}n},$$

которое в то время вычислить было невозможно. По его собственным словам, де Муавр выбрал $n = 10\,000$ и заметил, что вычисление «возможно только путем приложения труда поистине гигантского, а лучше сказать – невозможно». И через несколько страниц читаем (Diaconis and Zabell 1991):

> «Из-за этого человек, которого я похвалил выше, спросил меня, нельзя ли придумать какой-нибудь метод, посредством которого было бы возможно определить этот член бинома, не прибегая к умножению, или, что в конечном итоге привело бы к тому же результату, сложению логарифмов. Я отвечал, что, с его позволения, попробую посмотреть, что можно сделать, в его присутствии, хотя питал мало надежд на успех. Когда он согласился на это, я принялся за работу и в течение одного часа очень близко подошел к решению следующей задачи».

А мы теперь познакомимся с весьма впечатляющими плодами этого часа работы де Муавра.

8.3. Аппроксимация невозможного

Человеком, достойным похвалы, в глазах де Муавра был, конечно, Камминг, которому в 1730 г. предстояло публичное признание; он только что возвратился из путешествия в земли племени чероки в Южной Калифорнии, привез с собой его вождей и был пожалован аудиенцией у короля; впрочем, к тому времени кредиторы уже преследовали его по пятам. А в задаче III из книги V «Miscellanea» де Муавр представил фантастическое утверждение, что для четных n

$$\frac{1}{2^n}\binom{n}{\frac{1}{2}n} \approx 2\frac{21}{125}\left(1-\frac{1}{n}\right)^n \frac{1}{\sqrt{n-1}}.$$

Его доказательство угнездилось среди других аналитических результатов в книге VI, и мы представим его в современной форме. Оно довольно подробное, и в нем используется несколько методов, которые, несмотря на вполне ожидаемую известность де Муавра, были тогда еще совсем новыми. Нам будет удобно считать, что $n = 2m$. Итак, нам нужно найти выражение для

$$
\begin{aligned}
p_m &= \frac{1}{2^{2m}}\binom{2m}{m} \\
&= \frac{1}{2^{2m}}\frac{2m(2m-1)(2m-2)\cdots(2m-(m-2))(2m-(m-1))}{m(m-1)(m-2)\cdots 3.2.1} \\
&= \frac{1}{2^{2m-1}}\frac{(m+1)(m+2)(m+3)\cdots(m+(m-2))(m+(m-1))}{(m-1)(m-2)\cdots 3.2.1} \\
&= \frac{1}{2^{2m-1}}\frac{m+1}{m-1}\frac{m+2}{m-2}\cdots\frac{m+(m-2)}{m-(m-2)}\frac{m+(m-1)}{m-(m-1)} \\
&= \frac{1}{2^{2m-1}}\prod_{i=1}^{m-1}\frac{m+i}{m-i} \\
&= \frac{1}{2^{2m-i}}\prod_{i=1}^{m-2}\frac{m+i}{m-i}\times\frac{2m-1}{1} \\
&= \frac{2m-1}{2^{2m-1}}\prod_{i=1}^{m-2}\frac{1+i/m}{1-i/m}.
\end{aligned}
$$

Кажущееся излишним отщепление последнего члена от произведения окажется важнейшим этапом последующего доказательства сходимости.

Возьмем натуральный логарифм обеих частей:

$$
\begin{aligned}
\ln p_m &= \ln(2m-1) - (2m-1)\ln 2 + \sum_{i=1}^{m-2}\ln\frac{1+i/m}{1-i/m} \\
&= \ln(2m-1) - (2m-1)\ln 2 + 2\sum_{i=1}^{m-2}\sum_{k=1}^{\infty}\frac{1}{2k-1}\left(\frac{i}{m}\right)^{2k-1} \\
&= \ln(2m-1) + (-2m+1)\ln 2 + 2\sum_{k=1}^{\infty}\frac{1}{(2k-1)m^{2k-1}}\sum_{i=1}^{m-2} i^{2k-1},
\end{aligned}
$$

где внутренняя сумма – частный случай логарифмического ряда Ньютона:

$$
\ln\frac{1+x}{1-x} = 2\sum_{k=1}^{\infty}\frac{x^{2k-1}}{2k-1}.
$$

На последнем шаге выделен компонент, являющийся суммой степеней последовательных целых чисел. К нему можно применить знаменитый результат Якоба I Бернулли, представившего явное выражение для этой суммы, которое в наших обозначениях можно записать так:

$$
2k\sum_{i=1}^{m-2} i^{2k-1} = \sum_{j=0}^{2k-1}\binom{2k}{j}B_j(m-1)^{2k-j},
$$

где $B_0 = 1$, $B_1 = -\frac{1}{2}$, $B_2 = \frac{1}{6}$, $B_3 = 0$, $B_4 = -\frac{1}{30}$, ... – *числа Бернулли* (все последующие числа с нечетными номерами равны 0). Подстановка этой суммы в наше выражение дает

$$\begin{aligned}
\ln p_m &= \ln(2m-1) + (-2m+1)\ln 2 \\
&\quad + 2\sum_{k=1}^{\infty} \frac{1}{(2k-1)m^{2k-1}} \left\{ \frac{1}{2k} B_0 (m-1)^{2k} + \frac{B_1}{1!}(m-1)^{2k-1} \right. \\
&\quad\quad + \frac{B_2}{2!}(2k-1)(m-1)^{2k-2} \\
&\quad\quad \left. + \frac{B_4}{4!}(2k-1)(2k-2)(2k-3)(m-1)^{2k-4} + \cdots \right\} \\
&= \ln(2m-1) + (-2m+1)\ln 2 \\
&\quad + 2mB_0 \sum_{k=1}^{\infty} \frac{1}{2k(2k-1)} \left(\frac{m-1}{m}\right)^{2k} \\
&\quad + 2B_1 \sum_{k=1}^{\infty} \frac{1}{2k-1} \left(\frac{m-1}{m}\right)^{2k-1} + \frac{2B_2}{2!m} \sum_{k=1}^{\infty} \left(\frac{m-1}{m}\right)^{2k-2} \\
&\quad + 0 + 2\frac{B_4}{4!}\frac{1}{m^3} \sum_{k=1}^{\infty} (2k-2)(2k-3) \left(\frac{m-1}{m}\right)^{2k-4} + \cdots .
\end{aligned}$$

Упростим выражение, положив $t = (m-1)/m$, и получим

$$\begin{aligned}
\ln p_m &= \ln(2m-1) + (-2m+1)\ln 2 + 2mB_0 \sum_{k=1}^{\infty} \frac{1}{2k(2k-1)} t^{2k} \\
&\quad + 2B_1 \sum_{k=1}^{\infty} \frac{1}{2k-1} t^{2k-1} + \frac{2B_2}{2!m} \sum_{k=1}^{\infty} t^{2k-2} + 0 \\
&\quad + 2\frac{B_4}{4!}\frac{1}{m^3} \sum_{k=1}^{\infty} (2k-2)(2k-3) t^{2k-4} + \cdots .
\end{aligned}$$

Мы будем рассматривать ряды по отдельности, первый из них требует привлечения ряда Ньютона, для чего мы применим интегрирование по частям:

$$\begin{aligned}
\int_0^t \ln\frac{1+x}{1-x}\,\mathrm{d}x &= \left[x \ln\frac{1+x}{1-x}\right]_0^t - \int_0^t x\frac{2}{1-x^2}\,\mathrm{d}x \\
&= t\ln\frac{1+t}{1-t} + \ln(1-t^2) \\
&= (1+t)\ln(1+t) + (1-t)\ln(1-t) \\
&= \int_0^t 2\sum_{k=1}^{\infty} \frac{x^{2k-1}}{2k-1}\,\mathrm{d}x = 2\sum_{k=1}^{\infty} \frac{t^{2k}}{(2k-1)2k},
\end{aligned}$$

а это означает, что

$$2mB_0\sum_{k=1}^{\infty}\frac{t^{2k}}{(2k-1)2k} = m\{(1+t)\ln(1+t)+(1-t)\ln(1-t)\}$$
$$= m\left\{\frac{2m-1}{m}\ln\frac{2m-1}{m}+\frac{1}{m}\ln\frac{1}{m}\right\}$$
$$= (2m-1)\ln\frac{2m-1}{m}-\ln m.$$

Со вторым рядом все просто:

$$2B_1\sum_{k=1}^{\infty}\frac{1}{2k-1}t^{2k-1} = B_1\ln\left(\frac{1+t}{1-t}\right) = -\tfrac{1}{2}\ln(2m-1).$$

С первой из двух оставшихся компонент мы справимся следующим образом:

$$\frac{2B_2}{2!m}\sum_{k=1}^{\infty}t^{2k-2} = \frac{2}{2m}\times\frac{1}{6}\frac{1}{1-t^2}$$
$$= \frac{m^2}{6m(2m-1)} = \frac{m}{6(2m-1)}$$
$$\xrightarrow[m\to\infty]{}\frac{1}{12}.$$

А со второй – так:

$$2\frac{B_4}{4!}\frac{1}{m^3}\sum_{k=1}^{\infty}(2k-2)(2k-3)t^{2k-4}$$
$$= \frac{2}{4!}\times-\frac{1}{30}\times\frac{1}{m^3}\sum_{k=1}^{\infty}\frac{\mathrm{d}^2}{\mathrm{d}t^2}t^{2k-2}$$
$$= -\frac{1}{360m^3}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\sum_{k=1}^{\infty}t^{2k-2} = -\frac{1}{360m^3}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\frac{1}{1-t^2}$$
$$= -\frac{1}{360m^3}\frac{2+6t^2}{(1-t^2)^3} = -\frac{1}{360m^3}m^4\frac{2m^2+6(m-1)^2}{(2m-1)^3}$$
$$= -\frac{1}{360}\frac{m(2m^2+6(m-1)^2)}{(2m-1)^3}$$
$$\xrightarrow[m\to\infty]{}-\frac{1}{360}.$$

Мы не будем продолжать (хотя де Муавр включил еще два члена), а напишем

$$\begin{aligned}
\ln p_m &= \ln(2m-1) + (-2m+1)\ln 2 + (2m-1)\ln\frac{2m-1}{m} - \ln m \\
&\quad - \tfrac{1}{2}\ln(2m-1) + \frac{1}{12} - \frac{1}{360} + \left(\frac{1}{1260} - \frac{1}{1680} + \cdots\right) \\
&= \ln(2m-1) - 2m\ln 2 + \ln 2 + (2m-1)\ln(2m-1) \\
&\quad - (2m-1)\ln m - \ln m - \tfrac{1}{2}\ln(2m-1) \\
&\quad + \frac{1}{12} - \frac{1}{360} + \left(\frac{1}{1260} - \frac{1}{1680} + \cdots\right) \\
&= -\tfrac{1}{2}\ln(2m-1) + 2m\ln(2m-1) - 2m\ln 2 + \ln 2 - 2m\ln m \\
&\quad + \frac{1}{12} - \frac{1}{360} + \left(\frac{1}{1260} - \frac{1}{1680} + \cdots\right) \\
&= (2m - \tfrac{1}{2})\ln(2m-1) - 2m\ln 2m + \ln 2 \\
&\quad + \frac{1}{12} - \frac{1}{360} + \left(\frac{1}{1260} - \frac{1}{1680} + \cdots\right).
\end{aligned}$$

Запись нелогарифмической части в виде логарифма дает

$$\begin{aligned}
\ln p_m &= (2m - \tfrac{1}{2})\ln(2m-1) - 2m\ln 2m + \ln 2 \\
&\quad + \frac{1}{12} - \frac{1}{360} + \frac{1}{1260} - \frac{1}{1680} + \cdots \\
&= (2m - \tfrac{1}{2})\ln(2m-1) - 2m\ln 2m + \ln 2 + \ln A,
\end{aligned}$$

где $\ln A = \frac{1}{12} - \frac{1}{360} + \frac{1}{1260} - \frac{1}{1680} + \cdots$, и потому $A = \exp(\frac{1}{12} - \frac{1}{360} + \frac{1}{1260} - \frac{1}{1680} + \cdots)$.

Потенцируем обе части и получаем:

$$\begin{aligned}
p_m &= 2A(2m-1)^{2m-1/2}(2m)^{-2m} \\
&= \frac{2A}{\sqrt{n-1}}\left(1 - \frac{1}{n}\right)^n \approx 2\frac{21}{125}\frac{1}{\sqrt{n-1}}\left(1 - \frac{1}{n}\right)^n,
\end{aligned}$$

где рациональное приближение получено путем использования все более точных оценок $2A$: 2.1738, 2.1676, 2.1695 и 2.1682, получаемых путем включения дополнительных членов ряда. Де Муавр остановился на числе $2.168 = 2\,^{21}/_{125}$.

Затем он вычислил выражение, аппроксимирующее вероятности исходов, далеких от центрального, точнее, находящихся на расстоянии l от него. Опираясь на описанный выше метод, он пришел к удивительно симметричному выражению:

$$\ln\frac{p_{m+l}}{p_m} = (m + l - \tfrac{1}{2})\ln(m + l - 1) + (m - l + \tfrac{1}{2})\ln(m - l + 1),$$

вывод которого мы оставляем читателю. Но, несмотря на красоту, это выражение нам не понадобится, потому что если рассматривать только первый член каждого разложения, то имеем по определению:

$$
\begin{aligned}
&\ln \frac{p_{m+l}}{p_m} \\
&= \ln \frac{1}{2^{2m}} \binom{2m}{m+l} \times \frac{2^{2m}}{\binom{2m}{m}} = \ln \frac{(2m)!}{(m+l)!(m-l)!} \times \frac{m!m!}{(2m)!} \\
&= \ln \frac{m!m!}{(m+l)!(m-l)!} = \ln \frac{m(m-1)\cdots(m-l+1)}{(m+l)(m+l-1)\cdots(m+1)} \\
&= -\ln \frac{(m+l)(m+l-1)\cdots(m+1)}{m(m-1)\cdots(m-l+1)} \\
&= -\ln \frac{m+l}{m}\frac{m+1}{m-1}\cdots\frac{m+l-1}{m-l+1} \\
&= -\ln \left(1+\frac{l}{m}\right)\left(\frac{1+1/m}{1-1/m}\right)\left(\frac{1+2/m}{1-2/m}\right)\cdots\left(\frac{1+(l-1)/m}{1-(l-1)/m}\right) \\
&= -\left\{\left(\frac{l}{m}+\cdots\right)+2\left(\frac{1}{m}+\cdots\right)+2\left(\frac{3}{m}+\cdots\right)\cdots 2\left(\frac{l-1}{m}+\cdots\right)\right\} \\
&\approx -\left\{\frac{l}{m}+\frac{2}{m}(1+2+3+\cdots+(l-1))\right\} \\
&= -\left\{\frac{l}{m}+\frac{2}{m}\frac{l}{2}(l-1)\right\} \\
&= -\frac{l^2}{m} = -\frac{2l^2}{n}.
\end{aligned}
$$

Все это означает, что

$$p_{m+l} \approx p_m \mathrm{e}^{-2l^2/n}.$$

И у нас появляется намек на функцию e^{-x^2}, определяющую форму нормальной кривой. Намек становится еще более явственным, если заменить исходное выражение p_m чем-то более естественным, а для этого мы перенесемся в 12 ноября 1733 г. – день, когда де Муавр представил группе друзей статью всего на семи страницах; называлась она «Approximato as Summam Terminorum Binomii $(a+b)^n$ in Series Expansi». Это очень редкое издание, до нас дошло всего две копии, и, как следует из названия, она посвящена той же самой проблеме и приводит в порядок выражение, выведенное им в 1730 г.; материал был повторен во втором и третьем издании «Доктрины случайностей». В поправке использовалось представление постоянной e в виде предела, а A было заменено на B, где $AB = e$, и в итоге получилось

$$
\begin{aligned}
p_m &= \frac{2\mathrm{e}}{B\sqrt{n-1}}\left(1-\frac{1}{n}\right)^n \\
&\approx \frac{2\mathrm{e}}{B\sqrt{n}} \times \mathrm{e}^{-1} = \frac{2}{B\sqrt{n}},
\end{aligned}
$$

где $\ln B = 1 - \frac{1}{12} + \frac{1}{360} - \frac{1}{1260} + \cdots$, в чем, как заметил де Муавр, его «ученый друг Джеймс Стирлинг» узнал $B = \sqrt{2\pi}$. Тогда

$$p_m \approx \frac{2}{\sqrt{2\pi n}} \quad \text{и} \quad p_{m+l} \approx \frac{2}{\sqrt{2\pi n}} \mathrm{e}^{-2l^2/n}.$$

И наконец, вернемся к *Miscellanea*:

> «Если члены бинома поместить по вертикали, на равном расстоянии друг от друга под прямым углом к прямой линии и над ней, то крайние члены будут располагаться на некоторой кривой. Так построенная кривая имеет две точки перегиба, по одной с каждой стороны от максимального члена».

Он выяснил, что эти точки перегиба находятся (приближенно и для больших n) там, где $l = \pm\frac{1}{2}\sqrt{n}$, и по ходу дела нашел естественную меру дисперсии случайной величины; то, что он называл ее *модулем*, определяется как $\sqrt{n}$. Если мы возьмем стандартное нормальное приближение к симметричной биномиальной случайной величине, то будем иметь распределение $N(\frac{1}{2}n, \frac{1}{4}n)$ со стандартным отклонением $\frac{1}{2}\sqrt{n}$: его модуль является вариантом нашего стандартного отклонения. Оставался неразрешенным важный вопрос: применить связь между дискретным и непрерывным к выполнению невозможных никаким иным способом вычислений, например:

$$P(\tfrac{1}{2}n - d \leqslant X \leqslant \tfrac{1}{2}n + d) \approx \sum_{l=-d}^{d} \frac{2}{\sqrt{2\pi n}} \mathrm{e}^{-2l^2/n} \approx \frac{4}{\sqrt{2\pi n}} \sum_{l=0}^{d} \mathrm{e}^{-2l^2/n}$$

с помощью приближения

$$\frac{4}{\sqrt{2\pi}} \int_0^{d/\sqrt{n}} \mathrm{e}^{-2y^2}\, \mathrm{d}y,$$

получающегося, если положить $y^2 = l^2/n$ в сумме. Конечно, эту экспоненциальную функцию невозможно проинтегрировать в конечной форме, поэтому перед де Муавром (и сегодня перед нами) открывалось два пути: использовать разложение в ряд Тейлора и интегрировать почленно или воспользоваться приближенной формулой площади под кривой. Последний подход он назвал «Искусством механических квадратур, впервые изобретенным сэром Исааком Ньютоном», а выбранный им метод мы сегодня назвали бы *правилом ⅜ Симпсона* (если оно вообще как-то называется):

$$\int_a^b f(x)\, \mathrm{d}x \approx \tfrac{3}{8}h\{f(x_0) + f(x_1) + f(x_2) + f(x_3)\}, \text{ где } h = \frac{b-a}{3}.$$

Имея эти формулы, де Муавр выполнил вычисления для $n = 3600$ и $l = \frac{1}{2}\sqrt{n}$ и показал, что если $X \sim B(3600, \frac{1}{2})$, то

$$P(1800 - \tfrac{1}{2}\sqrt{3600} \leqslant X \leqslant 1800 + \tfrac{1}{2}\sqrt{3600}) = P(1770 \leqslant X \leqslant 1830) \approx 0.682688.$$

То есть 68 % распределения лежит на расстоянии не более одного стандартного отклонения от среднего; он также вычислил, что 0.95428 = 95 % лежит на расстоянии не более $l = \pm 2\frac{1}{2} \times \sqrt{n}$ (два стандартных отклонения) от среднего

и что 0.99874 > 99 % лежит на расстоянии не более $l = \pm 3½ \times \sqrt{n}$ (три стандартных отклонения) от среднего. Мы полагаем, что читатель впечатлен точностью этих вычислений и их провидческой природой.

В результате этих вычислений стала понятна важность кривой с уравнением e^{-x^2}, а вместе с этим началось восхождение нормальной кривой на роль доминирующей статистической кривой. Второй этап этого эволюционного процесса был совершенно иной.

8.4. Кривые ошибок

Имея набор измерений, собранных для приближения какого-то значения, мы, естественно, начинаем думать, как лучше ими воспользоваться. Нужно ли выбрать одно, которое, в силу условий измерения, с максимальной вероятностью является наиболее точным, а значит, «лучшим» приближением к неизвестному значению? Или нужно считать, что все измерения равноправны, и взять их медиану или если имеются повторения, то моду? Или нужно вычислить среднее? Или мы остановимся на каком-то процессе, который подразумевает применение сразу нескольких описанных выше идей? Или выберем что-то совсем другое? Такие вопросы волновали ученых с того времени, как наука получила право на такое название, а первые серьезные попытки разрешить эту трудную загадку были предприняты астрономами – первыми учеными, для которых повторяющиеся точные измерения имели важнейшее значение. Начиная с античных времен и до Кеплера и Тихо Браге ученые применяли различные подходы – смесь разных казавшихся разумными методов, – но только Галилей первым наметил систематический подход к изучению ошибок. В своем «Диалоге» (к которому мы еще вернемся в следующей главе) он сформулировал следующие разумные предположения.

- Существует только одно число, равное расстоянию от звезды до центра Земли: истинное расстояние.
- Все наблюдения подвержены ошибкам, связанным с наблюдателем, приборами и другими условиями наблюдений.
- Наблюдения распределены симметрично относительно истинного значения, т. е. ошибки распределены симметрично относительно нуля.
- Малые ошибки встречаются чаще, чем большие.
- Вычисленное расстояние есть функция от прямых угловых наблюдений такая, что малые изменения наблюдений могут привести к большим изменениям расстояния.

На этом минимальном фундаменте было возведено величественное здание нынешней обстоятельной теории ошибок, которая и составляет второй этап эволюции нормальной кривой. В общем виде фундаментальный вопрос звучит так: если из данных наблюдений выведено некоторое число, то с какой степенью уверенности мы можем использовать его в качестве истинного значения измеряемой величины? В количественном выражении нас интересует, какова вероятность, что отклонение выбранной оценки от истинного значения измеряемой величины находится в приемлемых пределах? И следует ли в качестве оценки брать среднее результатов измерений?

> «Вашей светлости хорошо известно, что метод, применяемый астрономами с целью уменьшить ошибки вследствие несовершенства инструментов и органов чувств, заключающийся в вычислении Среднего нескольких наблюдений, не был принят в общем, но что некоторые персоны, пользующиеся большим авторитетом, придерживаются и даже публично выражают мнение о том, что на одно наблюдение, выбранное с должным тщанием, можно полагаться так же, как на Среднее многих наблюдений».

Так писал английский математик Томас Симпсон (автор правила Симпсона) тогдашнему президенту Королевского общества в 1755 г., и так думал, например, весьма влиятельный Роберт Бойль (автор закона Бойля). Именно в этом письме Симпсон высказался в пользу среднего нескольких «наблюдений», принимающего во внимание их относительную частоту. Приведенный им пример был искусственным, чтобы не слишком усложнять математику; он взял дискретное распределение, в котором ошибки принимали значения $-v, \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots, v$ для некоторого v, сначала с вероятностями, пропорциональными $r^{-v}, \ldots, r^{-3}, r^{-2}, r^{-1}, r^0, r^1, r^2, r^3, \ldots, r^v$, а затем пропорциональными $r^{-v}, 2r^{1-v}, 3r^{2-v}, \ldots, (v+1)r^0, \ldots, 3r^{v-2}, 2r^{v-1}, r^v$ для некоторого r. Точки на рис. 8.1 показывают это распределение ошибок во втором случае, когда $r = 1$. Эта структура позволила ему использовать геометрическую прогрессию, а с ее помощью он смог вывести выражения для вероятности того, что среднее наблюдений будет отклоняться от истинного измеряемого значения не более чем на заданную величину. Отсюда он построил пример, в котором эта вероятность была больше, чем соответствующая вероятность одного наблюдения. Идея была сформулирована, но впоследствии он пошел дальше. Во второй статье, написанной всего двумя годами позже[1], его отношение к приемлемости среднего изменилось. Он начинает словами:

> «Хотя метод, применяемый Астрономами для уменьшения ошибок вследствие несовершенства инструментов и органов чувств, заключающийся в вычислении среднего нескольких наблюдений, весьма полезен и используется почти универсально, тем не менее, насколько мне известно, до сих пор не было никаких попыток обосновать его».

Далее следует перефразирование предыдущей статьи, но с одним важным дополнением; для него он поместил рисунок, из которого вывел непрерывную кривую, изображенную на рис. 8.1, сопроводив это следующими словами:

> «Но теперь я покажу, как можно вычислить вероятность, когда ошибка принимает любое значение, целое или дробное… Пусть прямая AB представляет весь интервал, в который предположительно должны попадать наблюдения, и представим себе, что он разбит на очень большое число очень малых частей».

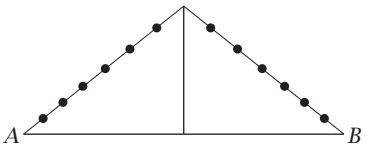


Рис. 8.1. Распределение ошибок Симпсона с $r = 1$

[1] Датирована 1757 г. и вошла в его сборник «Miscellaneous Tracts».

И далее он ни больше ни меньше как построил первое непрерывное распределение ошибок, но эстафета была еще далека от завершения. Палочку принял Лагранж в своем первом мемуаре, относящемся к вероятности, который был опубликован в пятом томе его трудов «Miscellanea Taurinesia», охватывающем период с 1770 по 1773 г. Первые десять задач, составляющих эту работу, сильно напоминают работу Симпсона, поскольку в них рассматриваются удобно сконструированные дискретные распределения ошибок, но в последней части он переходит к непрерывным распределениям. Кульминацией является обсуждение двух распределений ошибок с функциями плотности вероятности $\varphi(x) = K\sqrt{c^2 - x^2}$ для $-c \leqslant x \leqslant c$ и $\varphi(x) = K \cos x$ для $-\frac{1}{2}\pi \leqslant x \leqslant \frac{1}{2}\pi$. Квадратура обеих функций была по силам математическому анализу. Таким образом, обработка ошибок перешла в область непрерывного, но еще была далека от практического применения. Как для комбинации человеческих и инструментальных ошибок определить естественное распределение? Какова должна быть его форма? Должна ли она меняться в зависимости от обстоятельств?

Первая основательная попытка дать ответ на эти вопросы выпала на долю Пьера-Симона Лапласа с его *первым распределением ошибок*. Хотя разрабатывать этот подход он начал в 1772 г., его отчетливые очертания видны только в задаче 3 из статьи 1774 г. (Laplace 1986), где он хочет получить наилучшую оценку неизвестной величины всего по трем наблюдениям. Его подход можно кратко описать следующим образом.

На рис. 8.2 истинное значение V оценивается по трем наблюдениям a, b, c; положив $p = b - a$ и $q = c - b$, мы можем обозначить ошибки x, y, z, где $y = p - x$ и $z = p + q - x$; эти ошибки подчиняются неизвестному распределению $\varphi(x)$. В предположении, что измерения независимы, вероятность наблюдения того, что произошло в действительности, записывается как произведение $\varphi(x)\varphi(p - x)\varphi(p + q - x)$, и мы ищем значение x (и значит, и V), которое его максимизирует. Но, как замечает Лаплас, «необходимо знать $\varphi(x)$. Но при бесконечном числе возможных функций как выбрать наилучшую?».

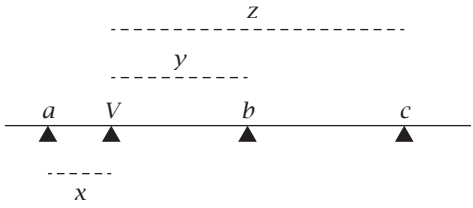


Рис. 8.2. Подход Лапласа к оцениванию

Какова бы ни была ее форма, кривая ошибок должна являться следствием множества разумных предположений, стандарт которых задал Галилей: симметрия относительно 0 и убывание по обе стороны от него. Общая площадь под кривой должна быть равна 1. Какое еще условие добавить, чтобы можно было выстроить рассуждение, ведущее к естественной кривой? Для этого Лаплас выдвинул *принцип безразличия*, или *принцип недостаточного обоснования*:

«Таким образом, не только ординаты точек на кривой, но и разности этих ординат должны убывать по мере удаления... А поскольку у нас нет причин предполагать для ординат иной закон, нежели для их разностей, отсюда следует, что мы должны, сообразуясь с правилами вероятностей, предположить, что отношение двух бесконечно малых соседних разностей равно отношению соответствующих ординат. Стало быть, имеем

$$\frac{\mathrm{d}\varphi(x+\mathrm{d}x)}{\mathrm{d}\varphi(x)} = \frac{\varphi(x+\mathrm{d}x)}{\varphi(x)},$$

откуда

$$\frac{\mathrm{d}\varphi(x)}{\mathrm{d}x} = -m\varphi(x) \quad \text{что дает} \quad \varphi(x) = \tfrac{1}{2}m\mathrm{e}^{-mx}.$$

Мы можем заполнить незначительные пробелы в его рассуждениях с помощью рис. 8.3, на котором изображены две касательные в близких точках, так что $\mathrm{d}\varphi = (\mathrm{d}\varphi/\mathrm{d}x)\,\mathrm{d}x = \varphi'(x)\,\mathrm{d}x$ в любой точке, и, следовательно, его принцип безразличия дает

$$\frac{\mathrm{d}\varphi(x+\mathrm{d}x)}{\mathrm{d}\varphi(x)} = \frac{\varphi(x+\mathrm{d}x)}{\varphi(x)} \to \frac{\varphi'(x+\mathrm{d}x)\,\mathrm{d}x}{\varphi'(x)\,\mathrm{d}x} = \frac{\varphi(x+\mathrm{d}x)}{\varphi(x)}$$
$$\to \frac{\varphi'(x)}{\varphi(x)} = \frac{\varphi'(x+\mathrm{d}x)}{\varphi(x+\mathrm{d}x)}.$$

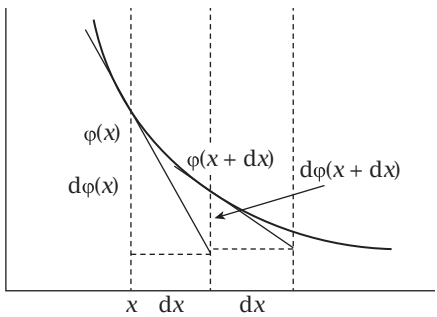


Рис. 8.3. Принцип безразличия Лапласа

Это означает, что $\varphi'(x)/\varphi(x) = -m$, постоянной, и $\varphi(x) = Ae^{-mx}$. Из требований симметрии и равенства 1 площади под кривой вытекает, что

$$\frac{1}{2} = \int_0^\infty A\mathrm{e}^{-mx}\,\mathrm{d}x = -\frac{A}{m}[\mathrm{e}^{-mx}]_0^\infty = \frac{A}{m},$$

и значит, so $\varphi(x) = \tfrac{1}{2}me^{-mx}$. Это уравнение правой части кривой, и значит, снова в силу симметрии ее продолжение должно описываться уравнением $\varphi(x) = \tfrac{1}{2}me^{-m|x|}$, а вся кривая показана на рис. 8.4.

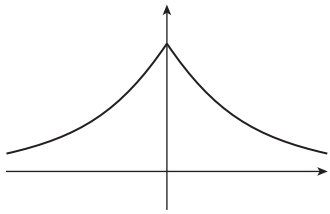


Рис. 8.4. Распределение Лапласа

Эта кривая существует по сей день и описывает *распределение Лапласа*. С помощью подручных инструментов он сумел вывести выражение

$$V \sim p + \frac{1}{m}\ln(1 + \tfrac{1}{3}\mathrm{e}^{-mp} - \tfrac{1}{3}\mathrm{e}^{-mq})$$

в предположении, что $p > q$, где p, q и неизвестная величина m взаимозависимы. Но даже потрясающей способности Лапласа к манипулированию символами оказалось недостаточно для применения его методов к сколь угодно большим наборам данных, и в 1777 г. он предпринял второй подход к кривой ошибок, применив рассуждения, куда более сложные, чем в описанном выше. На самом деле большая часть его математических работ вплоть до 1781 г. была посвящена развитию этих идей. Мы не можем поведать полную историю в этом кратком очерке, а ограничимся конечным выводом, что новая кривая ошибок должна иметь уравнение

$$\varphi(x) = \frac{1}{2a}\ln\frac{a}{|x|},$$

где $-a \leqslant x \leqslant a$ ограничивают величину ошибок. Ее график показан на рис. 8.5, а связанные с ней трудности не устраняются наличием вертикальной асимптоты, проходящей через 0.

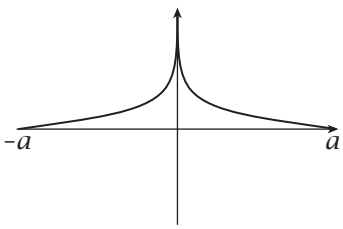


Рис. 8.5. Второе распределение Лапласа

Несмотря на всю свою сложность, эта попытка стала шагом назад в эволюции кривой ошибок, а финал потребовал от Лапласа еще больших мыслительных усилий – и вмешательства, пожалуй, величайшего из работавших тогда математиков.

8.5. Настоящая кривая ошибок

Этим математиком был несравненный Карл Фридрих Гаусс, но, прежде чем переходить к его вкладу, выведем нормальную кривую как естественную кривую ошибок в контексте, который, как мы надеемся, является естественным для нее окружением.

Рассмотрим многократные попытки поразить цель 0 на рис. 8.6. Безуспешные попытки иллюстрируются отклонениями по горизонтали x и отклонениями по вертикали y. Эти ошибки подчиняются непрерывному распределению вероятностей $\varphi(x)$ и предположительно одинаковому с ним распределению $\varphi(y)$.

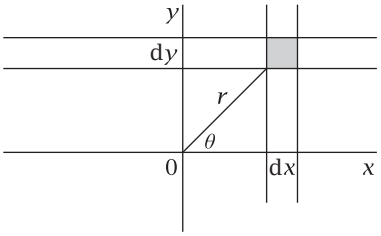


Рис. 8.6. Оценка ошибки

Сформулируем три естественных предположения:

- ошибки в двух взаимно перпендикулярных направлениях независимы;
- ошибки не зависят от ориентации;
- величины ошибок убывают по мере удаления от начала координат.

Вероятность, что ошибка по вертикали лежит в бесконечно узкой вертикальной полосе, равна $\varphi(x)\mathrm{d}x$, и аналогично вероятность, что ошибка по вертикали лежит в бесконечно узкой горизонтальной полосе, равна $\varphi(y)\mathrm{d}y$. В силу предположения о независимости вероятность, что попытка поразить мишень окажется в заштрихованном прямоугольнике, равна $\varphi(x)\mathrm{d}x \times \varphi(y)\mathrm{d}y$. Но она также не зависит от ориентации, поэтому ее можно записать в виде $g(r)\mathrm{d}x\mathrm{d}y$, откуда следует, что $g(r) = \varphi(x)\varphi(y)$. Имеем, следовательно, $\varphi(x)\varphi(y) = g(\sqrt{x^2 + y^2})$. Полагая $y = 0$, получаем $\varphi(x)\varphi(0) = g(x)$, откуда $\varphi(x)\varphi(y) = \varphi(\sqrt{x^2 + y^2})\varphi(0)$, что можно переписать в виде

$$\frac{\varphi(x)}{\varphi(0)} \times \frac{\varphi(y)}{\varphi(0)} = \frac{\varphi(\sqrt{x^2 + y^2})}{\varphi(0)}.$$

Положив $\mathrm{P}(x) = \ln(\varphi(x)\varphi(0))$, будем иметь $P(x) + P(y) = P(\sqrt{x^2 + y^2})$ – функциональное уравнение для $P(x)$, определяющее ее с точностью до постоянной, хотя его форма и не вселяет оптимизма. Но его можно улучшить, определив еще одну функцию $Q(x) = P(\sqrt{x})$, в результате чего получится новое функциональное уравнение $Q(x) + Q(y) = P(\sqrt{x}) + P(\sqrt{y}) = P(\sqrt{x + y}) = Q(x + y)$. Следовательно, $Q(x)$ – линейная функция и должна иметь вид $Q(x) = kx$ (так как $Q(0) = 0$). Более строго: по индукции получаем $Q(nx) = nQ(x)$ и далее рассуждаем следующим образом (вспомним, что $\varphi(x)$ и, следовательно, $P(x)$ и $Q(x)$ непрерывны):

- $x = 1 \to Q(n) = nQ(1) = kn$ и, следовательно, $Q(x) = kx$ для $x \in \mathbb{N}$;
- $x = \frac{p}{q} \to qQ\left(\frac{p}{q}\right) = Q\left(q \times \frac{p}{q}\right) = Q(p) = kp \to Q\left(\frac{p}{q}\right) = k\frac{p}{q}$

  и, следовательно, $Q(x) = kx$ для $x \in \mathbb{Q}^+$;
- если $x \in \mathbb{R}^+$, то пусть $\{x_i\}$ – последовательность рациональных чисел такая, что $x_i \to x$ при $i \to \infty$. Тогда, поскольку $Q(x)$ непрерывна,

$$Q(x) = Q\left(\lim_{i\to\infty} x_i\right) = \lim_{i\to\infty} Q(x_i) = \lim_{i\to\infty} kx_i = k\lim_{i\to\infty} x_i = kx.$$

И мы заключаем, что $Q(x) = kx$ для всех положительных x, а в силу симметрии – вообще для всех x. Таким образом, $P(\sqrt{x}) = kx$, откуда следует, что $P(x) = kx^2$, где k – постоянная. Тогда $\varphi(x) = p(0)e^{kx^2}$, а кроме того, поскольку ошибки уменьшаются с увеличением x, $k < 0$, мы можем написать, что $\varphi(x) = Ae^{-kx^2/2}$. Зная, что полная вероятность равна 1, и применяя стандартные приемы, находим $A = \sqrt{k/2\pi}$, откуда $\varphi(x) = \sqrt{k/2\pi}e^{-kx^2/2}$, и мы нашли форму нормальной кривой в контексте обработки ошибок. Но это не то воплощение, в котором она явилась впервые.

1 января 1801 г., работая в своей обсерватории в Палермо, итальянский священнослужитель и авторитетный астроном Джузеппе Пьяцци обнаружил новое небесное тело в созвездии Тельца – и оно двигалось. Это была точно не звезда, возможно, комета, а быть может, новая планета между Марсом и Юпитером, существование которой давно подозревали, потому что его предсказывало правило Тициуса–Боде. Пьяцци назвал объект *Цererой Фердинанда* в честь римской богини земледелия Цереры и короля Сицилии Фердинанда, но вскоре по политическим причинам упоминание короля было убрано из названия, и осталось просто Церера. Пьяцци выполнил девятнадцать полных измерений, последнее 11 февраля, после чего объект скрылся за Солнцем. Отчасти для того, чтобы вы могли оценить точность, в табл. 8.1. приведены основные элементы первого, среднего и последнего измерений.

Таблица 8.1. Измерения положения Цереры, выполненные Пьяцци

1801	Долгота	Широта
1 января	53°23′06:38″	3°06′45:16″
22 января	53°39′11:58″	1°42′28:80″
11 февраля	56°26′28:20′	0°35′55:02″

Бросается в глаза отсутствие расстояний от объекта до Земли или Солнца. Очевидная проблема, которую лихорадочно пытались решить многие выдающиеся астрономы того времени, – предсказать, появится ли снова небесная сфера Цереры и, стало быть, можно ли будет продолжить изучение элементов ее орбиты. Астрономы делали предсказания разного качества, но Церера упрямо оставалась «ненаблюдаемым бродягой». Эти слова принадлежат Гауссу, которому в 1801 г. было 24 года. В сентябре этого года он увидел эти данные, а к ноябрю закончил первое вычисление орбиты Цереры, которое опублико-

вал в декабрьском номере «Monatliche Correspondenz» – ежемесячного журнала по астрономии. 7 декабря основатель журнала, венгерский астроном Франц Ксавер фон Цах, вновь наблюдал Цереру точно в том месте, которое предсказал Гаусс, – совсем не там, где предсказывали многие другие. 31 декабря немецкий астроном Вильгельм Ольберс подтвердил этот результат собственными наблюдениями, согласующимися с предсказанием Гаусса. 1801 г. стал *annus mirabilis*[1] для Гаусса; много лет спустя он писал, что в тот год ему в голову приходило так много важных идей, что он не успевал их толком обдумывать. Вышла его основополагающая работа по теории чисел, «Disquisitiones Arithmeticae», где был доказан знаменитый квадратичный закон взаимности; также был опубликован получивший широчайшую известность результат о возможности построения правильного 17-угольника циркулем и линейкой – и это далеко не все. После этого последнего успеха он стремительно ворвался в элитарную группу гениев, и этот эпитет заслуженно сопровождает его имя и по сей день. Именно на основе трех наблюдений в табл. 8.1 Гаусс произвел свои вычисления, пользуясь законами Кеплера и сферической тригонометрией; его вычисления – убедительный пример различия между «элементарным» и «простым» в математике[2]: эту работу можно классифицировать как *элементарную*, но вряд ли кто-то рискнет назвать ее *простой*. По счастью, ее детали нам не важны; заметим лишь, что, сделав начальное предсказание орбиты по трем наблюдениям, он затем использовал остальные наблюдения, чтобы уточнить его. Точного совпадения он не достиг, но улучшил результаты – как укладчик ковров подгоняет положение ковра под размеры комнаты. После обнаружения и последующего прослеживания орбиты Цереры Гаусс получил дополнительные данные для оценки и воспользовался ими для уточнения своих методов:

> «Но когда в нашем распоряжении имеется более длинный ряд наблюдений, охватывающий несколько лет, на их основе можно вывести больше нормальных положений; по этой причине мы не можем обеспечить наивысшую точность, если вынуждены выбирать всего три или четыре положения для определения орбиты, пренебрегая всеми остальными. Но в таком случае если ставится задача достичь наибольшей точности, то мы должны собрать и использовать возможно большее число точных положений. Тогда, конечно, данных будет больше, чем необходимо для определения неизвестных величин, но все эти данные подвержены ошибкам, пусть и небольшим, поэтому в общем случае идеально удовлетворить им всем будет невозможно. Нет никакой причины, по которой мы должны были бы среди этих данных рассматривать какие-то шесть как абсолютно точные; напротив, исходя из принципов вероятности, мы должны предполагать, что большие или меньшие ошибки могут присутствовать во всех данных без разбора; при этом, вообще говоря, меньшие ошибки встречаются чаще, чем большие. Очевидно, что орбита, которая точно удовлетворяет шести наблюдениям, больше или меньше отклоняясь от остальных, должна рассматриваться как менее согласная с принципами исчисления вероятностей, чем та, которая, немного отличаясь от этих шести наблюдений, гораздо лучше согласуется с остальными».

---

[1] Год чудес (*лат.*).

[2] Элементарный – не требующий солидных математических познаний для чтения работы. Простой: для понимания не нужно больших математических способностей.

Точку зрения Гаусса можно пояснить на примере трех элементов данных, показанных на рис. 8.7. Мы будем предполагать, что «орбита» объекта является прямой линией и что на график нанесены три измеренных положения. Какую прямую принять за орбиту: AB, BC, AC или ни одну из них? Ответ, конечно, – ни одну, потому что третья точка не должна оказывать никакого влияния. Гораздо более предпочтительно взять прямую, на которую надлежащее влияние оказывают все точки, пусть даже она не проходит ни через одну из них. Но что понимать под «надлежащим»?

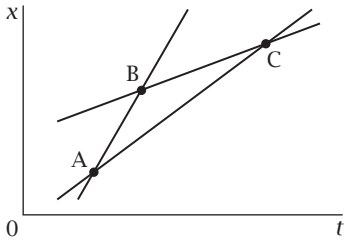


Рис. 8.7. Линия наилучшего соответствия

Предположим, что имеется три точки с координатами

$$\{(t_1, x_1), (t_2, x_2), (t_3, x_3)\},$$

где $\{t_i\}$ – время, которое предполагается измеренным с высокой точностью, а $\{x_i\}$ – положения, при измерении которых возможны ошибки наблюдения. Уравнение линейной орбиты имело бы вид $x = a + bt$ с ошибками:

$$\{(a + bt_1 - x_1), (a + bt_2 - x_2), (a + bt_3 - x_3)\},$$

и мы ищем такие значения параметров, которые максимизируют вероятности этих трех ошибок, а это значит, что прежде всего нам нужно знать их распределение. И Гаусс поддержал использование *среднего* как наиболее репрезентативной статистики:

> «Гипотеза, которую обыкновенно считают аксиомой и заключающаяся в том, что если некоторая величина определена посредством нескольких прямых наблюдений, произведенных при схожих условиях и с одинаковой тщательностью, то среднее арифметическое всех наблюдений – наиболее вероятное значение, является если не абсолютно строгой, то, по крайней мере, близка к этому, и было бы безопаснее всего придерживаться ее».

С этим убеждением он представил первое доказательство того, что нормальная кривая является естественным выбором для кривой ошибок, пусть и предварительным, зато полезным. Перефразировать это можно следующим образом. Предположим, что истинное значение, которое мы пытаемся измерить, но никогда не будем знать точно, равно v, и для его оценки мы производим n измерений $x_1, x_2, x_3, \ldots x_n$. Тогда соответствующие ошибки измерения равны $(x_1 - v)$, $(x_2 - v)$, $(x_3 - v)$, …, $(x_n - v)$, и мы предполагаем, что они подчиняются неизвестному распределению, определяемому дифференцируемой функцией $\varphi(x)$. Исходя из этого, строим так называемую *функцию правдоподобия* (вслед за Лапласом):

$$\Omega = \varphi(x_1 - v)\varphi(x_2 - v)\varphi(x_3 - v)\cdots\varphi(x_n - v) = \prod_{i=1}^{n}\varphi(x_i - v)$$

в виде функции от v, которая отражает вероятность независимого возникновения такого множества значений. Мы, конечно, попытаемся установить вид функции $\varphi(x)$ и сделаем обычные предположения в этом направлении: небольшие ошибки вероятнее больших, так что значение $\varphi(0)$ доставляет максимум; ошибки по обе стороны истинного значения равновероятны, так что $\varphi(-x) = \varphi(x)$; и, так как должны быть учтены все возможности, полная площадь под графиком $\varphi(x)$ равна 1.

Предположение о том, что при наличии нескольких измерений среднее является наиболее вероятным значением измеряемой величины, интерпретировалось в том смысле, что Ω достигает максимума, когда

$$v = \bar{v} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Следующий естественный шаг – продифференцировать по v и приравнять производную нулю. Для этого проще всего перейти к логарифмам:

$$\ln\Omega = \ln\prod_{i=1}^{n}\varphi(x_i - v) = \sum_{i=1}^{n}\ln\varphi(x_i - v)$$

и, следовательно,

$$\frac{1}{\Omega}\frac{\mathrm{d}\Omega}{\mathrm{d}p} = \sum_{i=1}^{n}\frac{-\varphi'(x_i - v)}{\varphi(x_i - v)},$$

откуда

$$\sum_{i=1}^{n}\frac{\varphi'(x_i - \bar{v})}{\varphi(x_i - \bar{v})} = 0.$$

Положим $f(x) = \varphi'(x)/\varphi(x)$. Эта функция нечетна, и мы имеем $\sum_{i=1}^{n} f(x_i - \bar{v}) = 0$.

Поскольку x_i произвольны, мы можем выбрать их, как нам удобно, и если взять $x_1 = a$ и $x_2 = x_3 = x_4 = \ldots = x_n = a - nb$, где a и b произвольны, то будем иметь

$$\begin{aligned}
\bar{v} &= \frac{1}{n}\{a + (n-1)(a - nb)\} = a - (n-1)b,\\
x_1 - \bar{v} &= a - \{a - (n-1)b\} = (n-1)b,\\
x_i - \bar{v} &= (a - nb) - \{a - (n-1)b\} = -b.
\end{aligned}$$

$$\begin{aligned}
\sum_{i=1}^{n} f(x_i - \bar{v}) &= f(x_1 - \bar{v}) + \sum_{i=2}^{n} f(x_i - \bar{v})\\
&= f((n-1)b) + (n-1)f(-b) = 0.
\end{aligned}$$

А так как $f(x)$ нечетна, то

$$f((n-1)b) = -(n-1)f(-b) = (n-1)f(b).$$

И мы возвращаемся к прежнему функциональному уравнению, означающему, что $f(x) = \varphi'(x)/\varphi(x) = kx$, которое решается методом разделения переменных, в результате чего получается $\varphi(x) = Ae^{kx^2}$. А используя тот факт, что площадь под кривой равна 1, мы окончательно находим $\varphi(x) = (h/\sqrt{\pi})e^{-h^2x^2}$, где h – положительная постоянная, которую Гаусс называл *точностью процесса измерения*.

При помощи этого распределения ошибок мы можем измерить вероятность независимого возникновения всех трех ошибок:

$$\Omega = \left(\frac{h}{\sqrt{\pi}}\right)^3 \exp\left(-h^2 \sum_{i=1}^{3} (a + bt_i - x_i)^2\right).$$

Если мы хотим, чтобы она была максимальной, то $\sum_{i=1}^{3}(a + bt_i - x_i)^2$ должна быть минимальной, и мы пришли к частному случаю того, что теперь называется методом *наименьших квадратов*, – его история долгая и сложная, и мы ее опустим. Если рассматривать это выражение как функцию двух переменных a и b, продифференцировать ее и приравнять частные производные нулю, то мы получим уравнения, которые можно лаконично записать в матричной форме:

$$\begin{pmatrix} \sum 1 & \sum t_i \\ \sum t_i & \sum t_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum x_i \\ \sum x_i t_i \end{pmatrix}$$

и легко решить; в литературе они называются *нормальными* уравнениями. Нормальные уравнения, с которыми имел дело Гаусс, были куда более трудными, но благодаря им орбиту Цереры удалось вычислить еще точнее, как и ее истинную сущность: она оказалась не новой пятой планетой, а первым астероидом. Эта часть истории завершается новым появлением на сцене Лапласа, который, познакомившись с выводом Гаусса, предложил свой собственный, более строгий, в качестве альтернативы. Нормальная кривая заняла свое место как кривая ошибок измерения, и были разработаны относящиеся к этому теории.

Последний из трех этапов эволюции нормальной кривой связан с нашим первым выводом кривой ошибок. Он начинается (в повествовательной форме) на стр. 19 и 20 подробного обзора, вышедшего в 1850 г., – обзора, занимающего 57 страниц журнала «The Edinburgh Review», за авторством известного английского астронома и физика сэра Джона Гершеля (Herschel 1850). А книге, которой был посвящен обзор, суждено было стать весьма влиятельной – благодаря как благодатной роли обзора, написанного авторитетным ученым, так и исключительному влиянию самого автора. Да и книга далеко не ординарная.

8.6. Нормальное распределение

Если бы мы питали надежду сколько-нибудь подробно рассказать о том, как вездесущая нормальная кривая влияет на нашу повседневную жизнь, то нам не хватило бы печатных страниц и пришлось бы без особой на то необходимости обсуждать работы Фишера, Пирсона, Йетса, Тьюки, Госсета и др., но мы ограничимся только вкладом бельгийского ученого-универсалиста

Ламберта Адольфа Жака Кетле (1796–1874), чьи основополагающие исследования, касающиеся нормальной кривой, задали образец людям куда более известным, включая Флоренс Найтингейл, Чарльза Дарвина и разносторонне одаренного Фрэнсиса Гальтона, двоюродного брата Дальтона.

Современный индекс массы тела (ИМТ), интерес к которому рос вместе с обхватом талии населения преуспевающих западных стран, – вовсе не новое изобретение. У него есть и другое название, которое сейчас почти не употребляется: *индекс Кетле*[1], в честь Адольфа Кетле, астронома, математика, статистика и социолога (см. Eknoyan 2008). ИМТ – лишь одна из предложенных им количественных характеристик человека, и именно благодаря ему и его уникальному кругу интересов мы находим первое письменное подтверждение того, что кривая ошибок родом из XIX в. могла быть предтечей нынешней нормальной кривой. Именно, закон, определяющий разброс производимых человеком наблюдений, можно было бы привлечь для характеристики самого человека: его веса, роста и комбинации того и другого, показателя тучности. А также всех прочих характеристик человеческой деятельности: частоты и типов преступлений, возраста вступления в брак и наступления смерти и т. д. и т. п. Проведенный анализ данных о преступности в Париже позволил ему предположить, что совершившая противоправное деяние «хорошо образованная женщина старше тридцати лет, добровольно взявшая на себя ответственность за преступление против личности», скорее всего, избежит обвинения или иного результативного осуждения. На самом деле не будет слишком большим преувеличением предположить, что Кетле видел форму нормальной кривой во всех природных явлениях, равно как не будет преувеличением предположить, что он трезво оценивал новизну своих идей и связанные с ними трудности:

> «Я не скрываю от себя многочисленных трудностей, связанных с применением математических методов к анализу явлений. Это исследование, пока еще новое, я знаю, насторожило многих читателей, которым показалось, что они увидели в нем тенденцию к материализации того, что принадлежит благородному духу человека».

Его влияние – благодаря широкому распространению и личностям тех, кто ему подвергся – было значительным и в целом позитивным, и он по праву считается основателем количественных общественных наук. Самой известной его публикацией, сборником выполненных к тому моменту работ, является вышедшая в 1835 г. книга «Трактат о человеке и развитии его способностей», именно в ней мы находим ИМТ и многое другое. Но за деталями его вклада в предмет этой главы мы обратимся в другое место.

Во впечатляющем перечне его должностей, почетных званий и наград мы находим упоминание о том, что одно время он был преподавателем в Королевской читальне Брюсселя, а затем в 1830-х гг. занимал должность наставника принцев Эрнста и Альберта Саксен-Кобург-Готских; младший брат, Альберт, стал мужем королевы Виктории в 1840 г., а старший, Эрнст, в 1844 г. вступил на престол после смерти отца. Когда мальчики вернулись в Германию,

[1] Название ИМТ (англ. BMI) он получил в 1972 г.; термин предложил Ансель Киз.

чтобы завершить образование, уроки продолжились по переписке, и многие из них впоследствии вошли в книгу «Письма, адресованные Его Королевскому высочеству Великому герцогу Саксен-Кобург-Готскому, по теории вероятностей применительно к нравственным и политическим наукам», изданную в 1844 г. Это именно та книга, из которой взята приведенная выше цитата и которую так пространно обозревал Гершель, и именно в этом совершенно неожиданном месте мы находим упоминание об использовании кривой ошибок в контексте, не связанном с ошибками измерения.

Из 46 писем, вошедших в книгу, для нас представляют интерес в основном письма XX и XXI; первое озаглавлено «К вопросу о том, является ли среднее арифметическое истинным средним размера человека». Объясняя юному (но с научным складом ума) принцу свои мысли по этому поводу, Кетле предложил ему рассмотреть скульптуру, которую он называет «Гладиатор», и подумать о ее основных размерах, в частности об обхвате груди. «Гладиатор» являет пример совершенной формы, а любая попытка измерить обхват груди обязательно сопровождается ошибкой; но проведите тысячу измерений, и среднее будет очень близко к истинной величине. Кроме того, измерения, меньше всего отличающиеся от среднего, составят самую многочисленную группу, а по мере удаления от среднего измерений будет становиться все меньше: «Если последовательность групп нанести на график, то получится кривая вероятностей». Далее он плавно переходит к тщательному копированию «Гладиатора» тысячью разных скульпторов и говорит, что и тогда измерения будут подчиняться закону вероятностей.

Оставив позади этот промежуточный шаг, Кетле приступает к кульминации, потребовавшей не столь плавного перехода, – к солдатам шотландской армии. Любому статистику нужны данные, и Кетле черпал свои из различных государственных отчетов, равно как из многих других источников, а также был благодарным получателем многочисленных и разнообразных наборов данных, которые присылали его корреспонденты. Особенно его интересовали отчеты военного ведомства, которые он считал источником непротиворечивых и надежных данных. Каким-то образом ему в руки попал том 13 «Эдинбургского медицинского журнала» за 1817 г. Среди различных статей (чтение многих из них требует от читателя мучительных усилий) на стр. 260 имеется одиннадцать таблиц с измерением обхвата груди (с корреляцией по весу) для каждого из полков шотландской милиции; всего 5731 измерение, хотя, согласно Кетле, их было 5738. Его таблица воспроизведена в нашей табл. 8.2, а ниже приводится его комментарий к ней:

> «И теперь я задаю вопрос, не будет ли верным делом побиться об заклад, что лицо, имеющее мало опыта в измерении человеческого тела, совершит ошибку на целый дюйм при измерении обхвата груди в 40 дюймов? Что ж, признавая вероятие такой ошибки, 5738 измерений одного человека, безусловно, образуют группы не с большей регулярностью – по порядку величины, – чем измерения 5738 шотландских солдат; если предъявить нам оба ряда, не указав их происхождения, то мы вряд ли сможем сказать, какой ряд получен в результате обмера 5738 разных солдат, а какой – в результате обмера одного человека, но менее умелым измерителем и более грубыми инструментами.

> Приведенный мной пример заслуживает, думается мне, пристального внимания: он показывает, что результаты действительно выглядят так, как если бы обмеренные грудные клетки принадлежали одному и тому же лицу, если угодно, с идеальным сложением, но пропорции которого мы устанавливаем в результате достаточно продолжительных испытаний. Если бы это не было законом природы, то измерения (при всем их несовершенстве) не образовывали бы группы с вызывающей удивление симметрией, навязанной им законом вероятностей».

Таблица 8.2. Измерения обхвата груди в шотландской милиции

Измерение обхвата груди	33	34	35	36	37	38	39	40
Число мужчин	3	18	81	185	420	749	1073	1079
Измерение обхвата груди	41	42	43	44	45	46	47	48
Число мужчин	934	658	370	92	50	21	4	1

Утверждению об эквивалентности одного измерения 5738 солдат и 5738 измерений одного солдата с трудом можно поверить, пусть даже последние выполнены «менее умелым измерителем и более грубыми инструментами». Идеальный «Гладиатор» был заменен *l'homme moyen* – средним человеком – в качестве фундаментальной теоретической конструкции, идеальным индивидуумом, являвшимся общественным центром притяжения, который стал постоянной парадигмой для статистических и социальных исследований. Не каким-то одним средним человеком, а теоретическим идеалом в каждом конкретном контексте; в данном случае это шотландский солдат и измерение обхвата груди. Идея пронизывает наш современный язык: *средний мужчина* нуждается в 250 г углеводов ежедневно, в *средней семье* 2.4 детей... В письме XXI юному принцу предлагается рассмотреть возможный обман в опубликованных статистических данных о росте 100 000 французских новобранцев; в наборе данных было слишком много записей о росте меньше 5 футов 2 дюйма, что не ложилось в аппроксимацию Кетле нормального распределения, и «власти снисходительно относятся к случаям, когда вместо мужчин походящего роста призываются менее рослые».

Своим энтузиазмом Кетле заразил широкую аудиторию, и нормальная кривая быстро стала главной моделью явлений природы вообще и человека в частности. Разнообразные вклады многих других ученых, которые мы не рассмотрели, в совокупности обеспечили нормальной кривой выдающееся место среди статистических распределений. Одним из основных применений является приближение биномиального распределения, да и для обработки ошибок она по-прежнему используется. Сознавая безнадежность любых попыток осветить подлинную важность нормальной кривой, мы закончим художественно расположенными словами У. Дж. Юдена, взятыми из его книги «Эксперимент и измерение», изданной Национальным бюро стандартов США в 1984 г. Он писал, что является вла-

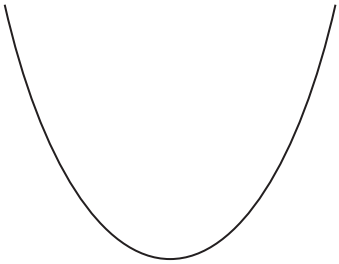
дельцем собственного печатного станка, – и вот нашел ему достойное применение[1]:

THE
NORMAL
LAW OF ERROR
STANDS OUT IN THE
EXPERIENCE OF MANKIND
AS ONE OF THE BROADEST
GENERALISATIONS OF NATURAL
PHILOSOPHY ♦ IT SERVES AS THE
GUIDING INSTRUMENT IN RESEARCHES
IN THE PHYSICAL AND SOCIAL SCIENCES AND
IN MEDICINE AGRICULTURE AND ENGINEERING ♦
IT IS AN INDISPENSABLE TOOL FOR THE ANALYSIS AND THE
INTERPRETATION OF THE BASIC DATA OBTAINED BY OBSERVATION AND EXPERIMENT

[1] Нормальный закон распределения ошибок выделяется в опыте человечества как одно из самых широких обобщений натурфилософии ♦ Он служит направляющим инструментом в исследованиях по физическим и общественным наукам, в медицине, сельском хозяйстве и технике ♦ Это незаменимый инструмент для анализа и интерпретации базовых данных, полученных путем наблюдения и эксперимента. – *Прим. перев.*

Глава 9

Цепная линия



Почему именно эта кривая?

Во-первых, она, вероятно, ввела в заблуждение Галилея и, безусловно, обманула многих до и после него. Во-вторых, это важная проверка возможностей только что возникшего математического анализа. В-третьих, будучи нарисована на вертикальной стене, она может использоваться для вычисления логарифмов. В-четвертых, эта форма повсеместно встречается в архитектуре и весьма важна в строительстве. В-пятых, она была названа в честь будущего президента США. И наконец, эта форма могла бы дать интересный эффект при строительстве дорог.

9.1. Вопрос симметрии

С кривыми связаны две симметрии, которые в терминах определяющей функции $f(x)$ имеют вид $f(-x) = f(x)$ или $f(-x) = -f(x)$; первое условие означает зеркальную симметрию относительно оси y, вторая – симметрию относительно поворота на 180° вокруг начала координат. Поскольку четные степени x удовлетворяют первому условию, а нечетные – второму, для описания такого поведения применяются соответственно термины *четная* и *нечетная* функция. Отметим также, что функция $\cos x$ четная, $\sin x$ – нечетная и т. д. Очевидно, что далеко не все кривые обладают такой симметрией, но существует изящное построение того, что можно было бы с должным основанием назвать четной и нечетной частями кривой. Пусть $f(x)$ – произвольная функция с подходящей областью определения; обозначим $g(x) = \frac{1}{2}(f(x) + f(-x))$ и $g(x) = \frac{1}{2}(f(x) - f(-x))$.

Тогда $g(x)$ – четная функция, а $h(x)$ – нечетная; более того, $f(x) = g(x) + h(x)$, поэтому мы вполне можем назвать $g(x)$ четной частью $f(x)$, а $h(x)$ – ее нечетной частью. Иначе говоря, любую функцию, которая сама по себе не является ни четной, ни нечетной, но имеет подходящую область определения, можно представить в виде суммы четной и нечетной функции. Например,

$$f(x) = x^2 + x + 1 = (x^2 + 1) + x = g(x) + h(x).$$

На рис. 9.1 график этой функции – сдвинутая на единицу сплошная парабола – представлен в виде суммы пунктирной параболы, симметричной относительно оси y, и пунктирной прямой, симметричной относительно начала координат.

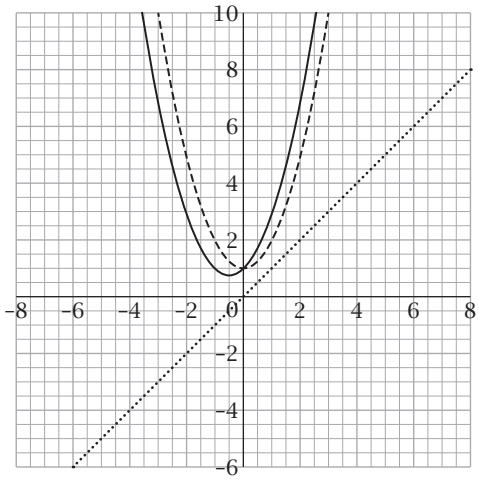


Рис. 9.1. Четная и нечетная части $f(x) = x^2 + x + 1$

Такой фокус можно провернуть с мириадами функций, но, что и не удивительно, особенно важные результаты дает экспоненциальная функция $f(x) = e^x$, они показаны на рис. 9.2. Симметричная относительно начала координат часть напоминает график тангенса, но, конечно, им не является, а симметричная относительно оси y часть похожа на параболу, но только похожа: уравнения этих кривых имеют вид $g(x) = \frac{1}{2}(e^x + e^{-x})$ и $h(x) = \frac{1}{2}(e^x - e^{-x})$. В современной терминологии это две основные гиперболические функции, $\operatorname{ch} x$ и $\operatorname{sh} x$ соответственно, и в этой главе нас будет интересовать именно $\operatorname{ch} x$, а точнее более общая функция $a \operatorname{ch}(x/a) = \frac{1}{2}a(e^{x/a} + e^{-x/a})$ для $a > 0$, поскольку именно она описывает самую общую форму кривой: *цепную линию*. Если бы мы по ошибке решили, что это парабола, то оказались бы в недурной компании: Галилей, похоже, тоже так думал, равно как многие до и после него. Но не в связи с четными и нечетными функциями, да и вообще вне всякой связи с функциями, а в контексте экстраполяции одного истинного динамического наблюдения на статическое, которое истинным не является.

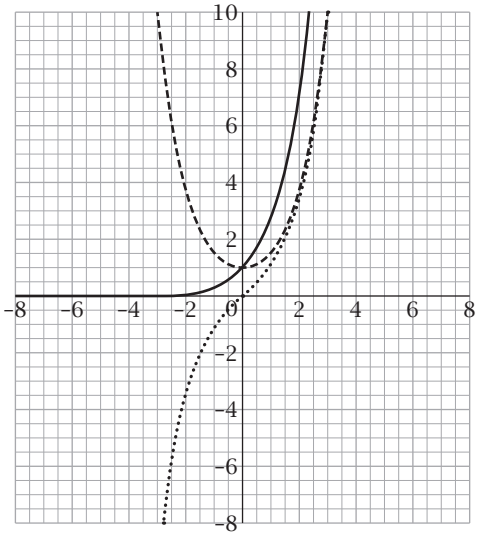


Рис. 9.2. Четная и нечетная части $f(x) = e^x$

9.2. Исторические ошибки

В 1633 г. римская инквизиция внесла книгу Галилея «Диалог о двух главнейших системах мира» в «Индекс запрещенных книг», совершив при этом величайшую ошибку – отрицание системы Коперника в пользу система Птолемея: Земля считалась неоспоримым центром Вселенной, а любое иное утверждение – ересью; Галилей был принужден согласиться с этим, хотя впоследствии ему приписывались слова «а все-таки она вертится» (о Земле). Чудо, что его последующую и последнюю публикацию «Беседы и математические доказательства, касающиеся двух новых отраслей науки» («Беседы») удалось издать в 1638 г. в Голландии, далеко от лап инквизиции. Обе книги построены в форме диалога между персонажами (форма не новая и вполне обычная) и содержат основополагающие научные принципы и выводы, на которые мы сегодня опираемся: вовсе не случайно Галилей не раз аттестуется как *отец* ..., где вместо многоточия следует подставить название отрасли науки. Не стоит удивляться тому, что в этих муках рождения современного научного метода и глубоких идей встречались и неверные суждения, о чем было известно и самому Галилею. В конце третьего дня «Бесед» после теоремы VI Галилей заставил своего персонажа Сальвиати высказаться довольно поэтически[1]:

> «Сальвиати. Но столь глубокие соображения относятся уже к учениям более высоким, чем наше. Для нас будет достаточно, если мы уподобимся менее искусным рабочим, выламывающим и добывающим из карьера мрамор, из которого впоследствии опытные скульпторы могут создать удивительные образы, скрывавшиеся под грубой и бесформенной корой».

[1] Перевод С. Н. Долгова.

Поскольку Сальвиати представляет тогдашнего Галилея, интересно, кого в сравнении с ним можно было бы по праву назвать опытным. И тем не менее во втором дне, правильно доказав, что траектория брошенного тела в условиях, когда сопротивлением воздуха можно пренебречь, является параболой, он ошибся, вложив в уста Сальвиати следующий комментарий:

> «Сальвиати. Существует много способов начертить такую линию, но я познакомлю вас только с двумя наиболее простыми. Один из них действительно изумителен, так как, пользуясь им, я в меньшее время, чем то, за которое другие вычерчивают на бумаге четыре или шесть окружностей разного диаметра, могу начертить тридцать-сорок параболических линий, не менее тонких, точных и правильных, чем упомянутые окружности. У меня имеется бронзовый шарик весьма правильной формы, величиною не более ореха. Брошенный на металлическое зеркало, лежащее не совсем горизонтально, но несколько наклонно, так что при движении он может по нему катиться, производя при этом легкое давление, шарик этот оставляет след в виде тонкой и правильной параболической линии…
>
> Другой способ начертить искомую параболу на призме состоит в следующем. Вобьем в стену два гвоздя на одинаковой высоте над горизонтом и на таком расстоянии друг от друга, чтобы оно равнялось двойной ширине прямоугольника, на котором желательно построить полупараболу; между одним и другим гвоздем подвесим тонкую цепочку, которая свешивалась бы вниз и была такой длины, чтобы самая низкая точка ее находилась от уровня гвоздя на расстоянии, равном длине призмы. Цепочка эта, свисая, расположится в виде параболы, так что, отметив ее след на стене пунктиром, мы получим полную параболу, рассекаемую пополам перпендикуляром, проведенным через середину линии, соединяющей оба гвоздя».

Первая часть диалога, конечно же, правильна, а вот вторая ошибочна. Не то чтобы полностью неверна, для практических целей разницы нет, но все же имеется расхождение между параболой и формой свободно свисающей цепи, как показано на рис. 9.3. Галилей выбрал отношение длин 2:1, и если мы возьмем отрезок $-1 \leqslant x \leqslant 1$, то уравнение параболы будет иметь вид $y = x^2$, а ее график показан пунктирной линией; форма же, принимаемая свисающей цепью, на том же отрезке показана сплошной линией. Различие крохотное, но в то же время огромное. Но Галилей оставил себе лазейку. Те же идеи обсуждаются еще раз в конце четвертого дня:

> «Сальвиати. Скажу вам более: мы должны с удивлением и удовольствием констатировать, что канат, натянутый в большей или меньшей степени, располагается по линии, весьма близкой к параболе. Сходство столь велико, что если вы начертите на вертикальной плоскости параболическую линию и, рассматривая ее в обратном положении, т. е. обратив вершину ее вниз, а основание, параллельное горизонту, вверх, подвесите цепочку, укрепив концы ее в конечных точках основания начерченной параболы, то вы увидите (укорачивая или удлиняя цепочку, смотря по надобности), что она очень близко подходит к параболе; при этом совпадение ее с параболой наблюдается тем большим, чем меньше кривизна параболы, т. е. чем более последняя растянута, так что при параболе, описываемой при угле наклона в 45°, цепочка почти точно совпадает с параболой.

Сагредо. Таким образом, при помощи подобной тонко сработанной цепочки можно в одну минуту начертить на плоскости большое количество параболических линий?

Сальвиати. Конечно, и притом с немалой пользой, как я вам сейчас покажу».

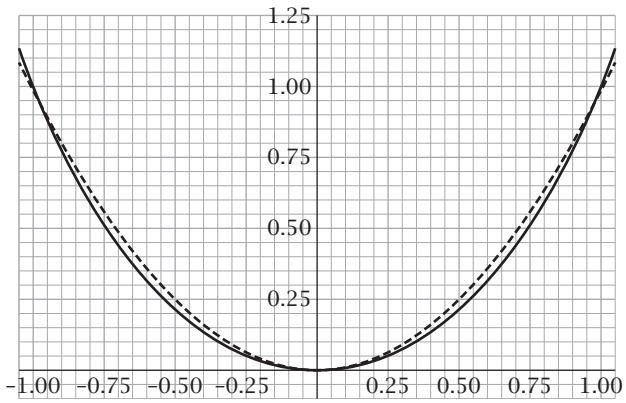


Рис. 9.3. Небольшое расхождение

Никакого «сейчас» на этот счет в диалоге нет, но Галилей перешел от точного совпадения к приближенному: почему он сохранил в диалогах обе степени убежденности, мы можем только гадать, но знать очень хотелось бы. Если еще раз взять отрезок $-1 \leqslant x \leqslant 1$, то парабола $y = \frac{1}{2}x^2$ в обеих конечных точках наклонена под углом 45° и показана пунктирной линией на рис. 9.4. Отношение длин теперь равно 2:½ = 4:1, и свисающая цепь снова показана сплошной линией.

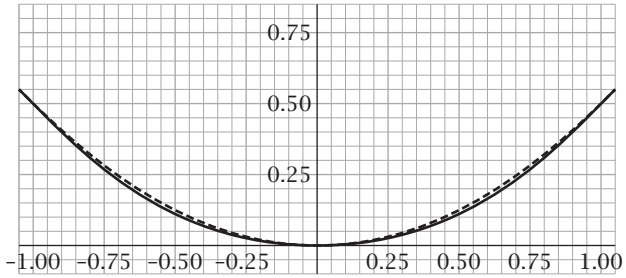


Рис. 9.4. Еще меньшее расхождение

Но у великого ученого были основания так считать, пусть даже и не слишком убедительные. Что бы Галилей ни думал на самом деле о форме свисающей цепи, предположение о том, что это парабола, было согласно с общепринятым мнением: и до, и после него все считали, что цепь свисает по параболе, хотя (по очевидным причинам) доказать это никому не удавалось. Например, друг и корреспондент Рене Декарта фламандский математик Исаак Бекман просил его подтвердить эту гипотезу, но у нас нет никаких свидетельств, что Декарт ему ответил. В конце концов конические сечения были очень хорошо

изучены и отлично послужили изучению природы: это и окружность в своих многочисленных проявлениях, и эллипсы в законах Кеплера, и, как мы видели, собственная работа Галилея, установившего, что брошенное тело совершает движение по параболе. Коническое сечение было очевидным кандидатом, а парабола – самой очевидной формой, которую могла бы принять свисающая цепь. Для любой физической задачи выбор конического сечения в качестве кандидата на ее решение всегда стоял на первом месте – эта точка зрения еще раз подчеркнута в третьем дне «Бесед», где излагается теорема XXII, утверждающая, что путь по дуге окружности из ее высшей точки в низшую быстрее, чем по соединяющей их хорде, и что

> «из сказанного можно заключить, что быстрейшее движение от одной конечной точки до другой происходит не по кратчайшей линии, каковой является прямая, а по дуге окружности».

Это неправда, такой кривой является *брахистохрона*.

Ошибочно считая, что цепь принимает ту же форму, по которой движется брошенное тело, Галилей пал жертвой оптического обмана и общего мнения, но не недостаточно острого аналитического ума – и на старуху бывает проруха. У него просто не было математического аппарата, который позволил бы прийти к правильному выводу, да и в любом случае локально цепная линия, как мы видели, очень похожа на параболу, а в результате подбора параметров это сходство можно сделать разительным – по крайней мере, в пределах длины цепи. Только не по годам развитому семнадцатилетнему Христиану Гюйгенсу[1] предстояло показать, что эта кривая не парабола, и, более того, определить дополнительные ограничения, при которых она таковой является. В датированном 28 октября 1646 г. письме преподобному отцу Марину Мерсенну, человеку, который, как мы уже упоминали, вел легендарную переписку с европейскими мыслителями, юный Гюйгенс (1638–1656) обещал

> «в следующем письме прислать вам доказательство того, что свисающая веревка или цепь не принимает форму параболы, и определить, каким должно быть давление на математическую нить или цепь, чтобы они приняли эту форму; я нашел это доказательство недавно».

После полного энтузиазма ответа Мерсенна Гюйгенс выполнил свое обещание в письме, датированном ноябрем, где представил свои рассуждения в виде последовательности из девяти предложений, основанных на идеях Евклида, центральной из которых была идея подобных треугольников. От начальной модели, где роль цепи играла невесомая нить, к которой через равные промежутки были подвешены одинаковые веса, он перешел к модели, где точечные веса были заменены массивными прямолинейными отрезками. Говоря современным языком, он понял, что парабола определяется любыми тремя неколлинеарными точками на цепи, и доказал, что другие точки не могут лежать на этой параболе. Теперь что касается давления, которое заставило бы свиса-

[1] Который пользовался методами, почерпнутыми из изучения трудов Симона Стевина.

ющую цепь принять форму параболы: распределим веса равномерно по цепи, измеряя расстояние не вдоль самой цепи, а вдоль горизонтальной прямой под ней, тогда доказательство вытекает из его предложений 11 и 12. В следующем разделе мы рассмотрим важные следствия из этой поправки.

Его доказательства можно признать корректными, но они многословны вследствие смешения риторики и символизма в представлении рассматриваемого случая, требующих от читателя принять показанный путь как единственно разумный. Но избыток риторики в математическом рассуждении опасен тем, что может завести не туда. Исправив одну историческую ошибку, Гюйгенс допустил другую в последнем предложении, которое он провозгласил под названием «*Manifestum.φανερδ*» – латинское и греческое слова, означающие «очевидно». То, что показалось ему очевидным, изображено на рис. 9.5, копирующем его собственный чертеж:

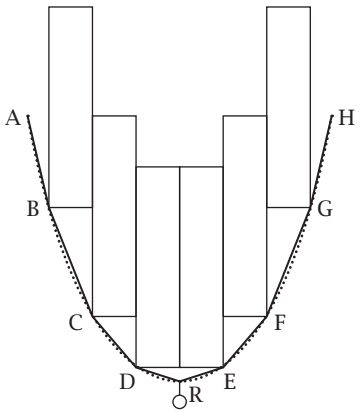


Рис. 9.5. Ошибка Гюйгенса

> «Отсюда ясно, что если бы на струне <…> можно было бы разместить маленькие бусины или параллелепипеды одинакового веса, размера и формы, то точки *A*, *B*, *C* и так далее давят на струну, и все они находятся на одной и той же параболе».

И снова преимущественно словесная аргументация убедила Гюйгенса, но на этот раз сбила его с пути истинного. На самом деле параллелепипеды толкают струну вовне в направлении нормали в каждом углу, где имеет место касание, и это деформирует струну в дугу окружности. Спустя двадцать два года, став старше и гораздо мудрее, он в заметке на полях признал свою ошибку, добавив комментарий «это ниоткуда не следует и неверно».

С годами доказательства того, что эта кривая не парабола, стали яснее: одно из таких дал малоизвестный немецкий ученый Иоахим Юнгиус, но, пожалуй, самым известным является доказательство, которое в 1673 г. привел иезуит, отец Игнас Гастон Парди, применив более элегантные методы (Pardies 1725, стр. 280–81). После этого такое удобное и разумное предположение, что свисающая цепь принимает формулу параболы, было порублено на мелкие кусочки: научный мир теперь знал, что это не та форма, но не знал, какая та.

9.3. Опознанная кривая

Проблема назрела и перезрела. Запас математических кривых существенно расширился, но по-прежнему оставалось неизвестным, какая из них (если она вообще существует) описывает форму свисающей цепи, хотя этот вопрос продолжали задавать себе авторитетные математики. Если не коническое сечение, то что? Расплывчатый ответ гласил, что это должна быть *механическая кривая*; ее нельзя построить циркулем и линейкой, но она порождена природой (см. главу 5).

Сегодня довольно обычное дело – распространить задачу в надежде, что про нее узнает кто-то, способный ее решить, или, еще того лучше, – собрать особенно трудные задачи и назначить премию за их решение, как в случае проблем тысячелетия, каждая из которых оценена в 1 млн долл. Так же было и в XVII в. с той существенной разницей, что распространение происходило в форме вызовов, которыми обменивались тогдашние именитые академики: словесно, в письме или обыкновенно и более публично – посредством публикации в журнале. Одним таким журналом, особенно влиятельным, потому что в числе его учредителей и постоянных авторов значился великий Лейбниц, был ежемесячный «Acta Eruditorum»[1], хотя, быть может, наиболее известна его роль как главного органа, оказывающего поддержку Лейбницу в знаменитом споре о приоритете изобретения математического анализа[2]. В июньском номере за 1696 г. Иоганн I Бернулли призвал читателей найти кривую быстрейшего спуска (брахистохрону, а не галилееву дугу окружности), по которой свободно движущаяся частица перемещается из более высокой точки в более низкую, а в майском номере за 1697 г. появилось не только решение самого Иоганна, но также решения Лейбница маркиза де Лопиталя и брата Бернулли Якоба I[3]. Спустя несколько лет в майском номере того же журнала за 1690 г. Якоб I представил свое решение задачи об *изохроне*: кривой, двигаясь по которой под действием силы тяжести, тело опускалось бы до заданного уровня за время, не зависящее от своего начального положения; он также и поставил свою задачу перед научным сообществом, а особенно перед Лейбницем и его новым исчислением: определить форму свисающей цепи (Bernoulli 1690). И как часто бывает с автобусами, подъезжающими один за другим после долго отсутствия, в июне 1691 г. появилось три решения: Лейбница, Иоганна I Бернулли и Христиана Гюйгенса. Гюйгенс вернулся к этой задаче через 45 лет с методами 45-летней давности; он не желал осваивать новомодное исчисление.

Верно ли, что Лейбниц, один из создателей математического анализа, и Бернулли, один из его самых ревностных приверженцев, применили всю его мощь? В обоих случаях – да, но читателю их доказательств будет извинительно думать иначе. Предложенные ими решения были построены в стиле Гюйгенса: евклидовы, геометрические, со сложными чертежами, запутанными рассуждениями

---

[1] В переводе «Отчеты о трудах ученых».

[2] Журнал «Transactions of the Royal Society» играл ту же роль по отношению к Ньютону в Англии.

[3] Ньютон тоже анонимно решил эту задачу, что дало повод для знаменитого отклика Бернулли «tanquam ex undue leonem» (по когтям узнают льва).

и странными для нас обозначениями. В то время такая задача считалась решенной, если было предъявлено построение кривой, и именно этот факт определял их подход. Не было никаких видимых признаков анализа, и ни в одном из построений не прослеживалась связь с цепью, свободно подвешенной в двух точках, закрепленных на одном уровне. Мы опустим решение Гюйгенса, объясним решение Лейбница, являющееся образцом для решений такого рода, и подробно остановимся на математическом анализе, скрытом в подходе Бернулли.

В путанице линий на чертеже Лейбница и на нашем рис. 9.6 скрыта цепная линия с соответствующей меткой и еще одна кривая, названная *логарифмической*, которую вообще-то следовало назвать экспоненциальной[1]. Чертеж сопровождался пояснениями:

> «Пусть дана неопределенная прямая линия ON, параллельная горизонту, а также OA, перпендикулярный к ней отрезок, равный O3N, а из точки 3N проведен вертикальный отрезок 3N3ξ, который относится OA как D к K. Найдем среднее пропорциональное 1N1ξ (отрезков OA м 3N3ξ); затем среднее пропорциональное 1N1ξ и 3N3ξ, а затем среднее пропорциональное 1N1ξ и OA; во время поиска вторых средних пропорциональных, а по ним третьих пропорциональных будем двигаться вдоль кривой 3ξ-1ξ A-1(ξ)-3(ξ) таким образом, что, когда берутся равные интервалы 3N1N, 1NO, O1(N), 1(N)3(N) и т. д., ординаты 3N3ξ, 1N1ξ, OA, 1(N)1(ξ), 3(N)3(ξ) образуют непрерывную геометрическую прогрессию, касаясь кривой, которую я обычно называю логарифмической. Итак, отложив равные отрезки ON и O(N), восставим из N и (N) перпендикуляры NC и (N)(C), равные полусумме Nξ и (N)(ξ), так что точки C и (C) будут лежать на цепной кривой FCA(C)L, на которой мы можем геометрически определить столько точек, сколько пожелаем».

Прозрачным этот текст не назовешь, но, разбираясь в том, что современному взгляду кажется мешаниной линий, и привлекая подкрепляющее рассуждение, можно увидеть его общую схему, включающую экспоненциальную кривую, которую мы записали бы в виде $y = ar^{x/a}$ и которая проходит через точки $(-a, ar^{-1})$, $(0, a)$, (a, ar), где a – постоянная, которую мы впоследствии положим равной единице, чтобы не связываться с делением и потому что тогда логарифм будет равен 0. Отрезки слева от чертежа, обозначенные K и D, определяют конкретный показатель степени $r = d/k$, равный отношению их длин. Тут есть еще скрытая тонкость, о которой мы поговорим ниже. Лейбниц заметил, что если точки (x_1, y_1) и (x_2, y_2) лежат на этой кривой, то точка $(\frac{1}{2}(x_1 + x_2, \sqrt{y_1 y_2})$ также лежит на ней; это просто следствие правил для показателей степеней. Теперь мы можем продолжить кривую как угодно далеко, применяя этот факт повторно. Поэтому каждая пара точек на кривой $(x, ar^{x/a})$ и $(-x, ar^{-x/a})$ порождает пару точек на цепной линии $(\pm x, \frac{1}{2}(ar^{x/a} + ar^{-x/a})$, так что цепная линия строится точка за точкой.

Теперь что касается связи со свисающей цепью:

> «Эту кривую можно построить и провести очень просто с помощью физического построения, а именно подвесив струну, а лучше небольшую цепочку (переменной длины)».

Остается только догадываться почему.

---

[1] Обе надписи – современное добавление к чертежу.

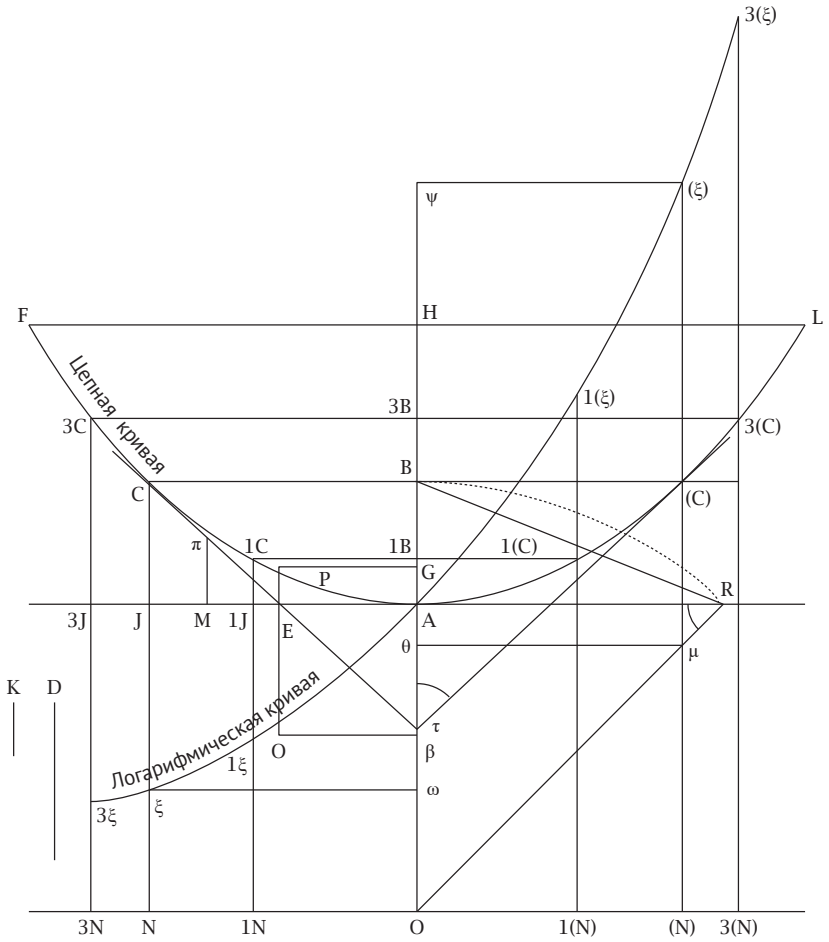


Рис. 9.6. Лейбниц, цепная линия и свисающая цепь

Наверное, полушутя он превратил свое доказательство в физический метод вычисления логарифмов:

> «А коль скоро вы умеете определять эту кривую, вы можете найти все средние пропорциональные и все логарифмы, которые вам нужны, а также выполнить квадратуру гиперболы. Если цепная кривая физически построена, путем подвешивания струны или цепочки, то вы можете построить с ее помощью столько средних пропорциональных, сколько захотите, и найти логарифмы чисел или числа по логарифмам. Если вы ищете логарифм числа Oω, т. е. логарифм отношения OA к Oω, а логарифм OA (который я выбрал в качестве единицы и который также буду называть параметром) считается равным нулю, то должны взять третье пропорциональное Oψ между Oω и OA; затем отложите на оси абсцисс полусумму OB отрезков Oω и Oψ, а соответствующая ордината BC или ON на цепной кривой будет искомым логарифмом предложенного числа».

Итак, подвесим нашу цепь на фоне вертикально расположенного листа миллиметровки, на которой для справки нарисован пунктирной линией график экспоненциальной функции $y = ar^x$, как показано на рис. 9.7. Для удобства увеличим масштаб, как на рис. 9.8, где с точками, обозначенными заглавными буквами, ассоциированы соответствующие строчные буквы: мы ищем логарифм числа *w*, расположенного на вертикальной оси под A.

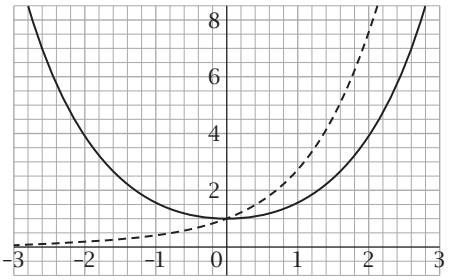


Рис. 9.7. Экспонента и цепная линия

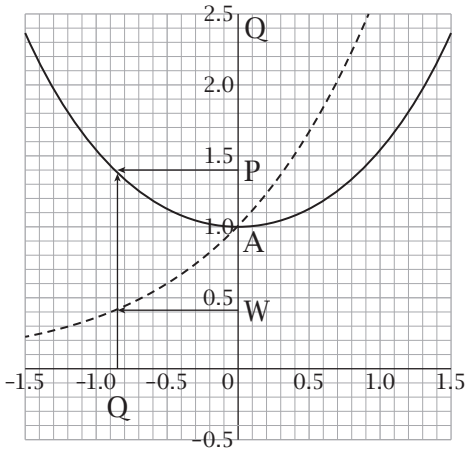


Рис. 9.8. Вычисление логарифмов

Для того чтобы найти логарифм *w* с помощью экспоненциальной кривой, нам нужно только сместиться по горизонтали до пересечения с кривой, тогда проекция Q точки пересечения на горизонтальную ось даст искомое число. Однако мы должны найти Q, используя только цепную линию, поэтому будем искать точку на цепной линии, расположенную выше Q на той же вертикали, а для этого Лейбниц предписывает нам найти среднее *a* и *третьего пропорционального w* и *a*. Мы уже встречали этот термин в главе 5, но напомним, что третье пропорциональное – это точка Q на вертикальной оси, для которой *w*, *a* и *q* образуют геометрическую прогрессию, т. е. мы ищем точку Q, используя отношение $w/a = a/q$, а точку P – с помощью $p = \frac{1}{2}(w + q)$: чтобы найти логарифм *w*, вычислим *q* и по нему вычислим *p*, которая находится на вертикальной оси, а затем сместимся поперек до пересечения с цепной линией, тогда координата *x* найденной точки будем искомым логарифмом.

Почему это предписание работает? Лейбниц ни словом не обмолвился о присущей ему тонкости, а мы можем описать метод и эту тонкость с помощью предыдущих обозначений. Пусть $w = ar^{\alpha/a}$ для некоторого α, тогда находим Q, пользуясь следующей импликацией:

$$\frac{ar^{\alpha/a}}{a} = \frac{a}{q} \rightarrow q = ar^{-\alpha/a},$$

откуда $p = \frac{1}{2}(ar^{\alpha/a} + ar^{-\alpha/a})$, где a – единица, $\alpha/a = \alpha$, т. е. α таково, каким и должно быть. Но если вернуться к экспоненциальной кривой, то Q нужно будет искать как $\log ar^{\alpha} = \log a + \alpha \log r = 0 + \alpha \log r = \alpha \log r$, поэтому отношение $r = d/k$ должно быть основанием логарифмов, а чтобы кривая была настоящей цепной линией, ее форма, как мы уже упоминали, должна описываться уравнением $y = \frac{1}{2}(e^{x/a} + e^{-x/a})$, так что основание d/k должно быть равно e. Конечно, Лейбниц это знал и в последующей подробной переписке с немецким бароном Рудольфом Кристианом фон Боденхаузеном, датированной августом 1691 г., описал свои новые методы математического анализа, лежащие в основе построения, а также признался, что опустил одну деталь построения: $d/k = 2.7182818$.

И это все, что мы хотели сказать о решении Лейбница. А как насчет Бернулли? Как мы отметили выше, первоначально его решение было представлено в таком же непрозрачном виде, как решение Лейбница, но его методы были раскрыты в заметках, составленных для маркиза де Лопиталя на основе «Лекций по интегральному исчислению», прочитанных в 1691–92 гг., когда он жил в Париже. Начнем с необходимой элементарной задачи по статике, которая позволяет вывести уравнение цепной линии.

Обозначим A нижнюю точку свисающей цепи, а P – точку, удаленную от нее на расстояние s вдоль цепи, тогда участок AP находится в равновесии под действием трех сил, показанных на рис. 9.9: горизонтальной силы натяжения T_0, приложенной в точке A, его веса mg (а поскольку цепь однородна, $mg = k_1 s$) и касательной силы T, приложенной в точке P, которая составляет некоторый угол с горизонталью. Из уравнений равновесия получаем $T_0 = T \cos \psi$ и $k_1 s = T \sin \psi$, откуда сразу же следует, что $s = k_2 \operatorname{tg} \psi$ – внутреннее уравнение кривой, и, полагая $dy/dx = \operatorname{tg} \psi$, получаем $dy/dx = ks$.

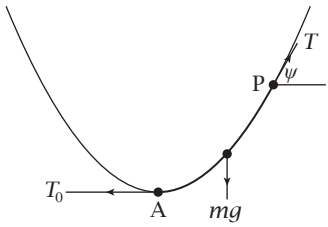


Рис. 9.9. Статика свисающей цепи

Для решения этого простого на первый взгляд дифференциального уравнения придется потрудиться, но, прежде чем рассказать, как Бернулли подошел к нему, разберемся с природой «давления» Гюйгенса, из-за которого свисающая цепь принимает форму параболы. Вместо того чтобы предполагать, что цепь тяжелая, предположим напротив, что она легкая, и через равные интер-

валы, измеряемые по горизонтали, к ней подвешены одинаковые веса; т. е. вес нагруженной цепи на расстоянии x по горизонтали от точки A равен k_1x. Теперь имеем $T_0 = T\cos\psi$ и $mg = k_1x = T\sin\psi$, откуда $x = k_2\,\mathrm{tg}\,\psi$, и, снова полагая $\mathrm{d}y/\mathrm{d}x = \mathrm{tg}\,\psi$, имеем $\mathrm{d}y/\mathrm{d}x = kx$, и значит, $y = kx^2 + c$. Парабола.

Что же до Бернулли, то он обозначал оси противоположно современному соглашению, так что для него $\mathrm{d}y/\mathrm{d}x = a/s$. Чтобы построить кривую, ему нужно было исключить s, и для этого он поступил следующим образом: $\mathrm{d}y = a\,\mathrm{d}x/s$ и значит, $(\mathrm{d}y)^2 = a^2(\mathrm{d}x)^2/s^2$, а так как $(\mathrm{d}s)^2 = (\mathrm{d}x)^2 + (\mathrm{d}y)^2$, имеем

$$(\mathrm{d}s)^2 = \frac{s^2(\mathrm{d}x)^2 + a^2(\mathrm{d}x)^2}{s^2}.$$

Это означает, что

$$\mathrm{d}s = \frac{\mathrm{d}x\sqrt{s^2 + a^2}}{s} \quad \text{и} \quad \mathrm{d}x = \frac{s\,\mathrm{d}s}{\sqrt{s^2 + a^2}},$$

что после интегрирования дает

$$x = \sqrt{s^2 + a^2} \quad \text{и} \quad s = \sqrt{x^2 - a^2}.$$

Дифференцируем:

$$\mathrm{d}s = \frac{x\,\mathrm{d}x}{\sqrt{x^2 - a^2}} = \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2}.$$

Возводя в квадрат и приводя подобные члены, получаем $x^2(\mathrm{d}y)^2 - a^2(\mathrm{d}y)^2 = a^2(\mathrm{d}x)^2$ и наконец

$$\mathrm{d}y = a\frac{\mathrm{d}x}{\sqrt{x^2 - a^2}}.$$

Для современного математика это дифференциальное уравнение, которое теперь нужно решить, – промежуточный шаг на пути к ответу. А тогда это была формула, которая показывала, как приращение y кривой связано с приращением x, а значит, давала способ поточечного построения графика кривой. С точки зрения математика XVII в., решение уже найдено. А как же быть с явным уравнением кривой? Для этого придется подождать еще 70 лет и понаблюдать за усилиями еще одного корифея.

9.4. Гиперболические функции

Иррациональность e была установлена в 1737 г. Эйлером (хотя этот результат он опубликовал только в 1742 г.), в связи с чем природа гораздо более старой математической постоянной π не давала покоя. Поэтому математики дружно праздновали, когда в конце XVIII в. Иоганн Ламберт доказал иррациональность π и закрыл вопрос. Простим ему, что он громогласно объявил об этом результате в 1761 г., хотя выглядел он как скромные первые две трети статьи, представленной Берлинской академии наук (Lambert 1761).

Последняя треть была посвящена совершенно другим вопросам, относящимся к сравнению *quantitiés circulaire* и *quantitiés transcendentes logarithmique*, – сейчас

мы употребляем термины *круговые* и *гиперболические функции*. Он не первым оставил свой след на этом песке: Эйлер (ну разумеется!) рассматривал выражения $\frac{1}{2}(e^x \pm e^{-x})$, но интересовался ими как лишь средством достижения цели, используя комплексные степени для вывода разложений синуса и косинуса в бесконечные произведения; еще один заметный вклад внес человек, воинские заслуги которого более значительны, чем его вклад в математику. Шевалье Франсуа Давье де Фонсене предположил, что связь между тем, что сейчас называется круговыми и гиперболическими функциями, проще установить, сравнив единичную окружность с единичной гиперболой. В основном он интересовался спорным вопросом о том, какой смысл можно приписать логарифмам отрицательных чисел, что неизбежно вело к комплексным числам – и неудобной идее отношения $1:\sqrt{-1}$. У Ламберта интерес был другой: он хотел сравнить то, что мы назвали бы натуральной параметризацией окружности и гиперболы. Поэтому прежде чем переходить к деталям рассуждения Ламберта, поясним, что в нашей терминологии его целью было найти параметризацию $x = \frac{1}{2}(e^u + e^{-u})$, $y = \frac{1}{2}(e^u - e^{-u})$ стандартной гиперболы $x^2 - y^2 = 1$, сравнить ее со стандартной параметризацией $x = \cos u$, $y = \sin u$ единичной окружности $x^2 + y^2 = 1$ и таким образом произвести на свет гиперболические функции, а затем изучить их свойства и наделить их тем, что он считал естественной нотацией. Итак, мы переходим к последней трети статьи 1761 г.

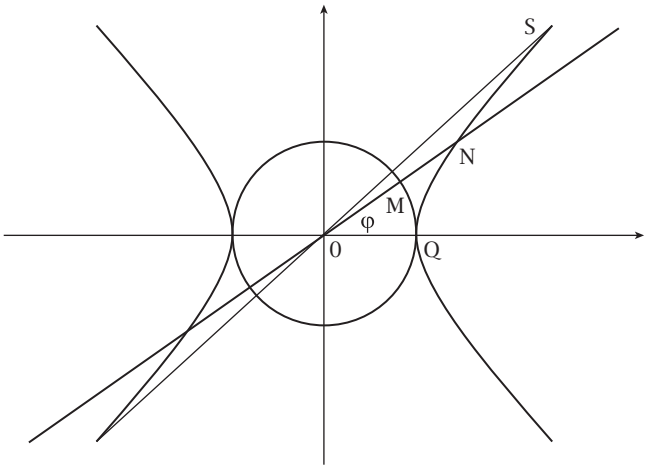


Рис. 9.10. Единичные окружность и гипербола

Поскольку нас интересует взаимосвязь между координатами точек на единичной окружности $x^2 + y^2 = 1$ и единичной гиперболе $x^2 - y^2 = 1$, то было бы естественно нарисовать их на одном чертеже; именно это мы и сделали на рис. 9.10, взятом из самой статьи Ламберта. Кроме двух конических сечений, мы видим прямую, проходящую через начало координат (с коэффициентом наклона меньшим 1), которая составляет угол φ с положительным направлением оси x и пересекает окружность в точке M, а гиперболу – в точке N. Имеется также прямая, изображенная более тонкой линией, с большим коэффициентом наклона, которая пересекает гиперболу в точке S; зачем она нужна, мы скоро узнаем. Обозначив вслед за Ламбертом координаты N (ξ, η), будем иметь $\operatorname{tg}\varphi = \eta/\xi$, и из определения гиперболы:

$$1 + \eta^2 = \xi^2 = \eta^2 \operatorname{ctg}^2\varphi \text{ и } \xi^2 - 1 = \xi^2 \operatorname{tg}^2\varphi.$$

Тогда имеем

$$\xi = \frac{1}{\sqrt{1 - \operatorname{tg}^2 \varphi}} \quad \text{и} \quad \eta = \frac{\operatorname{tg} \varphi}{\sqrt{1 - \operatorname{tg}^2 \varphi}}.$$

Мы получили параметризацию гиперболы в терминах угла φ, но это не то, что нам нужно. Чтобы продвинуться дальше, обозначим A площадь области ONQ между прямой, гиперболой и осью x и определим параметр $u = 2A$.

Мы уже имеем

$$\mathrm{ON}^2 = \xi^2 + \eta^2 = \frac{1}{1 - \operatorname{tg}^2 \varphi} + \frac{\operatorname{tg}^2 \varphi}{1 - \operatorname{tg}^2 \varphi} = \frac{1 + \operatorname{tg}^2 \varphi}{1 - \operatorname{tg}^2 \varphi},$$

и отсюда Ламберт сделал важный шаг по переходу к бесконечно малым. Вторая прямая OS круче ON на угол $\mathrm{d}\varphi$, и ONS можно аппроксимировать сектором круга радиуса ON (OS). Поскольку площадь сектора круга равна $\frac{1}{2}r^2\theta$, то

$$\mathrm{d}A = \tfrac{1}{2}\mathrm{ON}^2 \mathrm{d}\varphi \to \tfrac{1}{2}\,\mathrm{d}u = \frac{1}{2}\frac{1 + \operatorname{tg}^2 \varphi}{1 - \operatorname{tg}^2 \varphi}\,\mathrm{d}\varphi.$$

Поэтому

$$\mathrm{d}u = \frac{1 + \operatorname{tg}^2 \varphi}{1 - \operatorname{tg}^2 \varphi}\,\mathrm{d}\varphi.$$

Согласно правилу дифференцирования сложной функции,

$$\begin{aligned}\frac{\mathrm{d}\xi}{\mathrm{d}u} &= \frac{\mathrm{d}\xi}{\mathrm{d}\varphi}\frac{\mathrm{d}\varphi}{\mathrm{d}u} \\ &= -\frac{1}{2} \times \frac{1}{(1 - \operatorname{tg}^2 \varphi)^{3/2}} \times -2 \operatorname{tg} \varphi \sec^2 \varphi \times \frac{1 - \operatorname{tg}^2 \varphi}{1 + \operatorname{tg}^2 \varphi} \\ &= \frac{\operatorname{tg} \varphi}{\sqrt{1 - \operatorname{tg}^2 \varphi}} = \eta.\end{aligned}$$

Аналогично $\mathrm{d}\eta/\mathrm{d}u = \xi$, и мы имеем систему из двух дифференциальных уравнений. Сегодня мы продифференцировали бы и получили

$$\frac{\mathrm{d}\xi}{\mathrm{d}u} = \eta \to \frac{\mathrm{d}^2\xi}{\mathrm{d}u^2} = \frac{\mathrm{d}\eta}{\mathrm{d}u} = \xi,$$

а дальше использовали бы теорию дифференциальных уравнений, которой тогда не существовало. Вместо этого Ламберт представил ξ и η бесконечными рядами от u:

$$\xi = 1 + Au^2 + Bu^4 + Cu^6 + \ldots \text{ и } \eta = au + bu^3 + cu^5 + du^7 + \ldots,$$

аргументируя это тем, при $u = 0$ $\xi = 1$ и $\eta = 0$. А также тем, что если зеркально отразить прямую относительно оси x, то u становится отрицательным, ξ не изменяется, а η становится отрицательным, поэтому степени именно таковы, как показано выше. Дифференцируя один ряд и приравнивания результат другому, получаем:

$$\xi = 1 + \frac{u^2}{2!} + \frac{u^4}{4!} + \frac{u^6}{6!} + \cdots \quad \text{и} \quad \eta = u + \frac{u^3}{3!} + \frac{u^5}{5!} + \frac{u^7}{7!} + \cdots .$$

Зная ряд

$$e^u = 1 + u + \frac{u^2}{2!} + \frac{u^3}{3!} + \cdots ,$$

он смог написать

$$\xi = \tfrac{1}{2}(e^u + e^{-u}) \quad \text{и} \quad \eta = \tfrac{1}{2}(e^u - e^{-u}).$$

Требуемая параметризация гиперболы получена, и по ходу дела родились две новые функции, которые имели естественные аналогии с sin *u* и cos *u* и заслуживали каких-то имен. Ламберт окрестил их в статье 1768 г., когда вернулся к трансцендентным логарифмическим функциям и продолжил изучать их сходство с круговыми. Если мы вернемся к рис. 9.10 и взглянем на окружность, то увидим, что

$$\operatorname{tg}\varphi = \frac{MP}{OP} = \frac{\sin\varphi}{\cos\varphi},$$

тогда как для гиперболы

$$\operatorname{tg}\varphi = \frac{NR}{OR} = \frac{\xi}{\eta} = \frac{\operatorname{sinhyp}\varphi}{\operatorname{coshyp}\varphi}.$$

«По этой причине я буду называть абсциссу гиперболическим косинусом, а ординату – гиперболическим синусом», – писал Ламберт.

Короче говоря, таблицы круговых функций можно было заменить таблицами гиперболических. Но чтобы новый инструмент был полезен, требовалось составить эквивалент списка тригонометрических тождеств, которого так страшатся многие ученики старших классов, и Ламберт занялся этим делом в своей статье. В двух колонках перечисляются тригонометрические тождества и их гиперболические аналоги; многие из них, как он заметил, в точности совпадают, тогда как другие отличаются только знаком, например: $\cos^2 y - \sin^2 y = \cos 2y$ и $\operatorname{coshyp}^2 y + \operatorname{sinhyp}^2 y = \operatorname{coshyp} 2y$. Также он заметил, что это сходство распространяется и на результаты из математического анализа:

$$\frac{d}{dx}\operatorname{coshyp} x = \operatorname{sinhyp} x \quad \text{и} \quad \frac{d}{dx}\operatorname{sinhyp} x = \operatorname{coshyp} x.$$

По прошествии некоторого времени были приняты обозначения $\cosh x = \frac{1}{2}(e^x + e^{-x})$ и $\sinh x = \frac{1}{2}(e^x + -e^{-x})$:[1] не было стандартизовано только произношение. И форма свисающей цепи, как выяснилось, описывается вовсе не простой параболой, а трансцендентной функцией $y = \operatorname{ch} x = \frac{1}{2}(e^x + e^{-x})$ или, более общо, $y = a \operatorname{ch}(x/a)$, что дает нам важную аналогию с круговыми функциями.

В англоязычной литературе эта кривая называется *catenary*, но такое название появилось не сразу, а чтобы понять, как оно возникло, нам нужно перевернуть кривую вверх ногами.

[1] А в русскоязычной литературе еще короче: ch *x* и sh *x*. – *Прим. перев.*

9.5. Перевернутая цепь

Читатель, знакомый с законом упругости Гука (сила упругости пропорциональна деформации), удивится, что впервые он появился в форме анаграммы *ceiiinosssttuu*, которую и сейчас-то расшифровать трудно, а что уж говорить о XVII в. В эпоху ученых-энциклопедистов Роберт Гук был одним из самых выдающихся, и столь же выдающееся положение занимал среди научных сварливцев, по крайней мере в зрелые годы. Один из членов-основателей Королевского общества, он в 1665 г. был назначен куратором экспериментов и одновременно был профессором геометрии в Грэшем-колледже, топографом и весьма авторитетным архитектором, спроектировавшим ряд величественных зданий[1] и помогавшим Кристоферу Рену перестраивать Лондон после Большого пожара. В общем, занятой человек. Что до этих выстроенных в алфавитном порядке букв, то так он заявил свои интеллектуальные права на важный результат, который, однако, ждал своей очереди на публикацию и подробное объяснение. А очередь эта была велика. Эта анаграмма находится в конце 32-страничной статьи 1676 г., которую Гук озаглавил «Описание гелиоскопов и некоторых других приборов». Покончив с объяснениями различных научных приборов, он пишет:

> «Чтобы заполнить пустующую страницу, я добавил десяток из сотни Изобретений, которые я намерен опубликовать, хотя, возможно, не в таком порядке, а появятся у меня на то возможность и досуг; большинство из которых, я надеюсь, будут настолько же полезны человечеству, насколько сейчас они неизвестны и новы».

При таких-то достоинствах вряд ли можно счесть справедливым, что в тот же год он был выведен под именем сэра Николаса Гимкрака в качестве персонажа снискавшей успех комедии Томаса Шедуэлла «Виртуоз»; в этой сатире на Королевское общество были подвергнуты критике сумасбродства ученых мужей, преследующих бессмысленные научные цели в своих экспериментах. Не было ничего бессмысленного в третьем из этого десятка изобретений:

> «Истинная теория упругости или гибкости и частные проявления ее в различных предметах, в которых она обнаруживается, а также способ вычисления скорости тел, двигаемых ими. Ceiiinosssttuu».

Только через два года, в 1678 г., он расшифровал эту анаграмму как *ut tensio, sic vis*, или «каково удлинение, такова и сила». Интерпретируя эти слова как «удлинение пропорционально силе», мы узнаем в них знакомый закон Гука. А вот еще одна анаграмма в статье 1676 г., сопровождавшая второе из десятка описаний, заполнивших страницу:

> «Истинная Математическая и Механическая форма всех видов Арок для Возведения Зданий с описанием необходимых для них устоев. Проблема, к которой до сих пор не подступался, а тем более не решил никто из пишущих об Архитектонике, abccc ddeeeee f gg iiiiiiii llmmmmnnnnnooprr sssttttttuuuuuuuuux».

[1] Например, Королевский физический колледж, дом Монтагу и (хоть не столь благородно звучащую) Бетлемскую королевскую больницу, в просторечии Бедлам.

В этом случае Гук так и не представил решения загадки, которую разрешил некто Ричард Уоллер уже после смерти Гука, в 1703 г. В толстенном 570-страничном труде Уоллера, содержащем описание жизни и неопубликованные работы Гука (Waller 1705), мы находим расшифровку *Ut pendet continuum flexile, sic stabit contiguum rigidum inversum*, что в переводе означает «так же, как провисает гибкая нить, возводится твердая арка, только перевернутая». В комментариях Уоллер пишет, что это «линейная катенария» (*Linear Catenaria*). То есть Гук утверждал, что идеальная форма арки совпадает с формой свободно свисающей цепи, только перевернутой. И далее он замечает, что строительные материалы очень хорошо сопротивляются сжатию, но сравнительно плохо растяжению – прямая противоположность свисающей веревке: веревка провисает под действием чистого растяжения; переверните ее, измените материал, и конструкция будет поддерживать сама себя под действием чистого сжатия. Идеал приближается к реальности, когда на цепи, провисающей как цепная линия, закреплены грузы, заставляющие ее принимать форму, показанную на рис. 9.11, когда каждый участок цепи по отдельности подвергается растяжению. Замените участки цепи стержнями и переверните фигуру – получится рис. 9.12. Теперь на стержни действует боковые силы, и на рисунке показаны линии распора конструкции: если они находятся внутри нее, то все хорошо – иначе… Это не стало бы неожиданностью для шотландского математика Дэвида Грегори, который в 1697 г., т. е. до расшифровки Уоллера, предложил свое не вполне точное доказательство, основанное на анализе флюент и флюксий, в котором установил, что свисающая цепь принимает форму катенарии (цепной линии). Он также добавил такой комментарий:

> «В вертикальной плоскости, но будучи перевернутой, цепь сохранит свою форму, не падая, и потому будет образовывать очень тонкую арку, или форникс[1]: то есть бесконечно малые, твердые, отполированные сферы, расположенные в ряд на перевернутой катенарии, образуют арку, ни одна часть которой не будет выпирать наружу или внутрь под действием других частей, но если самые нижние части остаются прочными, то она будет поддерживать себя в силу своей формы. <…> Никакая кривая, кроме катенарии, не является формой истинной и правильной арки или форникса. А коли не падают арки другой формы, то это потому, что в их толще скрыта какая-то катенария».

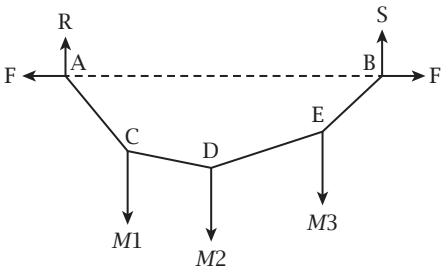


Рис. 9.11. Цепь с грузами

[1] Сводчатая или арочная конструкция.

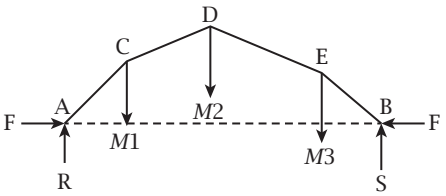


Рис. 9.12. Перевернутая цепь с грузами

В архивах Королевского общества есть записи о том, что Гук и сэр Кристофер Рен обещали представить математические доказательство в дополнение к физическим демонстрациям, которые провели для подкрепления своих теорий, но так и не представили. И если только в конце 1675 или в начале 1676 г. Гука не посетило какое-то озарение, то включение задачи о свисающей цепи в список ожидания было вызвано необходимостью: в его потаенном личном дневнике мы встречаем фразу «Загадка арки, свисающей гибкой нити (pendet continuum flexile)...» в записи от 26 сентября 1675 г. в воскресенье. Она наводит на мысль о том, что полное математическое обоснование по-прежнему не давалось ему. Но теория была вторична по отношению к практике, что и было продемонстрировано «бессмысленными» экспериментами в Королевском обществе, чему порукой конструкция купола (на самом деле одного из трех куполов) собора Святого Павла в Лондоне.

Быть может, самый известный пример практического применения этого принципа – анализ безопасности треснувшего купола базилики Святого Петра. Состояние конструкции здания давно вызывало опасения, и после нескольких лет исследований, ознакомившись с отчетами, папа Бенедикт XIV в 1743 г. пригласил знаменитого ученого Джованни Полени в Рим, чтобы тот изложил свое мнение. Полени пришел к выводу, что купол безопасен, для чего разделил его по вертикали на сходящиеся клинообразные элементы (дольки апельсина), как показано в верхней части его рисунка XIV (воспроизведенного на рис. 9.13), и взял легкую цепь, на которой закрепил 32 груза с весами, пропорциональными весам соответствующих участков элемента. В нижней части рис. 9.13 изображена цепь, свисающая, как идеальная цепная линия, а также цепь с грузами, показанными в виде кружочков; размер каждого кружочка обозначает его вес, и заметим, что самый большой находится в нижней части цепи, потому что соответствующие участки поддерживают гигантский фонарь. Пунктирная прямая в верхней части рисунка – положение цепи с грузами внутри клинообразного элемента; она полностью лежит внутри него. Купол оказался безопасен – он и до сих пор стоит. Этот исторический принцип важен и по сей день и формулируется в виде одной из *теорем взаимности* в строительной механике – *теоремы Хеймана о безопасности*, которая была сформулирована в 1966 г. и названа в честь выдающегося инженера Жака Хеймана.

Совсем не трудно найти примеры арок в виде цепной линии – имеющих такую форму ввиду требований конструкции или за ее архитектурную красоту, но мы избежим соблазна привести их длинный список, а также тягостной обязанности выбирать какую-то одну в качестве примера. Поклонник сюрреализма, архитектор Антонио Гауди пользовался своим знакомством с геометрическими

формами при создании многих функциональных, но при этом эстетически привлекательных и эксцентричных строений: церковь Саграда Фамилия и дом Каса Мила в Барселоне – два здания, в которых цепная линии встречается очень часто. Еще одно – церковь колонии Гуэль, в которой была завершена только крипта. В музее Саграда Фамилия есть удивительный экспонат – уменьшенная модель церкви, построенная из утяжеленных мешочками с дробью цепей, свисающих с потолка соседнего помещения, на котором он нарисовал план церкви: все это было сфотографировано, и фотография перевернута, чтобы показать общий вид, а дополнительно Гауди раскрасил ее, чтобы церковь предстала во всем своем воображаемом великолепии. Дух Полени присутствует повсюду.

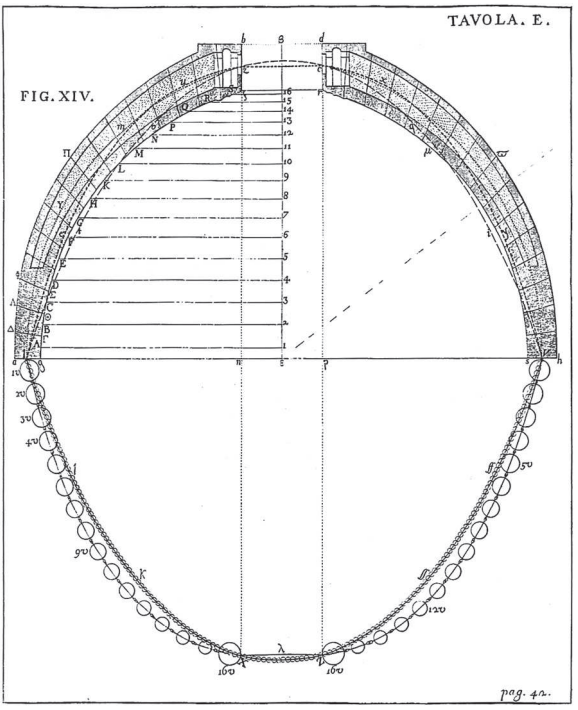


Рис. 9.13. Полени и базилика Святого Петра

Мы видели, что термин *catenaria* использовался с 1690 г., когда Гюйгенс придумал это слово в письме Лейбницу; неудивительно, что Гюйгенс образовал его от латинского *catena*, означающего *цепь*. Эта форма известна также под именами *chainette*, *chaine (corde) pendante* и несколькими другими, но переход к современной форме *catenary*, пусть и семантически незначительный, имеет совершенно неожиданный источник. Там, где есть арка, может быть и мост, и именно с ролью цепной линии в конструировании мостов связано первое упоминание ее современного названия. В 1788 г. Томас Джефферсон исправлял должность министра финансов США в качестве ступеньки на пути к избранию третьим президентом страны в 1801 г. Его командировали в Европу для участия в торговых переговорах, и остановился он в Париже, а человек, с которым он часто переписывался, политический активист и изобретатель Томас

Пейн, находился в Англии. Уже давно Пейн считал, что дерево и камень – материалы, из которых веками строились мосты, – следует заменить металлом, поэтому принял участие в проектировании железного места через реку Уир в городе Сандерленд. Это исследование вызвало интерес у разносторонне образованного Джефферсона, мнение которого было, в свою очередь, интересно Пейну. В переписке на тему будущей формы моста Джефферсон писал из Парижа 23 декабря 1788 г.:

> «Вы колеблетесь между цепной линией (catenary) и частью окружности, я недавно получил из Италии трактат о равновесии арок, написанный аббатом Маскерони. Выглядит как весьма ученая работа, с которой я еще не успел внимательно ознакомиться; но я нахожу, что из его демонстраций следует, что каждая часть цепной линии пребывает в идеальном равновесии. Но тогда было бы замечательно взять в новом эксперименте одну арку, так чтобы давление было одинаково в каждой ее точке».

Нашел ли этот необычайно занятой человек время изучить сообщение итальянского математика и священнослужителя Лоренцо Маскерони, мы не знаем, но работа «Nuove ricerche sull' equilibrio delle volte» (Новое исследование о равновесии сводов), вышедшая в 1785 г., насчитывает 144 страницы и, понятное дело, написана по-итальянски. Но похоже, что благодаря этому абзацу, написанному отцом-основателем США и главным автором Декларации независимости, слово «catenary» вошло в обиход как название формы, но не свисающей цепи, а неотъемлемой части архитектуры металлического моста[1].

Обоснование Маскерони использования цепной линии в качестве оптимальной формы арки встречается в виде задачи XIII в главе 2 его работы после тщательной подготовки математической конструкции, на которую опиралось рассуждение: его заключительное предложение имеет вид $dx : dy = s : a$, *уравнение цепной линии*.

Но мы закончим на негативной ноте и еще одним упоминанием Томаса Джефферсона. Знаменитые «Врата на Запад» – арка в Сент-Луисе, штат Миссури, – главная составная часть проекта, завоевавшего первое место в конкурсе 1947 г. на создание Национального мемориала Джефферсону, в память его «Луизианской покупки» у Франции большого участка земли, примерно удвоившего размер Америки. Эту арку часто называют впечатляющим примером использования перевернутой цепной линии, но на самом деле ее форма сложнее, чем может показаться. И цепной линии там нет. Во всяком случае, чистой. Ээро Сааринен, архитектор проекта, только добавил путаницы в своих комментариях[2]:

> «Арка не является ни параболой, ни цепной линией. Сначала мы работали с математическими фигурами, но в конечном итоге подправили на глазок. Я подозреваю, однако, что линия, образуемая цепью, звенья которой уменьшаются пропорционально сужению арки, оказалась бы очень близка к линии, на которой мы остановились… Арка не является истинной

[1] В русском языке слово «катенария» не употребляется (по крайней мере, в математике и строительстве). Хотя тело, образованное вращением цепной линии вокруг оси, называется катеноидом. – *Прим. перев.*

[2] Их можно найти в архиве Сааринена, хранящемся в Йельском университете.

> параболой, как часто утверждают. На самом деле она имеет форму цепной линии – формы, которую принимает свисающая цепь, – кривой, для которой равнодействующая распирающих сил расположена посередине между стойками арки. <...> [Арка является] абсолютно чистой формой, для которой линия сжатия проходит точно через центр линии конструкции перпендикулярно земле. Иными словами, это идеальная цепная линия».

На самом деле поперечное сечение арки на каждом уровне представляет собой равносторонний треугольник, размеры эти треугольников постепенно уменьшаются, а центры тяжести лежат на кривой, описываемой уравнением

$$y = 693.8597 - 68.7672 \operatorname{ch} 0.0100333x,$$

где $-299.2239 \leqslant x \leqslant 299.2239$[1]. Поскольку коэффициенты перед ch и перед x не равны, это уравнение не чистой цепной линии, а перевернутой *нагруженной*, или *сплюснутой*, цепной линии, общее уравнение которой имеет вид:

$$y = c - a\cosh\frac{x}{b} = c - \frac{a}{b}\left(b\cosh\frac{x}{b}\right).$$

Это перевернутая цепная линия, растянутая по вертикали с коэффициентом a/b и сплюснутая при условии, что $a/b < 1$. В данном случае a/b = 68.7672×0.0100333 = 0.6899619, т. е. чистая цепная линия сплюснута с коэффициентом 0.69; иначе говоря, она сплюснута по вертикали чуть меньше, чем на треть. Не чистая, но все же красивая.

9.6. Ухабистая дорога

Наше последнее путешествие по миру цепной линии уведет нас от ее роли в проектировании и возведении арок удивительной красоты и значимости к теоретически мыслимой, но весьма причудливой поверхности дороги.

Чтобы поднять наш центр тяжести, необходима работа, а для нее нужна энергия. Поэтому лучше бы не поднимать. Ясное дело, мы считаем само собой разумеющимся, что наш центр масс остается на постоянной высоте, когда мы едем на велосипеде по горизонтальной дороге, но это потому, что мы считаем само собой разумеющимся, что колеса велосипеда круглые. А если нет? Предположим, к примеру, что они квадратные. Какой тогда должна быть форма поверхности дороги, чтобы мы могли ехать «нормально»? Ответ, разумеется, – в форме цепной линии или последовательности цепных линий. Чтобы понять, почему это так, давайте сначала сформулируем критерии плавного движения.

- Центр колеса должен оставаться на постоянной высоте над горизонталью.
- В каждой точке контакта колесо должно касаться поверхности дороги.
- Центр колеса должен располагаться точно над точкой контакта.
- Расстояние между соседними участками поверхности дороги должно быть таким же, как длина стороны квадратного колеса.

[1] Единица измерения – фут.

Согласившись с этим, подведем итог на трех чертежах. На рис. 9.14 показано, как колесо завершает движение по одному участку дороги и начинает движение по следующему. Здесь обозначена ось x, длина стороны квадрата, равная $2a$, и постоянная высота h центра масс: понятно, что $h = a\sqrt{2}$.

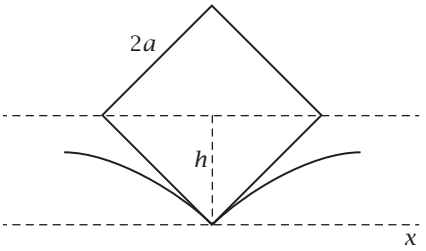


Рис. 9.14. Колесо на стыке двух участков

На рис. 9.15 показано колесо в верхней точке кривой, а также обозначена ось y и начало координат.

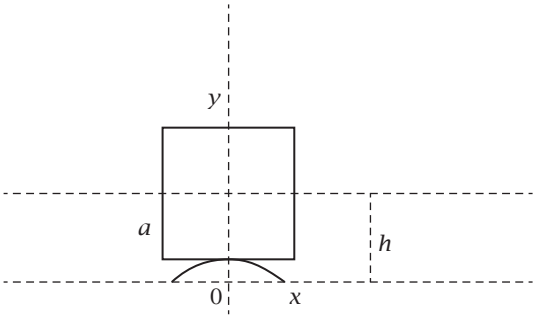


Рис. 9.15. Колесо в верхней точке

Наконец, на рис. 9.16 показано колесо в общем положении, когда оно составляет угол θ с горизонталью, а точка динамического контакта с кривой имеет координаты (x,y).

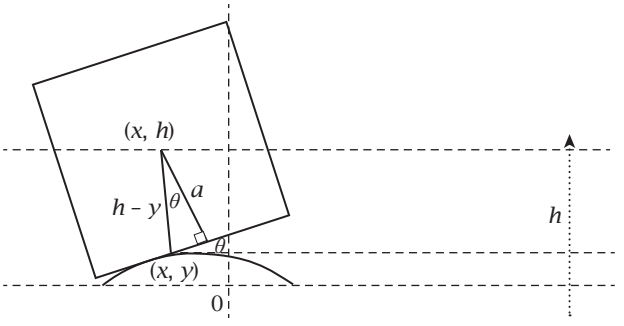


Рис. 9.16. Колесо в общем положении

Из рис. 9.16 имеем:

$$a = (h - y)\cos\theta = \frac{h - y}{\sec\theta} = \frac{h - y}{\sqrt{1 + \tan^2\theta}} = \frac{h - y}{\sqrt{1 + (\mathrm{d}y/\mathrm{d}x)^2}},$$

откуда

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \pm\sqrt{\left(\frac{h - y}{a}\right)^2 - 1},$$

и после интегрирования получаем

$$\int \frac{1}{\sqrt{(h - y)^2 - a^2}}\,\mathrm{d}y = \frac{1}{a}\int \mathrm{d}x,$$

$$-\cosh^{-1}\frac{y - h}{a} = \frac{x}{a} + c \rightarrow y = -a\cosh\left(\frac{x}{a} + c\right) + h.$$

Из рис. 9.15 видно, что, когда $x = 0$, $y = h - a$, откуда следует, что ch $c = 1$, так что $c = 0$. Тогда уравнение поверхности дороги имеет вид:

$$y = h - a\cosh\left(\frac{x}{a}\right) = a\sqrt{2} - a\cosh\left(\frac{x}{a}\right),$$

и мы имеем нашу (перевернутую) цепную линию.

Заметим, что в точке соединения двух участков дороги $y = 0$, и при $h = a\sqrt{2}$ имеем

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \pm\sqrt{\left(\frac{a\sqrt{2} - 0}{a}\right)^2 - 1} = \pm 1,$$

так что колесо точно касается дороги.

Что до длины участки, то если мы проинтегрируем на отрезке от $x = 0$ до $x = x_1$, где кривая соприкасается с горизонталью, то будем иметь

$$\begin{aligned} s &= 2\int_0^{x_1} \sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2}\,\mathrm{d}x = 2\int_0^{x_1} \sqrt{1 + \left(\frac{h - y}{a}\right)^2 - 1}\,\mathrm{d}x \\ &= 2\int_0^{x_1} \frac{h - y}{a}\,\mathrm{d}x = 2\int_0^{x_1} \cosh\frac{x}{a}\,\mathrm{d}x = 2\left[a\sinh\frac{x}{a}\right]_0^{x_1} \\ &= 2a\sinh\frac{x_1}{a}, \end{aligned}$$

где

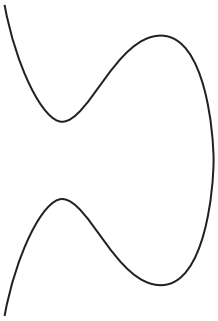
$$0 = a\sqrt{2} - a\cosh\left(\frac{x_1}{a}\right) \rightarrow \cosh\left(\frac{x_1}{a}\right) = \sqrt{2} \rightarrow \sinh\left(\frac{x_1}{a}\right) = 1.$$

Длина каждого участка в точности равна стороне квадрата – и мы можем ехать беспрепятственно. На самом деле колеса велосипеда могут иметь форму любого правильного выпуклого многоугольника – кроме треугольника, и, возможно, читатель захочет исследовать эту проблему.

На этом нетрадиционном средстве передвижения мы и добрались до конечной точки нашего путешествия.

Глава 10

Эллиптические кривые



Почему именно эти кривые?

Эллиптические кривые носят вводящее в заблуждение имя, а их история тянется со времен Древней Греции, через Индию и Аравийский полуостров в Европу, где наконец вливается в историю современной математики. Они встречаются в классической физике при анализе простого маятника (см. приложение F) и в современной физике, точнее теории струн, в топологии – в экзотической теории кобордизмов – и в теории чисел – от до сих пор не решенной обратной задачи Галуа до решения диофантовых уравнений (включая Великую теорему Ферма) и разложения больших целых чисел на простые множители. Они лежат в основе самого эффективного из современных компьютерных методов шифрования, и не будет преувеличением предположить, что их роль в современной науке обеспечивает им место в сонме самых важных кривых XXI в. Короче говоря, это многогранный математический инструмент, швейцарский нож среди математических кривых. Серж Лэнг, математик и весьма плодовитый автор, начал вступительное слово к одной из своих многочисленных монографий словами «Об эллиптических кривых можно писать бесконечно (это не угроза)» (Lang 1978): в нашем же распоряжении не так много страниц, чтобы рассказать историю, которая могла бы занять целые тома.

10.1. Эллиптическая неоднозначность

Мы начнем эту главу с уклончивых рассуждений на тему определения кривой, являющейся предметом нашего рассмотрения. Что такое эллиптическая кривая? Предлагаем три определения, которые удовлетворят только ученого читателя.

- Эллиптическая кривая E над полем K – это несингулярная кубическая кривая E над K вместе с заданной точкой $O \in E(K)$.
- Эллиптическая кривая – это проективное многообразие, изоморфное несингулярной кривой степени 3 в P^2 вместе с выделенной точкой $O \in E$.
- Эллиптическая кривая – это несингулярная проективная кривая рода 1, являющаяся абелевым многообразием относительно групповой операции.

Отсюда можно сделать вывод, что эллиптические кривые могут быть весьма сложными. Посыл один и тот же, хотя формулируется он разными словами, смысл которых нас не особенно волнует. Но все они говорят, что эллиптическая кривая – частный случай кубической кривой (не следует думать, что это кривая, описываемая кубической функцией от x) и, более того, что она естественно располагается в проективном пространстве, что позволяет связать с ней «бесконечно удаленную точку» и тем самым определить естественную алгебраическую структуру. Но это еще не все: это решительно нескладное подынтегральное выражение, особенно над множеством комплексных чисел, которое, следовательно, определяет *эллиптический интеграл*, обратная функция к которому имеет важнейшее значение. И к этим наблюдениям мы можем добавить еще кое-что. Главная трудность при определении кривой заключается в том, что определение зависит от того, кому задан вопрос, и мы постараемся избежать всех тонкостей, ограничившись определением, которое можно назвать «попсовым», но для наших целей его вполне достаточно. Поступая так, мы воспользуемся лишь немногими из инструментов, имеющихся в швейцарском ноже. Но начнем мы с бестолкового имени, потому что чем бы ни была эллиптическая кривая, это точно не эллипс.

Стандартное уравнение эллипса с большой полуосью a и малой полуосью b – $x^2/a^2 + y^2/b^2 = 1$, и, чтобы вычислить его площадь по площади единичного круга, нужно всего лишь произвести масштабирование: растяжение с коэффициентом a по оси x дает площадь πa, а последующее растяжение с коэффициентом b по оси y дает окончательную площадь πab. Можно вместо этого выполнить простое интегрирование, но ни линейное растяжение в двух направлениях, ни интегрирование не дадут нам элементарного выражения длины эллипса. Мы, конечно, можем найти для нее аппроксимацию, как поступил Исаак Ньютон, а вслед за ним Джон Валлис в XVII в., применив бесконечные ряды. Или как Рамануджан (Ramanujan 1962)[1] в веке двадцатом, который вывел фантастическое выражение длины эллипса:

$$\pi\left\{(a+b)+\frac{3(a-b)^2}{10(a+b)+\sqrt{a^2+14ab+b^2}}+\varepsilon\right\},$$

где $\varepsilon \sim 3ak^{20}/68719476736$. Или как многие другие между этими двумя математическими гениями. И тем не менее выражение в замкнутой форме – т. е. содержащее только обычные комбинации многочленов, рациональных, тригонометрических, логарифмических и экспоненциальных функций – отыскать невозможно (хотя строго этот факт был установлен только в 1883 г.)[2]. Так что же

[1] Полный вывод см. в работе Villain (2008).

[2] Доказано в работе Liouville (1833).

представляет собой этот особенно трудный интеграл? В декартовых координатах формула длины дуги имеет вид:

$$\int \sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2}\,\mathrm{d}x,$$

и после несложных алгебраических преобразований мы приходим к формуле длины эллипса:

$$4\int_0^a \frac{\sqrt{a^2 - k^2x^2}}{\sqrt{a^2 - x^2}}\,\mathrm{d}x,$$

где $k = \sqrt{a^2 - b^2}/a$ – эксцентриситет эллипса; эта величина не зависит от размера эллипса и описывает его форму, оставляя описание размера на долю числа a. Подставляя $x = at$, получаем формулу:

$$4a\int_0^1 \frac{\sqrt{1 - k^2t^2}}{\sqrt{1 - t^2}}\,\mathrm{d}t,$$

и если мы обозначим

$$\pi(k) = 2\int_0^1 \frac{\sqrt{1 - k^2t^2}}{\sqrt{1 - t^2}}\,\mathrm{d}t,$$

то длина эллипса с большой полуосью a и эксцентриситетом k будет равна $2\pi(k)a$, что обобщает формулу длины окружности, $2\pi r$. Если освободить числитель от иррациональности, то подынтегральное выражение примет вид $(1 - k^2t^2)/\sqrt{(1 - t^2)(1 - k^2t^2)}$; такие выражения обычно записываются в виде $R(t, \sqrt{(1 - t^2)(1 - k^2t^2)})$, это пример рационального выражения, полученного из t и квадратного корня с применением операций сложения, вычитания, умножения и деления, а они, в свою очередь, являются частными случаями более общего выражения $R(t, \sqrt{at^4 + bt^3 + ct^2 + dt + e})$. Многочлен четвертой степени под знаком корня (который путем замены переменной эквивалентен кубическому) резко отличается от любого выражения вида $R(t, \sqrt{at^2 + bt + c})$, которое всегда можно проинтегрировать (зачастую путем остроумной подстановки $u = \mathrm{tg}\,½t$): тот факт, что выражение

$$\frac{t^2 + t + 1 - (t^6 + t^5 + 2)\sqrt{t^2 + 2t + 1}}{t^4 - 6\sqrt{t^2 + 2t + 1}}$$

можно проинтегрировать в элементарных функциях, а $1/\sqrt{t^3 - 1}$ нельзя, – свидетельство непредсказуемой природы интегрирования[1].

Но как же справиться с этими сложными интегралами? С этим вопросом связан длинный список известных имен, но самым выдающимся среди них был француз Адриен-Мари Лежандр. Энциклопедический трехтомный труд, выходивший в 1811–1816 гг. и являющийся кульминацией 40 лет работы, включал и тот факт, что все множество интегралов такого вида может быть све-

[1] Но лучше оставить эту работу компьютеру.

дено к комбинации рациональных функций, элементарных трансцендентных функций и трех дополнительных *эллиптических интегралов*:

- первого рода, $\int \frac{1}{\sqrt{(1-t^2)(1-k^2t^2)}}\,\mathrm{d}t$;
- второго рода, $\int \sqrt{\frac{1-k^2t^2}{1-t^2}}\,\mathrm{d}t$;
- третьего рода, $\int \frac{1}{(1-nt^2)\sqrt{(1-t^2)(1-k^2t^2)}}\,\mathrm{d}t$.

Неудивительно, что вследствие тесной связи они получили общее название, а поскольку второй из них имеет отношение к измерению длины эллипса, то столь же понятно, какое это было название, сколь бы неуместным оно ни было[1]. Итак, мы имеем понятие *эллиптических интегралов*, но пока еще не *эллиптических кривых*. Для нас они скрыты внутри подынтегральных выражений. Рассмотрим наш интеграл для вычисления длины эллипса в его освобожденной от иррациональностей форме и член под знаком квадратного корня и запишем:

$$\begin{aligned} v^2 &= (1-t^2)(1-k^2t^2) = (t-1)(t+1)(kt-1)(kt+1) \\ &= k^2(t-1)(t+1)\left(t-\frac{1}{k}\right)\left(t+\frac{1}{k}\right). \end{aligned}$$

Правая часть – многочлен четвертой степени от t, имеющий различные корни $t = \pm 1, \pm 1/k$ (потому что $k \neq \pm 1$). Если мы теперь разделим обе части на $(t-1)^4$, то получим

$$\begin{aligned} \left(\frac{v}{(t-1)^2}\right)^2 &= k^2\frac{(t+1)(t-1/k)(t+1/k)}{(t-1)(t-1)(t-1)} \\ &= k^2\left(\frac{t-1+2}{t-1}\right)\left(\frac{t-1+1-1/k}{t-1}\right)\left(\frac{t-1+1+1/k}{t-1}\right) \\ &= k^2\left(1-\frac{\alpha}{t-1}\right)\left(1-\frac{\beta}{t-1}\right)\left(1-\frac{\gamma}{t-1}\right). \end{aligned}$$

Введя новые переменные $x = 1/(t-1)$ и $y = v/(t-1)^2$, мы приведем уравнение к виду $y^2 = ax^3 + bx^2 + cx + d$ с тремя различными вещественными корнями; многочлен четвертой степени сведен к кубическому, а мы приблизились к определению эллиптической кривой, которое собираемся принять.

Существует также естественная иерархия полиномиальных плоских кривых, выражаемая в терминах степени их декартовых уравнений, т. е. максимальной степени одного члена или произведения членов, степени которых складываются. Прямая линия степени 1, $hx + iy + j = 0$, является частным случаем конического сечения степени 2, $ex^2 + fy^2 + gxy + hx + iy + j = 0$, а оно – частным случаем кубических кривых степени 3,

$$ax^3 + by^3 + cx^2y + dxy^2 + ex^2 + fy^2 + gxy + hx + iy + j = 0,$$

[1] Например, граф Фаньяно (1682–1766) открыл, что длину дуги лемнискаты можно выразить в терминах эллиптического интеграла первого рода.

и на этом мы остановимся. Исаак Ньютон во всестороннем исследовании таких кубических кривых (Newton 1667) показал, что за счет правильного выбора осей эту общую форму можно без ограничения общности свести к одному из четырех типов, а впоследствии Карл Вейерштрасс привел их к компактной форме, которая теперь называется *(длинной) формой Вейерштрасса*, $y^2 + axy + by = x^3 + cx^2 + dx + e$. На рис. 10.1 показаны все четыре существенно различных случая этой кривой: несингулярная с тремя различными вещественными корнями, несингулярная с одним вещественным корнем, сингулярная с самопересечением и сингулярная с острием. Таким образом, в наших терминах кубическая кривая – не то, что проходят в школе, где зависимой переменной является y; она должна включать член y^2. Обычно длинную форму Вейерштрасса можно еще упростить: во-первых, дополним члены в левой части до полного квадрата и выполним замену переменной $y \to y + \frac{1}{2}(ax + b)$, придя тем самым к форме $y^2 = x^3 + ax^2 + bx + c$; а во-вторых, кубический многочлен в правой части можно привести к «неполной» форме, в которой квадратный член отсутствует, выполнив подстановку $x \to x + \frac{1}{3}a$[1], после чего мы приходим к приятной глазу *короткой форме Вейерштрасса* $y^2 = x^3 + ax + b$, которую Ньютон неудачно окрестил «расходящимися параболами». Мы будем требовать, чтобы кривая была несингулярной (т. е. ее дискриминант $4a^3 + 27b^2 \neq 0$), и тогда сможем еще на один шаг приблизиться к нашему определению эллиптический кривой: это кривая, описываемая уравнением $y^2 = x^3 + ax + b$, где $4a^3 + 27b^2 \neq 0$. На рис. 10.1a и b изображены типичные формы такой кривой, а на рис. 10c и d – кривые, которые мы отвергаем из-за нарушения условия несингулярности.

Для завершения нашего определения эллиптической кривой не хватает последнего компонента. Чтобы подвести к нему и заодно продемонстрировать типичные применения кривой в элементарной теории чисел, мы позволим себе небольшое отступление.

10.2. Проблемы, проблемы, проблемы

Теория чисел – раздел математики, который в своей алгебраической и аналитической ипостаси способен задавать кажущиеся простыми, но на самом деле невероятно трудные вопросы. Действительно, две из еще не решенных проблем тысячелетия берут свое начало именно там, и в конце этой главы мы еще вернемся к ним. Два имени в долгой истории этого предмета стоят особняком, когда речь заходит о ранних вопросах: Диофант, который, по имеющимся сведениям, жил в III в. н. э., и Пьер де Ферма, живший в XVII в. читатель Диофанта, решивший многие из поставленных им задач и сам поставивший много новых. Имя Диофанта сохранилось в веках в связи с поиском, зачастую трудным, а иногда внушающим ужас, целых или рациональных решений уравнений с целыми или рациональными коэффициентами: *диофантовых уравнений*, изучение которых стало благодатной почвой для развития математических методов и признанным испытательным полигоном для их практического применения.

---

[1] Этим приемом мы обязаны Никколо Тарталье, который изложил решение кубического уравнения в стихах (буквально).

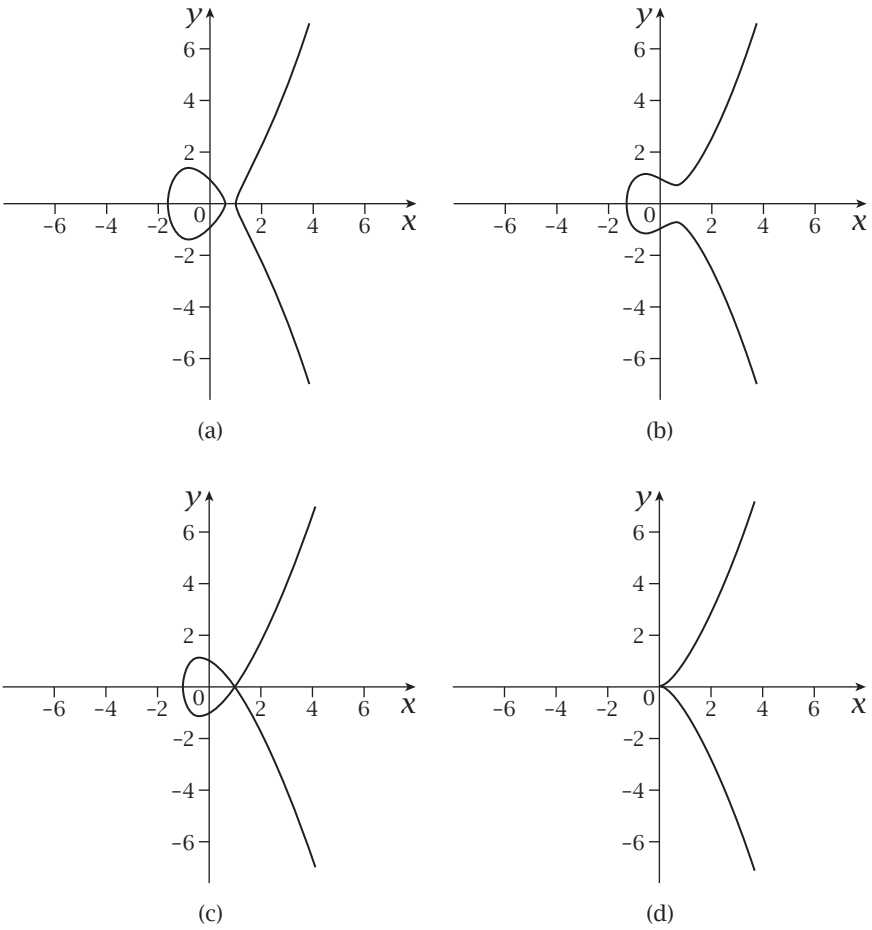


Рис. 10.1. Четыре типа кубических кривых

Наследие Диофанта составляют тринадцать книг «Арифметики», шесть из которых дошли до нас в их оригинальном греческом виде, а еще четыре – в переводе на арабский. И конечно же, сейчас нет недостатка в их переводах. Мы рассмотрим только две задачи.

***Арифметика*, книга 4, задача 24.** Данное число разложить на два числа и сделать, чтобы их произведение было кубом без стороны.

***Арифметика*, книга 6, задача 18.** Найти прямоугольный треугольник такой, чтобы его площадь, увеличенная на гипотенузу, образовала куб, а периметр был квадратом.

Подходя к этим и многим подобным задачам, Диофант был склонен брать то, что, без сомнения, считал типичными случаями, и выдвигать аргументы исходя из допущений, которые можно обобщить. В задаче 24 это выглядело так.

Пусть данное число равно 6, его первая часть равна m, а вторая, следовательно, $6 - m$; тогда их произведение равно $m(6 - m)$, и нам нужно решить уравнение

$$m(6 - m) = n^3 - n.$$

Допустим, что между m и n имеется линейная связь. Его первая догадка $n = 2m - 1$ приводит к кубическому уравнению $8m^3 - 11m^2 - 2m = 0$; отбрасывая корень $m = 0$, мы сводим его к квадратному уравнению с иррациональными корнями, которые нам не подходят. Вторая догадка – связь $n = 3m - 1$ дает куда более приемлемое уравнение $27m^3 - 26m^2 = 0$; снова отбрасывая возможность $m = 0$, мы остаемся с корнем $m = 26/27$, и две части числа 6 – это $(26/27, 136/27)$.

В задаче 18 Диофант обозначил площадь треугольника a, и, поскольку произведение двух перпендикулярных сторон равно $2a$, он положил одну из них равной 2, а другую a. Гипотенуза равна кубу минус a; условие на периметр означает, что 2 + куб = квадрат. Куб записывается в виде $(m - 1)^3$, а квадрат – в виде $(1½m + 1)^2$, а значит, первая степень m и свободный член в уравнении

$$2 + (m - 1)^3 = (1½m + 1)^2$$

взаимно уничтожаются. Это уравнение легко решается, $m = 21/4$, и мы получаем такую впечатляющую тройку для сторон треугольника:

$$(2, \frac{24121185}{628864}, \frac{24153953}{628864}).$$

Теперь перейдем к иезуиту Клоду Баше из Франции (1581–1638), получившему известность в первую очередь за свой перевод в 1621 г. первых шести книг «Арифметики» с греческого на латынь, поскольку на полях экземпляра именно этого перевода Ферма оставил знаменитое замечание, из которого родилась его прославленная Великая теорема. Интересная проблема, которую Баше оставил нам в наследство, формулируется так:

«Дано целое число c, какие рациональные решения имеет уравнение $y^2 - x^3 = c$?»

В том же году Баше дал своего рода ответ в виде оставляющей глубокое впечатление *формулы дублирования Баше*:

$$\left(\frac{a^4 - 8ac}{4b^2}, \frac{-a^6 - 20a^3c + 8c^2}{8b^3}\right),$$

которое генерирует следующее рациональное решение, если известно начальное решение (a, b). Разумеется, тут же возникает два вопроса: существует ли начальное рациональное решение для данного c, и, если да, будет ли этот процесс при повторении порождать новые решения? Второй вопрос приобретает особую остроту, если применить формулу к кривой $y^2 = x^3 + 1$ с начальной точкой (2,3) и к кривой $y^2 = x^3 - 432$ с начальной точкой (12, 36), – предлагаем читателю проверить самостоятельно.

Теперь перенесемся во времени к проблеме, которая, вероятно, исходит от французского математика Эдуарда Люка. В 1875 г. он поставил задачу: доказать, что единственный случай, когда сумма квадратов последовательных целых чисел сама является точным квадратом, имеет место для первых 24 чисел, сумма которых равна 70^2, т. е. $\sum_{r=1}^{24} r^2 = 70^2$.

Если сумму наглядно представить как сложение квадратных слоев пирамиды из пушечных ядер, как показано на рис. 10.2, а квадрат в правой части – как площадь квадрата, составленного из этих ядер, уложенных на плоском полу, то мы получаем *задачу о пушечных ядрах*.

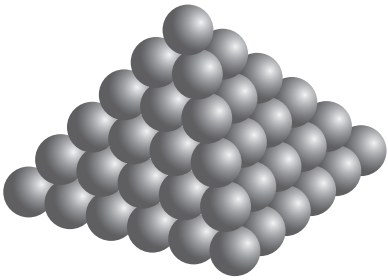


Рис. 10.2. Пирамида из пушечных ядер

Если в пирамиде n слоев, то ядер будет $1^2 + 2^2 + 3^2 + \ldots + n^2 = \frac{1}{6}n(n + 1)(2n + 1)$, а если мы хотим, чтобы из них можно было составить квадрат, то нужно найти такое целое положительное m, что $m^2 = \frac{1}{6}n(n + 1)(2n + 1)$. Конечно, имеется тривиальное решение $m = n = 1$, и нетрудно организовать поиск других решений с помощью компьютера; результат $n = 24$, $m = 70$ получается почти мгновенно. А вот найти еще одну пару – совсем другое дело, и после длительного ожидания возникает подозрение, что других целочисленных решений не существует.

Наконец, вернемся к более древним временам и рассмотрим последнюю задачу, которая может показаться незначительной, но на самом деле является глубокой и весьма трудной. Ранние ее формулировки можно найти у того же Диофанта или у арабских и индийских ученых, но мы поговорим о другом человеке, который первым познакомил с ней широкую публику, да к тому же оставил нам в наследство путаницу в терминологии: Леонардо Пизанском, известном также под именем Фибоначчи. В 1225 г., когда странствующий двор Чуда Мира, императора Священной Римской империи Фридриха II, посетил Пизу, император специально попросил встречи со знаменитым пизанским математиком Фибоначчи. Аудиенция была не простой мимолетной почестью, поскольку на ней магистр Джон из Палермо, придворный математик, предложил Фибоначчи три задачи, чтобы проверить, так ли он хорош, как о нем говорят. Одна из этих задач формулировалась так:

> «Найти квадратное число такое, что если к нему прибавить 5 или из него вычесть 5, то снова получится квадратное число».

Ее можно перефразировать: найти последовательность трех полных квадратов (целых или рациональных) такую, что разность между соседними равна 5.

Чтобы стало понятнее, приведем такую последовательность трех целых чисел, для которой разность между соседними равна 840:

$$\{529, 1369, 2209\} = \{23^2, 37^2, 47^2\}.$$

Но, памятуя о цели вопроса, следует ожидать, что задача сложнее, чем кажется на первый взгляд, и Фибоначчи нашел удивительное решение, состоящее из трех рациональных чисел (см. приложение G):

$$\left\{\frac{961}{144}, \frac{1681}{144}, \frac{2401}{144}\right\} = \left\{\left(\frac{31}{12}\right)^2, \left(\frac{41}{12}\right)^2, \left(\frac{49}{12}\right)^2\right\},$$

которое он сообщил (вместе с решениями двух других задач) императору позднее в документе, озаглавленном «Flos» (Цветы). Но задача показалась ему интересной, и он изложил более глубокие мысли в брошюре, изданной в том же году под названием «Liber Quadratorum» (Книга квадратов), где этот результат изложен в предложении 17. Именно благодаря ей возникла путаница в математической терминологии: если такая тройка существует, то разность между соседними членами он называл *конгруэнтным числом*, потому что тройка оказывается *congruum* (по латыни «сходящиеся вместе»). Как заметил Ричард Гай, «конгруэнтные числа, пожалуй, названы неудачно», но Фибоначчи жил задолго до переводов Евклида на английский язык, где фигурировали равные (конгруэнтные) треугольники, а также задолго до работы Гаусса по теории чисел с ее идеей сравнимости (конгруэнтности) по модулю простого числа, которой мы еще коснемся ниже в этой главе. Важным результатом из «Quadratorum» был тот факт, что целое конгруэнтное число имеет вид $pq(p + q)(p - q)$, когда $p + q$ четно, и число $4pq(p + q)(p - q)$, когда $p + q$ нечетно; отсюда Фибоначчи правильно предположил, что такое число делится на 24, и неправильно – что оно не может быть полным квадратом; этому факту пришлось еще четыреста лет ждать доказательства Пьером Ферма в 1659 г.

Вследствие обязательной делимости на 24 решение задачи должно включать дроби, и чрезвычайно остроумный прием, который употребил Леонардо, можно было использовать с мириадом разностей арифметической прогрессии, т. е. с мириадом конгруэнтных чисел.

Эти пять задач занимают важное место в долгой истории (алгебраической) теории чисел, какими бы случайно выбранными они ни казались. Но есть крепко связывающие их путы и оружие, более острое, чем то, что было использовано при попытке атаковать их. И это оружие – эллиптические кривые.

10.3. Общий взгляд

Если изменить буквы на стандартные обозначения переменных x и y, то задача 24 Диофанта сведется к поиску рациональных точек на кривой $y(6 - y) = x^3 - x$, а в задаче 18 то же самое требуется сделать для кривой $y^2 = (x - 1)^3 + 2$. В задаче Баше кривая уже задана, а найти требуется формулу, порождающую рациональные точки на ней. В задаче о пушечных ядрах возникает уравнение кривой $y^2 = \frac{1}{6}x(x + 1)(2x + 1)$, на которой ищутся целые точки. Остается проблема конгруэнтных чисел. Для случая разности между соседними числами 5 Фибоначчи было предложено найти рациональное число m такое, что последовательность $m^2 - 5$, m^2, $m^2 + 5$ состоит из полных квадратов. Если такое число существует, то мы можем образовать произведение $(m^2 - 5)m^2(m^2 + 5) = m^2(m^4 - 25)$, которое само должно быть полным квадратом; таким образом, мы пришли к поиску рациональных точек на кривой $y^2 = x(x^2 - 25) = x^3 - 25x$, где $x = m^2$. Но соответствие одностороннее: рациональное m, для которого такая последовательность существует, обязательно порождает рациональную точку на кривой, но у точки на кривой, координата x которой – квадрат рационального числа, координата y не обязана быть рациональной. Тем не менее мы в любом случае приходим к кубической кривой без сингулярностей, а значит, к нашему понятию эллиптической кривой, правда, в длинной, а не в короткой форме Вейерштрасса.

Переформулировав задачи в терминах эллиптических кривых, мы можем переформулировать и их решения.

Для своей задачи 24 Диофант выбрал подстановку, эквивалентную $x = 3y - 1$, и элементарный анализ показывает, что эта прямая является касательной к кривой $y(6 - y) = x^3 - x$ в точке $(-1, 0)$; кривая показана на рис. 10.3a, а ее пересечение с касательной на рис. 10.3b. Решая систему уравнений для нахождения точки пересечения, мы получаем уравнение $y^2(27y - 26) = 0$, а из него находим, что первая часть равна $y = 26/27$, а вторая $6 - y = 6 - 26/27 = 136/27$.

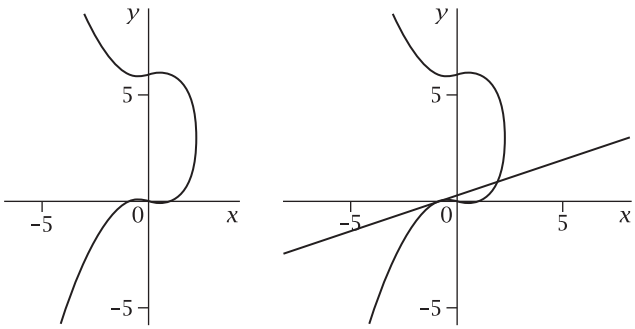


Рис. 10.3. (a) Задача 24 Диофанта. (b) Второе пересечение с касательной

Для задачи 18 прямая $y = 3/2x + 1$ является касательной к кривой $y^2 = (x - 1)^3 + 2$ в точке $(0, 1)$; кривая показана на рис. 10.4a, а ее пересечение с касательной на рис. 10.4b. И снова простейшие алгебраические преобразования показывают, что точка пересечения имеет абсциссу $x = 21/4$, откуда следует искомый ответ.

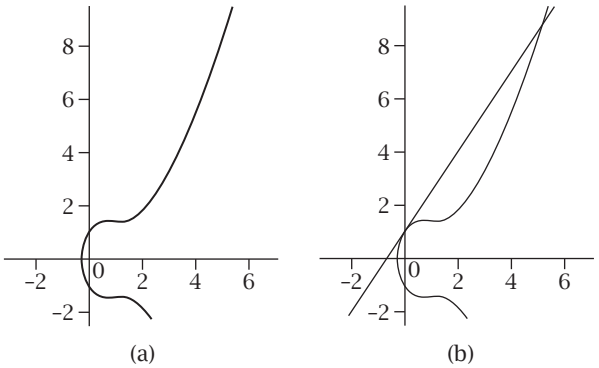


Рис. 10.4. (a) Задача 18 Диофанта. (b) Второе пересечение с касательной

Формула дублирования Баше получается в результате нахождения точки пересечения с касательной к кривой $y^2 = x^3 + c$ в точке (a, b). Детали решения таковы: продифференцировав уравнение кривой, найдем угловой коэффициент касательной $3x^2/2y$, откуда уравнение касательной в точке (a, b) имеет вид:

$$y - b = \frac{3a^2}{2b}(x - a).$$

Для нахождения второй точки пересечения касательной с кривой имеется кубическое уравнение

$$x^3 + c - \left(b + \frac{3a^2}{2b}(x - a)\right)^2 = 0,$$

которое, как мы знаем, имеет двойной корень $x = a$. Находим коэффициент $-9a^4/4b^2$ при x^2 в этом уравнении, и, согласно элементарной теории,

$$a + a + \alpha = \frac{9a^4}{4b^2}$$
$$\rightarrow \alpha = \frac{9a^4}{4b^2} - 2a = \frac{9a^4 - 8ab^2}{4b^2} = \frac{9a^4 - 8a(a^3 + c)}{4b^2} = \frac{a^4 - 8ac}{4b^2}.$$

Для нахождения y нужно просто выполнить подстановку.

В случае задачи о пушечных ядрах нет никакой подсказки о том, как использовать эллиптическую кривую $y^2 = \frac{1}{8}x(x + 1)(2x + 1)$, показанную на рис. 10.5a; касательная не показана, и не очевидно, какой она должна быть, – в трех бросающихся в глаза точках $\{(-1, 0), (-\frac{1}{2}), (0, 0)\}$ касательные вертикальны и во второй раз не пересекают кривую[1], однако же мы знаем, что точка (1, 1) лежит на кривой, и можем адаптировать стратегию анализа касательных (хотя в случае (1, 1) ее можно было бы применить), заметив, что хорда, соединяющая две рациональные точки на кривой, должна пересечь ее еще один раз в третьей рациональной точке.

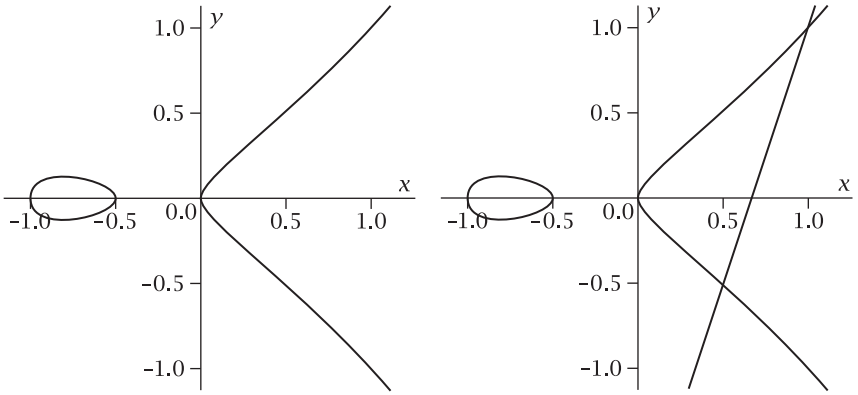


Рис. 10.5. (a) Задача о пушечных ядрах. (b) Пересечение с прямой $y = 3x - 2$

Прямая, проходящая через точки (0, 0) и (1, 1), имеет уравнение $y = x$, а поиск точек ее пересечения с кубической кривой приводит к уравнению $x^2 = \frac{1}{8}x(x + 1)$

[1] На самом деле пересекают, но об этом мы узнаем, когда завершим определение эллиптической кривой.

$(2x + 1)$ или $x^3 - \frac{3}{2}x^2 + \frac{1}{2}x = 0$, третий корень которого $x = \frac{1}{2}$, поэтому третья точка на кривой – $(\frac{1}{2}, \frac{1}{2})$, но она не целая. Точно повторять этот процесс не имеет смысла, поскольку все три точки, конечно, лежат на прямой $y = x$, но, в силу симметрии графика, точка $(\frac{1}{2}, -\frac{1}{2})$ также должна лежать на кривой. Соединяя ее с точкой (1, 1), получаем прямую $y = 3x - 2$, показанную на рис. 10.5b, и уравнение $(3x - 2)^2 = \frac{1}{6}x(x + 1)(2x + 1)$, или $2x^3 - 51x^2 + 73x - 24 = 0$. Мы знаем два его корня, а элементарная теория говорит, что сумма всех трех равна $\frac{51}{2}$, так что третий корень $x = 24$, что дает искомую точку (24, 70), расположенную далеко за границами рисунка. Мы нашли решение, из которого следует, что $1^2 + 2^2 + 3^2 + \ldots + 24^2 = 70^2$. Но является ли оно единственным? Можно ли с пользой продолжить этот процесс? Является ли альтернатива касательным плодотворной? Скоро увидим.

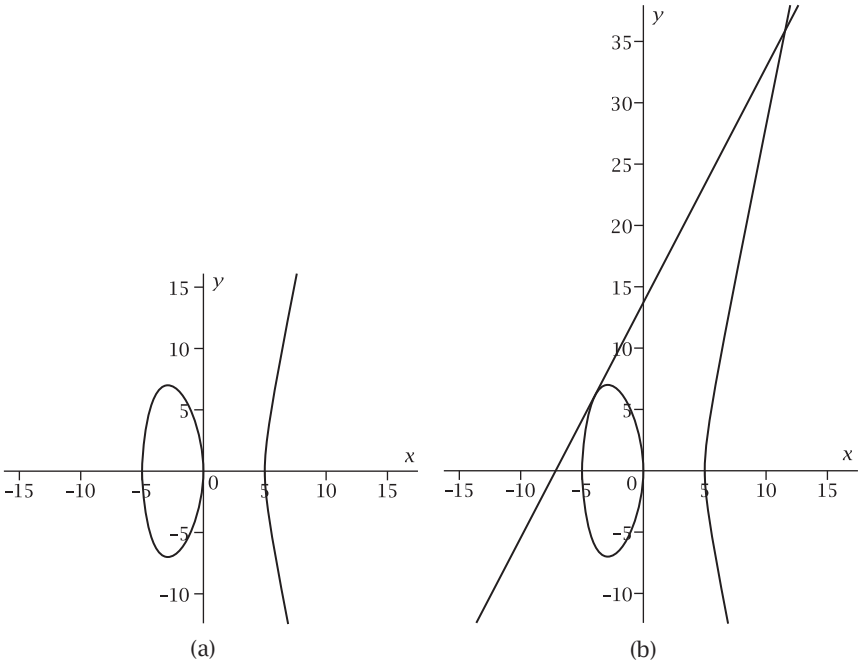


Рис. 10.6. (a) Кривая Фибоначчи и (b) ее пересечение с прямой $y = \frac{23}{12}x + \frac{41}{3}$

Наконец, из задачи Фибоначчи мы извлекли эллиптическую кривую $y^2 = x(x^2 - 25)$, показанную на рис. 10.6a.

Единственные очевидные целые точки на кривой – (0, 0), (5, 0), (–5, 0), но в качестве касательных все они бесполезны. Но мы предприняли поиск наудачу и быстро обнаружили, что $x = 4$ дает отрицательный полный квадрат, поэтому $x = -4$ дает точку (–4, 6) на кривой. Если мы поставим целью найти второе пересечение касательной в этой точке с кривой, то элементарный анализ и алгебра покажут, что касательная имеет уравнение $y = \frac{23}{12}x + \frac{41}{3}$ и пересекается с кривой в точке

$$\left(\frac{1681}{144}, \frac{62279}{1728}\right),$$

показанной на рис. 10.6b, и, таким образом, получается, что $^{1681}/_{144} \pm 5$ – точные квадраты, а поскольку

$$\frac{1681}{144} = \left(\frac{41}{12}\right)^2,$$

то мы имеем как раз решение, найденное Фибоначчи.

Вот на этом последнем вопросе мы и задержимся, потому что он заслуживает внимательного рассмотрения и дает возможность испытать мощь эллиптических кривых в элементарных, но важных задачах.

10.4. Проблема конгруэнтных чисел

Именно последняя задача Фибоначчи привела нас от головоломок к действительно трудным вопросам, и мы сделали наблюдение, которое было сделано задолго до Чуда Мира и очень далеко от его двора. Мы говорим, что если для рационального m и целого положительного n существует последовательность полных квадратов $m^2 - n$, m^2, $m^2 + n$, то n называется конгруэнтным числом. Арабские и индийские математики предложили другую формулировку: существование рациональных чисел таких, что $a^2 + b^2 = c^2$ и $\frac{1}{2}ab = n$, где n может быть рациональным. Геометрически мы ищем рациональные площади n прямоугольных треугольников с рациональными сторонами; такие треугольники не без оснований называются *рациональными*. Алгебраические преобразования

$$\begin{aligned} a^2 \pm 2ab + b^2 &= c^2 \pm 4n \\ \Leftrightarrow (a \pm b)^2 &= c^2 \pm 4n \\ \Leftrightarrow (\tfrac{1}{2}(a \pm b))^2 &= (\tfrac{1}{2}c)^2 \pm n = m^2 \pm n \end{aligned}$$

показывают, что если n – рациональное число, то существует взаимно однозначное соответствие между двумя множествами

$$\{(a, b, c) \in \mathbb{Q}^3 : a^2 + b^2 = c^2, \tfrac{1}{2}ab = n\}$$

и

$$\{(m^2 - n, m^2, m^2 + n) : (m, n) \in \mathbb{Q}^2\},$$

описываемое формулой

$$(a, b, c) \to ((\tfrac{1}{2}c)^2 - n, (\tfrac{1}{2}c)^2, (\tfrac{1}{2}c)^2 + n),$$

и наоборот:

$$(m^2 - n, m^2, m^2 + n) \to (\sqrt{m^2 + n} - \sqrt{m^2 - n}, \sqrt{m^2 + n} + \sqrt{m^2 - n}, 2m).$$

Частными случаями этой эквивалентности между последовательностями рациональных квадратов S и рациональных треугольников Δ являются:

$$n = 5: \quad S = \left\{\left(\frac{31}{12}\right)^2, \left(\frac{41}{12}\right)^2, \left(\frac{49}{12}\right)^2\right\} \quad \text{и} \quad \Delta = \left(\frac{3}{2}, \frac{20}{3}, \frac{41}{6}\right),$$

$$n = 6: \quad S = \{(\tfrac{1}{2})^2, (\tfrac{5}{2})^2, (\tfrac{7}{2})^2\} \quad \text{и} \quad \Delta = \{3, 4, 5\},$$

$$n = 7: \quad S = \left\{\left(\frac{113}{120}\right)^2, \left(\frac{337}{120}\right)^2, \left(\frac{463}{120}\right)^2\right\} \quad \text{и} \quad \Delta = \left\{\frac{35}{12}, \frac{24}{5}, \frac{337}{60}\right\}.$$

Поиск конгруэнтных чисел был трансформирован в вопрос: для каких рациональных чисел n существует рациональный треугольник площади n? Он получил название *задачи о конгруэнтных числах*, а ускользающая от понимания природа таких чисел сделала эту задачу одной из самых трудных и важных проблем современной теории чисел. На самом деле тест Фибоначчи задолго до него прошли древние арабские математики, оставившие еще в X в. рукописи, в которых утверждается, что числа 5, 6, 14, 15, 21, 30, 34, 65, 70, 110, 154, 190 – и 10 374 – конгруэнтные.

Важное наблюдение поможет нам устранить избыточность в определении конгруэнтного числа. Арабы допускали, что площадь рационального треугольника может быть рациональной, а не целой, но благодаря масштабированию мы видим, что оба случая эквивалентны. Если имеется рациональный треугольник площади n, то имеется такой треугольник площади k^2n для любого рационального k, и наоборот (нужно умножить треугольник на k или $1/k$). Это означает, что если задана рациональная площадь, то мы можем выделить из нее квадратную часть и таким образом получить более примитивную рациональную площадь, а затем обратить этот процесс, умножив на квадрат знаменателя, так что в итоге получится свободное от квадратов целое число, порождающее все семейство. Что-то уж слишком много слов для такого простого процесса; например, если мы смогли построить рациональный треугольник площади $^{18}/_{175}$, то, поскольку $^{18}/_{175} = (^3/_5)^2 \times {}^2/_7$, площадь $^2/_7$ тоже годится, после чего умножаем ее на 7^2 и приходим к 14 – примитивному свободному от квадратов конгруэнтному числу. Таким образом, в погоне за рациональными площадями рациональных треугольников мы можем ограничиться поиском среди (положительных) свободных от квадратов целых чисел и приходим к окончательному определению: *конгруэнтным числом* называется положительное, свободное от квадратов целое число, являющееся площадью какого-нибудь рационального треугольника. При таком определении нам нужно только сгенерировать все рациональные треугольники, а чтобы подойти к этой задаче систематически, мы можем воспользоваться методом Евклида генерирования примитивных[1] пифагоровых троек[2]: взять два положительных целых числа a, b таких, что:

- $a > b$,
- одно из них четно, другое нечетно;
- числа являются взаимно простыми,

тогда $(a^2 - b^2, 2ab, a^2 + b^2)$ – примитивная пифагорова тройка.

---

[1] Не имеющих общих множителей.

[2] Книга X, лемма 1, предваряющая предложение 29.

Например, возьмем $(a, b) = (5, 4)$, чтобы сгенерировать примитивный рациональный треугольник (9, 40, 41) площади $180 = 6^2 \times 5$, это означает, что 5 – конгруэнтное число благодаря треугольнику со сторонами (⅝, ⁴⁰⁄₉, ⁴¹⁄₆), который мы уже видели раньше. Итак, применим наш систематический метод: зафиксируем a, наложим на каждое b ограничения, воспользуемся приведенным выше выражением для генерирования троек и при необходимости освободимся от квадратов.

Первая из следующих далее строк была сгенерирована путем увеличения a и для каждого из них вычисления площади треугольника для допустимых значений b:

{6, 30, 60, 84, 210, 180, 210, 330, 630, 924, 546, 504, 1320, 1560, 840,
1386, 2340, 1224, 990, 2730, 3570, 1710, 2574, 4620, 5610, 5016, 2310,
1716, 7140, 7980, 3036, 4290, 7956, 10374, 10920, 8970, 3900, 2730,
7854, 11970, 14490, 11550, 4914, 6630, 12540, ...}

Теперь освободимся от квадратов, сохранив порядок:

{6, 30, 15, 21, 210, 5, 210, 330, 70, 231, 546, 14, 330,3 90,210, 154,
65, 34, 110, 2730, 3570, 190, 286, 1155, 5610, 1254, 2310, 429, 1785,
1995, 759, 4290, 221, 10374, 2730, 8970, 39, 2730, 7854, 1330, 1610,
462, 546, 6630, 3135, ...}

Затем переупорядочим числа в порядке возрастания и удалим дубликаты:

{5, 6, 14, 15, 21, 30, 34, 39, 65, 70, 110, 154, 190, 210, 221, 231, 286,
330, 390, 429, 462, 546, 759, 1155, 1254, 1330, 1610, 1785, 1995, 2310,
2730, 3135, 3570, 4290, 5610, 6630, 7854, 8970, 10374, ...}

Это все конгруэнтные числа, но процесс не совершенный. Мало того что исходный список не упорядочен, так еще в нем есть повторения и подозрительные лакуны. Прежде всего заметим, что числа 1, 2 и 3 отсутствуют в окончательном списке, но это правильно, потому что они не конгруэнтные; число 7 отсутствует, хотя оно конгруэнтное; 11 отсутствует и не конгруэнтное; 13 отсутствует, хотя конгруэнтное; 17 отсутствует и не конгруэнтное... Эта процедура и все ей подобные не годится для генерирования конгруэнтных чисел сколько-нибудь полезным способом: 13 появляется, когда $a = 325$ и $b = 36$, что дает площадь $1220649300 = 13 \times 9690^2$, а рациональный треугольник площади 13 – вот в таком впечатляющем виде:

$$\left(\frac{323}{30}, \frac{780}{323}, \frac{106921}{9690}\right).$$

23 появляется, когда $a = 24336$ и $b = 17689$, что дает площадь $2861397263088 = 23 \times 352716^2$, а рациональный треугольник площади 23 поражает воображение еще сильнее:

$$\left(\frac{391}{20748}, \frac{41496}{17}, \frac{42025}{352716}\right).$$

Ждать появления конгруэнтного числа 157 было бы бессмысленно, потому что рациональный треугольник такой площади имеет катеты

$$\left(\frac{6803298487826435051217540}{411340519227716149383203}, \frac{411340519227716149383203}{21666555693714761309610}\right),$$

а его гипотенуза равна

$$\frac{224403517704336969924557513090667486316094847204 1}{8912332268928859588025533178967163570016480830}.$$

Участвующие в поиске переменные равны

$$a = 443624018997429899709925$$

и

$$b = 166136231668185267540804.$$

Доктор Кармен Бруни из Университета Ватерлоо выполнил вычисление, показавшее, что при скорости поиска 100 000 чисел в секунду потребовалось бы 5.89×10^{34} лет для нахождения этого решения. Немецкий математик Дон Цагир сделал это[1], но, естественно, не методом полного перебора. Быть может, вас удивит, что он применил эллиптические кривые; точнее, он нашел рациональную точку на эллиптической кривой $y^2 = x^3 - 157^2x$.

Мы уже перешли от последовательности квадратов к специальному типу эллиптической кривой удобным, но несколько корявым способом. К счастью, существует более убедительная ассоциация, похожая на ту, с которой мы уже познакомились: пусть n – положительное число, тогда существует взаимно однозначное соответствие между двумя множествами:

$$\{(a, b, c) \in \mathbb{Q}^3 : a^2 + b^2 = c^2, \tfrac{1}{2}ab = n\}$$

и

$$\{(x, y. \in \mathbb{Q}^2 : y^2 = x^3 - n^2x, y \neq 0\},$$

описываемое формулой (существуют и другие):

$$(a, b, c) \to \left(\frac{nb}{c-a}, \frac{2n^2}{c-a}\right) \quad \text{и} \quad (x, y) \to \left(\frac{x^2 - n^2}{y}, \frac{2nx}{y}, \frac{x^2 + n^2}{y}\right).$$

Доказательство прямолинейное, достаточно простой подстановки. Соединив вместе все три интерпретации, – последовательность квадратов S, рациональный треугольник Δ и эллиптическую кривую E, – мы придем к рис. 10.7, на котором показаны точные переходы между интерпретациями.

[1] *http://people.mpim-bonn.mpg.de/zagier/files/mpim/89-23/fulltext.pdf* (in German).

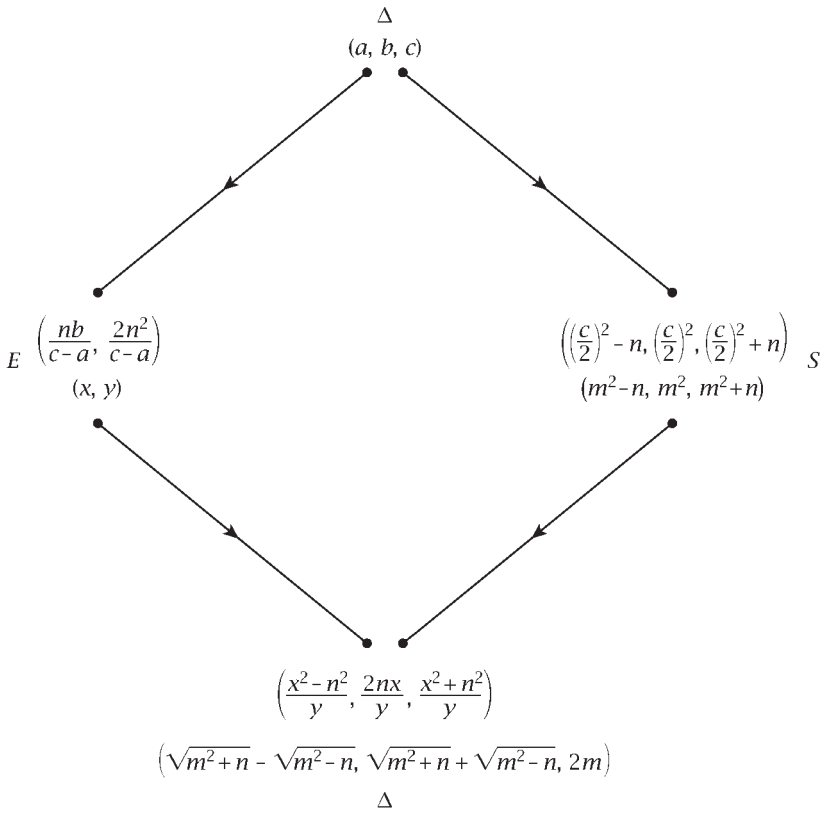


Рис. 10.7. Эквивалентность Δ, E и S

Чтобы понять, насколько продуктивной может быть эта ассоциация, начнем с канонического рационального треугольника (3, 4, 5), который устанавливает, что число $n = 6$ конгруэнтное. Ассоциация связывает его с эллиптической кривой $y^2 = x^3 - 36x$ и точкой (12, 36) на ней, и мы можем повторно применить процесс построения касательной из предыдущего раздела вместе с этими ассоциациями для порождения последовательностей квадратов и новых рациональных треугольников. Легко видеть, что уравнение касательной к этой эллиптической кривой в этой точке имеет вид $y = 1½x - 30$, а также что она во второй раз пересекает кривую в точке (25/4, 35/8), которая порождает еще один рациональный треугольник:

$$\left(\frac{7}{10}, \frac{120}{7}, \frac{1201}{70}\right)$$

площади 6, соответствующий трехчленной арифметической прогрессии квадратов:

$$\left\{\left(\frac{1151}{140}\right)^2, \left(\frac{1201}{140}\right)^2, \left(\frac{1249}{140}\right)^2\right\}$$

с разностью 6. И мы имеем (как выясняется) бесконечно повторяющийся процесс. Уравнение касательной к эллиптической кривой в точке (25/4, 35/8) имеет вид:

$$y = \frac{1299}{140}x - \frac{6005}{112},$$

и она снова пересекается с кривой в точке

$$\left(\frac{1442401}{19600}, \frac{1726556399}{2744000}\right),$$

которая соответствует рациональному треугольнику:

$$\left(\frac{1437599}{168140}, \frac{2017680}{1437599}, \frac{2094350404801}{241717895860}\right)$$

площади 6 и еще одной трехчленной арифметической прогрессии квадратов с разностью 6:

$$\left\{\left(\frac{1727438169601}{483435791720}\right)^2, \left(\frac{2094350404801}{483435791720}\right)^2, \left(\frac{77611083871}{483435791720}\right)^2\right\}$$

и т. д.

Наконец, как мы уже упоминали, Фибоначчи высказал предположение, что 1 не является конгруэнтным числом, но доказал этот факт (а также неконгруэнтность 2 и 3) только Ферма, воспользовавшись интерпретацией на основе треугольника и придуманным им методом «бесконечного спуска». Из неконгруэнтности 1 вытекают неожиданные следствия, три из которых приведены ниже.

- Никакое квадратное число не может быть конгруэнтным: если бы оно было таковым, то мы могли прийти к 1 путем деления.
- Произведение трех последовательных целых чисел не может быть полным квадратом: иначе существовали бы такие целые числа, что

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

- Число $\sqrt{2}$ иррационально: если бы $\sqrt{2}$ было рационально, то существовал бы рациональный треугольник $(\sqrt{2}, \sqrt{2}, 2)$ площади 1.

Разобравшись с геометрией эллиптической кривой и некоторыми ее теоретико-числовыми применениями, перейдем к арифметике на ней – и наконец-то дадим ее полное определение.

10.5. Арифметика

Мы уже говорили, что Ричард Гай посетовал на неудачное название эллиптических кривых. А одно математическое соглашение может еще и добавить путаницы в глазах неспециалиста, поскольку практики нередко берут обычные слова и переосмысливают их, наполняя техническим содержанием: *производная*, *интеграл*[1], *трансцендентный*… и *группа*. Анри Пуанкаре обобщил это

[1] *Integral* в английском языке означает «нечто целое». – *Прим. перев.*

явление, заметив, что «математика – это искусство давать одно и то же название разным вещам». Последний термин *группа* зарезервирован для множества элементов вместе со способом их комбинирования (бинарной операцией), которое содержит все необходимое для элементарной арифметики, алгебры и, в частности, решения уравнений. Точнее, требуется, чтобы множество E и бинарная операция (которая, чтобы не изобретать новый символ и в полном соответствии с философией повторного использования, обозначается знаком +) удовлетворяли следующим условиям:

- для любых $A, B \in E$ также $A + B \in E$, т. е. множество «замкнуто» относительно операции;
- существует $O \in E$ такой, что для любого $A \in E$ $A + O = O + A = A$; этот элемент называется «нейтральным»;
- для любого $A \in E$ существует такой элемент $-A \in E$, что $A + (-A) = (-A) + A = O$, т. е. для каждого элемента имеется обратный;
- для любых $A, B, C \in E$ $A + (B + C) = (A + B) + C$: операция «ассоциативна»;
- для любых $A, B \in E$ $A + B = B + A$: операция «коммутативна».

Последнее условие введено для удобства, а не по необходимости; группы, обладающие этим свойством, называются абелевыми[1].

Это понятие, кажущееся никак не связанным с нашим обсуждением эллиптических кривых, на самом деле играет ключевую роль, потому что объект, который до сих пор представлялся чисто геометрическим и определялся геометрическими построениями, параллельно существует в мире алгебры, т. е. его можно интерпретировать как группу и выполнять над ним основные арифметические операции. При условии, конечно, что мы умеем интерпретировать сумму двух точек на кривой, а в связи с этим описанное ранее построение касательной и секущей обретает куда более важную причину, потому что любая невертикальная прямая, которая пересекает кривую дважды, обязательно пересечет ее и в третий раз. При наличии двух точек A и B, находящихся в общем положении на рис. 10.8, третья точка пересечения, обозначенная A ∗ B, является очевидным кандидатом на роль их суммы, однако имеются препятствия. Аксиомы (абелевой) группы можно отнести к двум типам: две, в которых упоминается нейтральный и обратный элемент, и три, касающиеся замкнутости, ассоциативности и коммутативности. А наше зачаточное определение сложения противоречит обоим: не может быть нейтрального элемента, потому что условие A ∗ O = A означало бы, что O должно быть следующей точкой пересечения касательной к кривой в точке A и, стало быть, зависит от A, и к тому же мы имели бы O ∗ B = (A ∗ A) ∗ B = A ∗ (A ∗ B) = B, что, очевидно, является бредом, поэтому аксиома ассоциативности не выполняется.

Хотя на первый взгляд это кажется чудом[2], мы можем слегка видоизменить эту непригодную конструкцию, так чтобы выполнялись все аксиомы группы, для этого нужно выбрать на кривой произвольную, но фиксированную точку.

---

[1] В честь Нильса Абеля.

[2] Впрочем, при рассмотрении сквозь призму общей теории абелевых групп волшебство исчезает.

Обозначим O (говорящее обозначение, не правда ли?) эту фиксированную точку, которая показана на рис. 10.8, и определим сумму двух произвольных точек на кривой как A + B = O ∗ (A ∗ B), т. е. как следующее пересечение с кривой прямой линии, соединяющей O с A ∗ B (снова см. рис. 10.8). Легко видеть, что фиксированная точка O является нейтральным элементом, так как из коллинеарности O, A, O ∗ A следует, что O + A = O ∗ (O ∗ A) = A. Более того, точка A' = (O ∗ O) ∗ A является обратной к A, потому что A' + A = O ∗ (A' ∗ A) = O ∗ (((O ∗ O) ∗ A) ∗ A) = O ∗ (O ∗ O) = O, опять же в силу коллинеарности точек; тогда мы можем написать –A = (O ∗ O) ∗ A. Ассоциативность тоже имеет место, но доказательство этого факта утомительно, поэтому просим читателя поверить нам на слово.

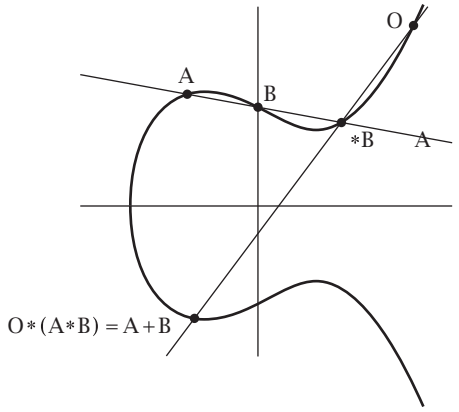


Рис. 10.8. Групповой закон на эллиптической кривой

Все сделано – если не считать важного вопроса о вертикальных прямых.

Если мы зафиксируем A и B (а значит, и A ∗ B) и будем двигать O дальше вдоль кривой, то переменная прямая, определяющая A + B, будет поворачиваться вокруг A ∗ B и приближаться к вертикали, никогда не достигая ее. Мы можем сделать прямую вертикальной, пополнив кубическую кривую «бесконечно удаленной точкой» и взяв ее в качестве нейтральной точки O; тогда вертикальные прямые будут пересекать кривую в трех точках, две из которых – A ∗ B и A + B – конечны, а обратной к точке будет ее отражение относительно горизонтальной оси.

И вот теперь мы, наконец, готовы дать полное определение эллиптической кривой (над множеством вещественных чисел). Эллиптической кривой называется множество:

$$\{(x, y) \in \mathbb{R}^2 : y^2 = x^3 + ax + b;\ 4a^3 + 27b^2 \neq 0\} \cup \{O\}.$$

Включение бесконечно удаленной точки можно оформить абсолютно строго (дорогой) ценой перехода в абстрактное пространство проективной плоскости, что, к счастью, выходит за рамки наших обязанностей и, к еще большему счастью, совершенно не нужно для наших целей.

Прежде чем продолжить, рассмотрим (несколько утомительные) детали вычисления суммы двух точек на кривой по их координатам. После того как

все элементарные алгебраические преобразования, связанные с различными типами пересечения, останутся позади, мы будем иметь в сухом остатке следующее.

Сумма двух точек $P_1 = (x_1, y_1)$ и $P_2 = (x_2, y_2)$ на эллиптической кривой $y^2 = x^3 + ax + b$ вычисляется по формулам:

- если $x_1 \neq x_2$, то $P_1 + P_2 = (x_3, y_3)$, где

$$\begin{cases} x_3 = \left(\dfrac{y_2 - y_1}{x_2 - x_1}\right)^2 - x_1 - x_2 \\ y_3 = -y_1 + \left(\dfrac{y_2 - y_1}{x_2 - x_1}\right)(x_1 - x_3); \end{cases}$$

- если $x_1 = x_2$, то:
  - если $y_1 = -y_2$, то $P_1 + P_2 = O$;
  - если $y_1 \neq -y_2$, то $P_1 + P_2 = 2P_1 = (x_3, y_3)$, где

$$\begin{cases} x_3 = \left(\dfrac{3x_1^2 + a}{2y_1}\right)^2 - 2x_1 \\ y_3 = -y_1 + \left(\dfrac{3x_1^2 + a}{2y_1}\right)(x_1 - x_3). \end{cases}$$

В первом случае речь идет о пересечении общего вида, во втором – о случае, когда пересекающая прямая является вертикальной или касательной к кривой.

Заметим, что арифметические групповые аксиомы гарантируют отсутствие неоднозначности. Например, $4P = P + P + P + P = 2P + 2P = P + 3P$ кажется очевидным, если говорить о манипуляциях символами, но за этими равенствами скрываются тонкости различных пересечений прямых с кривой, а эквивалентность продемонстрирована на рис. 10.9.

Вычислительные формулы позволяют нам взять любую эллиптическую кривую и любую начальную точку (или точки) на ней и произвести вычисления, как нам будет угодно. Например, для кривой $y^2 = x^3 - 5x + 8$ и точки $P(1, 2)$ мы можем рассмотреть последовательно генерируемые кратные. Начинается последовательность так:

$$2P = \left(-\frac{7}{4}, -\frac{27}{8}\right), \quad 3P = P + 2P = \left(\frac{553}{121}, -\frac{11950}{1331}\right), \quad \ldots.$$

Дроби, равные координатам кратных, становятся все более громоздкими. В табл. 10.1 показаны (в эстетически привлекательной форме) числители и знаменатели координат x точек {P, 2P, 3P, ... , 15P}, а в табл. 10.2 – то же самое для координат y (в обоих случаях знак минус приписывается числителю). Этот процесс генерирует все более сложные рациональные точки на кривой.

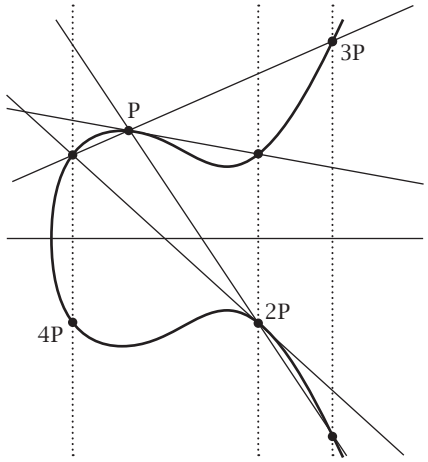


Рис. 10.9. Корректно определенные арифметические операции

Табл. 10.1. Координаты *x* кратных точки P(1, 2)

1
−7
553
45313
−19035719
2023813321
49242784704696l
2870103992322043393
−605072008062695815697327
4096232639504661713093461475З
901093218102649252044904107963254О1
−3062917124665740211193504994324000336883l9
9652333757946037439772059665978001789849562735636l
1470029049847141125711633299963290269471476525239512788З513
1863568237805708852575760933134572809631068778572203746362412153313
29527912871071204386688507595732530402719256283751766399612726094Ч1
11172733331684514306947905124320043169760918542492309216Ч
72076267683725311280022064313064365223932818287961
13373165024713309118563593855026673201000О
26863288840579508748932404959298О1
74777485548545062739932436Ч
41737822887561958289496l
34950386552770535О4
935706047761
17279102500
9357481
11664
121
4
1

Табл. 10.2. Координаты *y* кратных точки P(1, 2)

2
−27
−11950
8655103
8939659342
−439872200468869
109272278371852809561 8
13780060468451643045994142977
−941902861661639751136830734845315 0
−2462189989929456867409298574207561320461218 7
74940853885268286628149635581876788277525377623628839 8
42175374816324337346580531900355783380990006835862315528436607 9
−1177955498453725034735299819939981993978073114196045887485832173947247868259012882
17820788322320927710885517052354527183638385093878018495039756855584226045506936558525 3
11453927878769482382796952815378743985566207386732094754981069640997042247391427293641613194165085250
507398438140656172703799203729812959704355903108836959235782204215855125450268962508129831072178651 1
11809070773448643935992365672087813262745201301242953243725834478084845628358491992 88
61191124510095112244683016103855390061547624099374042290161215377202343459
15465049191152713044713196277349163469241815256541349801000000
13231907350484258765667140161519920484086746444909701
204483644802829789981943847558452163608437 52
269646438028348861920971568936899209
6533982622887348588544276992
90512623841412275 9
2271338023625000
2862453437 9
1259712
1331
8
1

Итак, мы дополнили нашу геометрическую реализацию эллиптической кривой алгебраической структурой. Теперь рассмотрим миры, в которых могут существовать такие кривые, и попутно упомянем еще одну неудачно названную математическую конструкцию.

10.6. Плодородные поля

Мы рисовали эллиптические кривые, неявное предполагая, что кривая состоит из бесконечного непрерывного множества точек, координатами которых являются пары вещественных чисел. Но, приглядевшись к вычислениям координат суммы двух точек и кратного точки на эллиптической кривой с рациональными коэффициентами, мы увидим, что если начать с точки с рациональными координатами, то в процессе будут порождаться только рациональные координаты и, следовательно, еще одна рациональная точка на кривой. Не всякая точка на кривой может быть достигнута выполнением такого процесса, а множество рациональных точек можно рассматривать как подгруппу всего множества точек на кривой. Это наблюдение важно, но не менее важно обобщение, частью которого оно является, и тут нам понадобится абстракция математического *поля*; еще одно общеупотребительное слово, обозначающее отнюдь не общеупотребительное понятие. Примером группы является множество целых чисел $\mathbb{Z}$ с операцией сложения, а примером поля – множество рациональных чисел $\mathbb{Q}$ с операциями сложения и умножения, которые взаимодействуют между собой, как и положено, с помощью закона дистрибутивности умножения относительно сложения. Имея эти базовые сведения, сформулируем результат, установленный Анри Пуанкаре примерно в 1900 г.

Пусть K – поле, а эллиптическая кривая E описывается уравнением вида $y^2 = x^3 + ax + b$, где $a, b \in K$. Обозначим $E(K)$ множество точек E с координатами, принадлежащими K. Тогда $E(K)$ – подгруппа всего множества точек E.

Например, если взять эллиптическую кривую с $a, b \in \mathbb{Q}$ и на ней точку с рациональной координатой x, то ясно, что соответствующее значение y^2 будет рациональным, но отнюдь не очевидно, что то же самое можно сказать про y. Результат Пуанкаре гласит, что все точки на кривой, обе координаты которых рациональны, образуют группу с правилом сложения, определяемым хордой и касательной, как мы интуитивно предполагали выше. Но это далеко не все. Пуанкаре выдвинул гипотезу, а Луис Морделл в 1922 г. доказал следующий результат.

Пусть E – эллиптическая кривая, описываемая уравнением $y^2 = x^3 + ax + b$, и $a, b \in \mathbb{Q}$. Тогда группа рациональных точек $E(\mathbb{Q})$ является конечнопорожденной.

Иными словами, существует конечное множество точек $P_1, P_2, P_3, \ldots, P_t \in E(\mathbb{Q})$, такое, что любую точку $P \in E(\mathbb{Q})$ можно записать в виде:

$$P = n_1P_1 + n_2P_2 + n_3P_3 + \ldots + n_tP_t \text{ для некоторых } n_1, n, n_3, \ldots, n_t \in Z.$$

Отсюда следует так много всего, что даже подумать о включении этого материала сюда немыслимо. Результатов и гипотез полно, как и очень трудной математики, и мы ограничимся всего одним результатом, доказанным Барри Мазуром в 1977 г., который можно интерпретировать следующим образом.

Возьмем рациональную точку P на рациональной эллиптической кривой и будем вычислять ее последовательные кратные. Если мы не достигли O, вычислив все кратные до 16P включительно, то не достигнем ее никогда.

В нашем примере (табл. 10.1 и 10.2) вычислены первые 15 кратных, и читатель может проверить, будет ли следующее кратное равно 0! Этот результат позволяет, в частности, надеяться, что будет сгенерировано бесконечное число рациональных точек на рациональной кривой, если начальная точка выбрана удачно. Чего

не наблюдается, если мы вернемся из поля рациональных чисел $\mathbb{Q}$ обратно в множество Z, полем не являющееся[1]. Целые точки на эллиптической кривой с целыми коэффициентами – звери гораздо более редкие, что и не удивительно, потому что при имеющихся правилах арифметических действий весьма маловероятно, что сумма целых точек или кратное целой точки будут иметь целые координаты.

Точно это стало ясно в 1928 г., когда немецкий математик Карл Зигель доказал следующий результат.

Пусть E – эллиптическая кривая, описываемая уравнением $y^2 = x^3 + ax + b$, где $a, b \in \mathrm{Z}$. Тогда E содержит лишь конечное множество точек $\mathrm{P} = (x, y)$ таких, что x, $y \in \mathrm{Z}$. Иными словами, $E(\mathrm{Z})$ конечно.

Таким образом, задача о пушечных ядрах имеет лишь конечное число решений, а в работе (Watson 1918) (с помощью эллиптических функций)[2] впервые было доказано, что два упомянутых выше решения единственные. Загадка разрешена, но связь с *решеткой Лича*, реализующей плотнейшую упаковку сфер в 24-мерном пространстве, определяет ее важный вклад в *24-битовый код Голея, исправляющий ошибки*, а через него в целостность электронной передачи данных. Но все это мы отложим в сторону, а поговорим о числовом поле совершенно другого вида, над которым можно определить эллиптическую кривую, и уже с этой стороны перейдем к вкладу в криптографию: секретную передачу сообщений.

Нас интересует конечное поле $F_p = \{0, 1, 2, 3, \ldots, p - 1\}$, где p – простое число[3], а арифметические действия производятся по модулю p. Этот математический термин описывает не что иное, как «арифметику часов», которую мы проходим в начальной школе и которая неявно присутствует, когда мы пользуемся 12- или 24-часовыми часами: просто нужно результат любого арифметического действия свести к остатку от его деления на соответствующий модуль. В табл. 10.3 показана таблица сложения по модулю 7, а в табл. 10.4 – таблица умножения по тому же модулю.

Таблица 10.3. Сложение по модулю 7

+	0	1	2	3	4	5	6
0	0	1	2	3	4	5	6
1	1	2	3	4	5	6	0
2	2	3	4	5	6	0	1
3	3	4	5	6	0	1	2
4	4	5	6	0	1	2	3
5	5	6	0	1	2	3	4
6	6	0	1	2	3	4	5

[1] На самом деле это *кольцо*.

[2] Впоследствии этот результат неоднократно был установлен с помощью «элементарных» средств.

[3] Оно обязательно должно быть простым, чтобы умножение обладало требуемыми свойствами.

Таблица 10.4. Умножение по модулю 7

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

Мы будем рассматривать эллиптическую кривую, определенную над конечным полем F_p, $y^2 = x^3 + ax + b$, где $a, b, x \in F_p$, и правая часть приведена по модулю p, а единственно допустимыми значениями y являются такие, для которых $y^2 \in F_p \Rightarrow y \in F_p$.

И именно в этом контексте мы были вынуждены проявить некоторую осторожность, когда преобразовывали уравнение эллиптической кривой из длинной формы Вейерштрасса в короткую (см. стр. 179 и 180). Двухэтапный процесс дополнения членов, содержащих y, до полного квадрата и избавления от члена, содержащего x^2, требует деления на 2 и 3 соответственно, но эти операции не определены соответственно в полях $F_2 = \{0, 1\}$ и $F_3 = \{0, 1, 2\}$. Если поле, над которым определена кривая, совпадает с одним из них, то нам придется удовлетвориться длинной формой, но это все же техническая деталь, а не какое-то внутреннее свойство кривой. Приведенное выше определение арифметических операций остается корректным для кривых над любым полем, а единственное отличие заключается в том, что все вычисления производятся по модулю некоторого простого числа p, а деление реализуется как умножение на обратный элемент.

Чтобы развить эту идею, мы возьмем одну эллиптическую кривую $y^2 = x^3 + x + 1$, но определим ее над тремя полями с $p = 5$, 7 и 11 соответственно.

Сначала положим $p = 5$, так что $F_5 = \{0, 1, 2, 3, 4\}$. Для любого $x \in F_5$ мы ищем такое $y \in F_5$, квадрат которого удовлетворяет кубическому уравнению, и для этой цели строим таблицу квадратов в F_5, показанную в табл. 10.5.

Таблица 10.5. Квадраты по модулю 5

n	0	1	2	3	4
n^2	0	1	4	4	1

Таблица 10.6. Точки на кривой по модулю 5

x	$x^3 + x + 1 \stackrel{?}{=} n^2$	**n**	**(x, y)**
0	1	1,4	(0,1), (0,4)
1	3	–	–
2	1	1,4	(2,1), (2,4)
3	1	1,4	(3,1), (3,4)
4	4	2,3	(4,2), (4,3)

Для каждого из пяти возможных значений x мы вычисляем значение кубического многочлена и определяем, какие их них являются квадратами (см. табл. 10.6).

Мы имеем, следовательно, восемь (конечных) точек, показанных на рис. 10.10, а также оригинальную эллиптическую кривую; визуально их ничто не связывает, поэтому здесь и далее мы несколько расширяем концепцию «быть кривой».

Теперь возьмем $p = 7$ и поле F_7. В табл. 10.7 и 10.8 и на рис. 10.11 показаны жалкие четыре точки.

И наконец, для $p = 11$ и поля F_{11} мы получаем 13 точек, как следует из табл. 10.9 и 10.10 и рис. 10.12.

Мы видим, что очень немногие (а то и вообще никакие) точки лежат на оригинальной кривой и что в таблицах они естественно перечислены как обратные пары, которые на графиках образуют вертикальные пары, симметричные относительно горизонтальной прямой $x = \frac{1}{2}p$, за исключением самообратных точек, например (2, 0) на рис. 10.12.

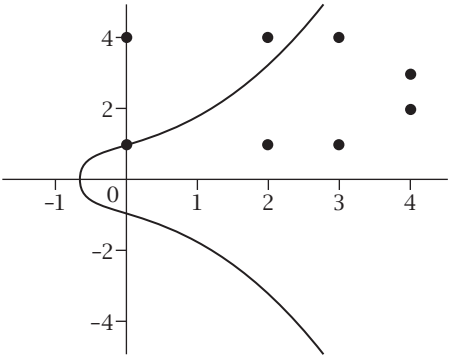


Рис. 10.10. Непрерывная и дискретная кривая над F_5

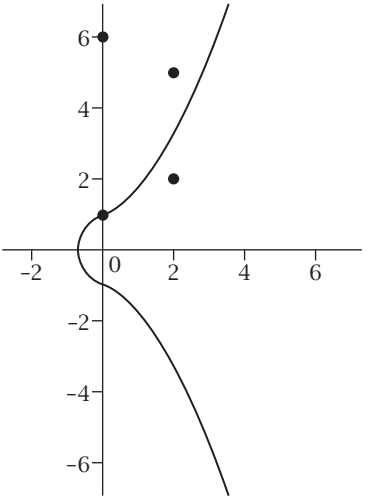


Рис. 10.11. Непрерывная и дискретная кривая над F_7

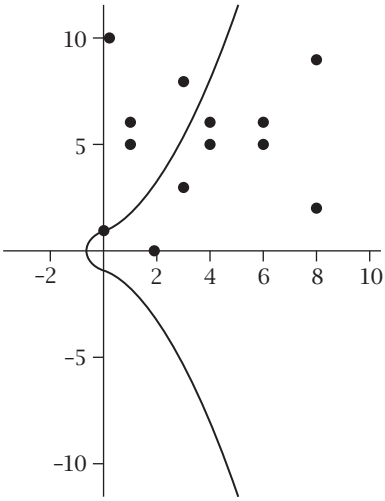


Рис. 10.12. Непрерывная и дискретная кривая над F_{11}

Таблица 10.7. Квадраты по модулю 7

n	0	1	2	3	4	5	6
n^2	0	1	4	2	2	4	1

Таблица 10.8. Точки на кривой по модулю 7

x	$x^3 + x + 1 \stackrel{?}{=} n^2$	n	(x, y)
0	1	1,6	(0,1),(0,6)
1	3	—	—
2	4	2,5	(2,2), (2,5)
3	3	—	—
4	6	—	—
5	5	—	—
6	6	—	—

Таблица 10.9. Квадраты по модулю 11

n	0	1	2	3	4	5	6	7	8	9	10
n^2	0	1	4	9	5	3	3	5	9	4	1

Таблица 10.10. Точки на кривой по модулю 11

x	$x^3 + x + 1 \stackrel{?}{=} n^2$	n	(x, y)
0	1	1,10	(0,1),(0,10)
1	3	5,6	(1,5),(1,6)
2	0	0	(2,0)
3	9	3,8	(3,3),(3,8)
4	3	5,6	(4,5),(4,6)
5	10	–	–
6	3	5,6	(6,5),(6,6)
7	10	–	–
8	4	2,9	(8,2),(8,9)
9	2	–	–
10	10	–	–

Надеемся, что наш подход к нахождению квадратов по модулю простого числа p помог вам разобраться, но он несистематичный. Гораздо более систематический подход состоит в том, чтобы воспользоваться одной из величайших теорем теории чисел, доказанной – что и не удивительно – одним из ее величайших столпов, Гауссом. Теорема берет начало в лемме Ферма, сформулированной в середине 1600-х гг., вновь появляется в виде недоказанной гипотезы Эйлера в 1744 г., затем снова в неполном доказательстве Лежандра в конце 1700-х гг. И наконец, 19-летний Гаусс дал полную формулировку и первое полное доказательство в 1797 г. Он был так горд этим результатом, что в конце концов предложил целых восемь его доказательств. Названная самим Гауссом *aureus theorema* (золотой теоремой), теперь она известна под названием *квадратичного закона взаимности Гаусса*. Основное ее содержание таково:

Пусть $p \neq q$ – нечетные простые числа. Тогда:

(i) если $p \equiv 1 \bmod 4$ или $q \equiv 1 \bmod 4$, то p является квадратом по модулю q тогда и только тогда, когда q является квадратом по модулю p;

(ii) если $p \equiv q \equiv 3 \bmod 4$, то p является квадратом по модулю q тогда и только тогда, когда q не является квадратом по модулю p.

К этой теореме есть два дополнения:

- если p – нечетное простое число, то -1 является квадратным корнем по модулю p тогда и только тогда, когда $p \equiv 1 \bmod 4$;
- если p – нечетное простое число, то 2 является квадратом по модулю p тогда и только тогда, когда $p \equiv 1, 7 \bmod 8$.

Существуют очевидные свидетельства потенциальной полезности этого важнейшего результата, но недостаток места не позволяет нам подробно рас-

сказать о них. Вместо этого наш путь ведет нас дальше, к основному применению эллиптических кривых.

Определив элементы каждой математической вселенной, мы должны взглянуть на применяемые к ним арифметические операции. Они, конечно, были определены нами ранее, но нужно внимательно следить за написанием выражений. Ниже приведен один из вариантов нотации.

Над полем F_p сумма $P_1 + P_2$ точек $P_1 = (x_1, y_1)$ и $P_2 = (x_2, y_2)$ на эллиптической кривой $y^2 = x^3 + ax + b$ записывается следующим образом.

- Если $x_1 \neq x_2$, то $P_1 + P_2 = (x_3, y_3)$, где

$$x_3 = [(y_2 - y_1)(x_2 - x_1)^{-1}]^2 - x_1 - x_2,$$

$$y_3 = -y_1 + (y_2 - y_1)(x_2 - x_1)^{-1}(x_1 - x_3).$$

- Если $x_1 = x_2$, то:
  - если $y_1 = -y_2$, то $P_1 + P_2 = O$;
  - если $y_1 \neq -y_2$, то $P_1 + P_2 = 2P_1 = (x_3, y_3)$, где

$$x_3 = ((3x_1^2 + a)(2y_1)^{-1})^2 - 2x_1,$$

$$y_3 = -y1 + (3x_1^2 + a)(2y_1)^{-1}(x_1 - x_3),$$

где все арифметические операции выполняются по модулю p.

Для иллюстрации приведем типичное вычисление при $p = 11$:

$$(8, 2) + (6, 6) \rightarrow \begin{cases} x_3 = [(6-2)(6-8)^{-1}]^2 - 8 - 6 \\ \quad = [4 \times 9^{-1}]^2 + 8 = (4 \times 5)^2 + 8 = 9^2 + 8 = 1, \\ y_3 = -2 + (6-2)(6-8)^{-1}(8-1) \\ \quad = 9 + 4 \times 5 \times 7 = 6. \end{cases}$$

И таким образом,

$$(8, 2) + (6, 6) = (1, 6).$$

Для интересующихся мы включили полную таблицу сложения для всех пар точек этой кривой, определенной над этим полем (см. табл. 10.11).

Наше графическое представление группы сопоставляет эллиптическую кривую, определенную над множеством вещественных чисел, с эллиптической «кривой», определенной только на положительной целочисленной сетке; для этой цели (в случае F_{11}) больше подходят представления на рис. 10.13a, а на рис. 10.13b показана описанная выше процедура сложения. Чтобы сложить две разные точки, соединим их прямой (разрывая ее с переносом на горизонтальной и вертикальной границах) и продолжим ее до тех пор, пока она не пройдет через следующую точку; отражение этой точки относительно горизонтальной линии симметрии и будет суммой. Седобородые старцы типа автора, наверное, помнят компьютерную игру «Астероиды», так что им такой заворот прямых знаком; думается нам, что и для более юных годами найдется подходящий эквивалент.

Таблица 10.11. Таблица сложения точек на кривой $y^2 = x^3 + x + 1$ над полем F_{11}

+	0	(0,1)	(0,10)	(1,5)	(1,6)	(2,0)	(3,3)	(3,8)	(4,5)	(4,6)	(6,5)	(6,6)	(8,2)	(8,9)
0	0	(0,1)	(0,10)	(1,5)	(1,6)	(2,0)	(3,3)	(3,8)	(4,5)	(4,6)	(6,5)	(6,6)	(8,2)	(8,9)
(0,1)	(0,1)	(3,3)	0	(4,5)	(2,0)	(1,5)	(6,6)	(0,10)	(8,2)	(1,6)	(3,8)	(6,5)	(8,9)	(4,6)
(0,10)	(0,10)	0	(3,8)	(2,0)	(4,6)	(1,6)	(0,1)	(6,5)	(1,5)	(8,9)	(6,6)	(3,3)	(4,5)	(8,2)
(1,5)	(1,5)	(4,5)	(2,0)	(3,3)	0	(0,1)	(8,2)	(1,6)	(6,6)	(0,10)	(4,6)	(8,9)	(6,5)	(3,8)
(1,6)	(1,6)	(2,0)	(4,6)	0	(3,8)	(0,10)	(1,5)	(8,9)	(0,1)	(6,5)	(8,2)	(4,5)	(3,3)	(6,6)
(2,0)	(2,0)	(1,5)	(1,6)	(0,1)	(0,10)	0	(4,5)	(4,6)	(3,3)	(3,8)	(8,9)	(8,2)	(6,6)	(6,5)
(3,3)	(3,3)	(6,6)	(0,1)	(8,2)	(1,5)	(4,5)	(6,5)	0	(8,9)	(2,0)	(0,10)	(3,8)	(4,6)	(1,6)
(3,8)	(3,8)	(0,10)	(6,5)	(1,6)	(8,9)	(4,6)	0	(6,6)	(2,0)	(8,2)	(3,3)	(0,1)	(1,5)	(4,5)
(4,5)	(4,5)	(8,2)	(1,5)	(6,6)	(0,1)	(3,3)	(8,9)	(2,0)	(6,5)	0	(1,6)	(4,6)	(3,8)	(0,10)
(4,6)	(4,6)	(1,6)	(8,9)	(0,10)	(6,5)	(3,8)	(2,0)	(8,2)	0	(6,6)	(4,5)	(1,5)	(0,1)	(3,3)
(6,6)	(6,6)	(6,5)	(3,3)	(8,9)	(4,5)	(8,2)	(3,8)	(0,1)	(4,6)	(1,5)	0	(0,10)	(1,6)	(2,0)
(8,2)	(8,2)	(8,9)	(4,5)	(6,5)	(3,3)	(6,6)	(4,6)	(1,5)	(3,8)	(0,1)	(2,0)	(1,6)	(0,10)	0
(8,9)	(8,9)	(4,6)	(8,2)	(3,8)	(6,6)	(6,5)	(1,6)	(4,5)	(0,10)	(3,3)	(1,5)	(2,0)	0	(0,1)

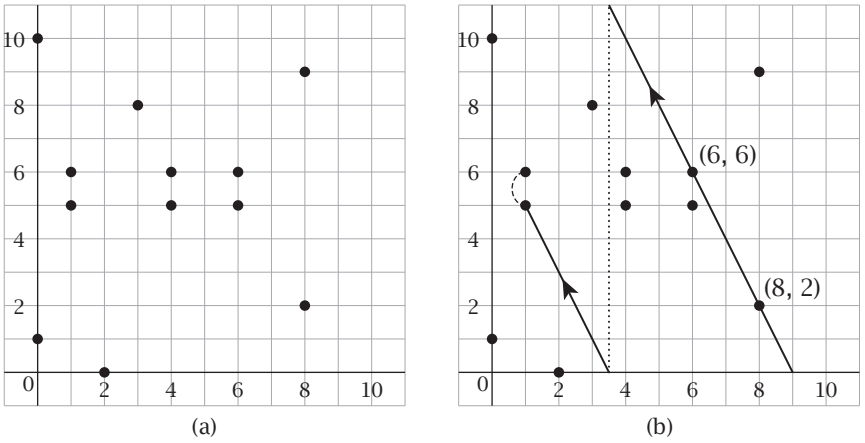


Рис. 10.13. (a) Эллиптическая кривая над F_{11}. (b) Пример сложения

Итак, мы видим, что в зависимости от поля (и кривой) количество точек на кривой существенно различается. Возникает очевидный вопрос: чему равно число точек $\#E(F_p)$ для заданного простого p и эллиптической кривой E?

В общем случае это открытая проблема, и лучшее, что можно сделать, – дать оценки, причем первая из них совсем проста. Если рассмотреть крайний случай, когда для каждого целого числа от 1 до $p - 1$ генерируется точка на графике и ни одна из этих точек не является самообратной, то всего будет $2p$ точек плюс бесконечно удаленная. Таким образом,

$$\#E(F_p) \leqslant 2p + 1.$$

Чтобы улучшить эту оценку, потребуется один из фундаментальных результатов соответствующей теории, который мы можем описать на интуитивном уровне. Внимательно приглядевшись к табл. 10.5, 10.7 и 10.9, можно заметить, что для каждого n числа n^2 и $(p - n)^2$ равны по модулю p, что с очевидностью следует из разложения $(p - n)^2 = p^2 - 2np + n^2$, и мы имеем естественное различие между элементами поля, которое приводит к важному определению.

Ненулевое число, сравнимое с квадратом по модулю p, называется *квадратичным вычетом* (КВ) по модулю p, а не сравнимое – *квадратичным невычетом* (КН) по модулю p. Из табл. 10.5 видно, что {1, 4} являются КВ, а {2, 3} – КН по модулю 5; из табл. 10.7 – что {1, 2, 4} являются КВ, а {3, 5, 6} – КН по модулю 7; из табл. 10.9 – что {1, 3, 4, 5, 9} являются КВ, а {2, 6, 7, 8, 10} – КН по модулю 11. На самом деле легко доказать, что:

если $p > 2$ простое, то существует ровно

$$\tfrac{1}{2}(p - 1) \text{ КВ и } \tfrac{1}{2}(p - 1) \text{ КН по модулю } p.$$

Зная это, мы можем предположить, что для больших p значения кубического выражения будут равномерно распределены между обеими возможностями: половина из них будет квадратичными вычетами. Если значение является КВ, то в общем случае будет сгенерировано два значения квадратичного выражения плюс бесконечно удаленная точка. То есть

$$\#E(F_p) \approx ½ \times p \times 2 + 1 = p + 1.$$

Этому предположению была придана строгость следующим результатом немецкого математика Гельмута Хассе в 1922 г.

Количество точек на эллиптической кривой $y^2 = x^3 + ax + b$ над конечным полем F_p удовлетворяет неравенству $|\#E(F_p) - (p + 1)| \leqslant 2\sqrt{p}$, откуда вытекает *интервал Хассе*:

$$p + 1 - 2\sqrt{p} \leqslant \#E(F_p) \leqslant p + 1 + 2\sqrt{p}.$$

То есть для больших p количество точек на кривой $\sim p$.

В (довольно большом) примере (Morain 2006), когда $p = 10^{2499} + 7131$ (это число действительно простое), а кривая имеет уравнение $y^2 = x^3 + 4589x + 91128$, $\#E(F_p) = p + 1 - t$, где выписанное ниже t мало по сравнению с p:

$$\begin{aligned}
&743119915223242448852593242753603560183870998734 53\\
&524190337127734742616052957456139347848273032219 28\\
&196336835756857318606333085947231334046337016503 47\\
&642609931708764997037635576407126373465428616355 30\\
&248560688747230776560970782387372349274130452135 88\\
&596512839070377985374614423232350452753408260919 26\\
&290612524515094221467986424645517930048054871163 60\\
&047431376657953293805586016188358341987968688933 91\\
&293204121353662006840136209644933588896320739874 00\\
&880836072043167819435435301254203874045015052903 92\\
&000068495427393032914624220033239147926141945021 24\\
&122343595679261259560456616043839757898379281360 25\\
&620011798249384004045008584520449871951575828394 36\\
&057153863826221227906256608278950318938988853308 12\\
&578313993269694618128437253459115977868025826425 29\\
&163013628536768647749494806629480269999989548358 31\\
&387765095297144723348697799906289840994365491033 56\\
&974032706070675024911460474847465294209029611323 03\\
&740576343364071957477085727098341529842061071267 56\\
&008468304449000961288194218319933018689619850760 29\\
&228733382357896594019878760506896270894774907173 66\\
&754410230986360942010122625495852602530360613170.
\end{aligned}$$

Почему мы заканчиваем раздел этим результатом и почему выбрали такое большое простое число в качестве примера? В развязке этой главы даются ответы именно на эти вопросы.

10.7. Криптография

Определив эллиптические кривые над конечными полями простого порядка, мы раскрыли еще один инструмент швейцарского ножа, поскольку теперь имеем аппарат, необходимый для их применения в криптографии.

Философия, стоящая за древней и вечно актуальной проблемой безопасных коммуникаций, эволюционировала от шифрования данных с помощью какой-то, предположительно секретной, системы до способа, в котором предполага-

ется, что о системе шифрования известно все, кроме ключа (или нескольких ключей), играющего основную роль в защите от вторжения.

В системе с *симметричным ключом* действуют стандартные криптографические персонажи Алиса и Боб, каждый из которых использует открытый процесс и общий закрытый ключ, что позволяет им общаться друг с другом, обезопасив зашифрованное сообщение от вездесущего подслушивающего противника, Евы. Схематически шифрование сообщения M ключом K, в результате которого создается зашифрованное сообщение C, можно обозначить $E_K(M) = C$, а процесс дешифрирования – $D_K(C) = M$. Как именно ключ используется для шифрования и дешифрирования, определяется выбранным алгоритмом. Принятый в настоящее время стандарт предполагает использование 128-битового ключа и называется AES (Advanced Encryption Standard – улучшенный стандарт шифрования) или, с большим уважением к персоналиям, Rijndael – по фамилиям двух бельгийских криптографов, придумавших его: Винсента Рэймана и Йоана Даймена. Составляющий его хитроумный набор процедур широко доступен в интернете, и мы не будем его здесь повторять, но после того как алгоритм шифрования и дешифрирования выбран, остается проблема выбора ключа, который был бы известен только сторонам A и B. И еще безопасной передачи этого ключа между сторонами.

Альтернативой схеме с симметричным ключом является схема с *открытым* ключом, когда и у Алисы, и у Боба имеется ключ, известный только им; схематически $E_{K_1}(M) = C$ и $D_{K_2}(C) = M$. Первая реализация шифрования с открытым ключом появилась в 1976 г.[1], когда Уолт Диффи и Мартин Хеллман изобрели метод (известный сейчас под названием DH) и его реализацию; в 1977 г. Рон Ривест, Ади Шамир и Леонард Адлеман (RSA) предложили альтернативный подход – но в обоих случаях в основе лежала вычислительная нереализуемость. В случае RSA авторы заметили, что, хотя умножение двух больших простых чисел не представляет проблемы для современных компьютеров, обратная задача – разложение большого составного числа на множители (факторизация) – может оказаться практически нерешаемой. В случае DH вычислительно нереализуемой стала *задача дискретного логарифмирования* (*discrete logarithm problem* – DLP). Конкретная постановка DLP зависит от окружения, но принцип заключается в том, что если $a = b^n$ (или $a = nb$) для известных a и b, то чрезвычайно трудно найти значение n. Мы не будем рассматривать воплощение этой идеи в протоколе DH, однако ее вариант в терминах точек на эллиптических кривых над конечным полем находится в центре внимания этого раздела, а вместо того чтобы каждый раз писать «задача дискретного логарифмирования на эллиптической кривой (elliptic curve discrete logarithm problem)», мы будем использовать труднопроизносимый, но стандартный акроним ECDLP: если на нашей эллиптической кривой взять начальную точку P и несколько раз сложить ее с самой собой, применяя метод касательной, так что в итоге получится nP, то при условии, что кривая, точка и простое число выбраны с необходимыми предосторожностями и n велико, будет исключительно трудно найти n, зная только P и nP. На основе этого предположения были придуманы методы безопасной передачи информации по небезопасному электронному

[1] Из многих альтернатив см. работы Steef et al. (2017) и King (2009).

каналу связи, хотя реализация, требующая многократного повторения, может оказаться непрактичной из-за слишком большого времени обработки.

Во-первых, предположим, что сообщение M, которое Алиса хочет передать Бобу, закодировано целым числом: это просто. Во-вторых, это целое число должно быть представлено точкой, вложенной в подходящую эллиптическую кривую. Это гораздо проще, чем можно подумать[1]. Итак, допустим, что все это проделано, и в результате получилась точка M(10, 12) на эллиптической кривой $y^2 = x^3 + 7742x + 734$, определенной над конечным полем простого порядка $p = 7901$. Вместе они определяют вторую точку на эллиптической кривой, P(5, 8), которая необязательно должна быть закрытой. Пока что единственная закрытая информация – это M, но Алиса, никому не сообщая, выбирает большое целое положительное число a, а Боб точно так же выбирает b, после чего они действуют следующим образом.

- Боб вычисляет точку bP и отправляет ее Алисе.
- Алиса вычисляет точки aP и M + a(bP) = M + abP и отправляет их Бобу.
- Зная aP, Боб вычисляет b(aP) = abP и вычитает результат из полученного M + abP, чтобы восстановить M.

Этот процесс обманчиво прост, а его безопасность опирается на предположение о том, что для вычисления abP = a(bP) = b(aP), зная P и член в скобках, Ева должна определить либо a, либо b, т. е. решить задачу ECDLP.

Например, предположим, что $a = 27$ и $b = 83$.

- Боб вычисляет bP = 83(5, 8) = (6602, 7629) и отправляет ее Алисе.
- Алиса вычисляет aP = 27(5, 8) = (213, 1529) и M = a(bP) = (10, 12) + 27(6602, 7629) = (10, 12) + (734, 475) = (4790, 5356) и отправляет результат Бобу.
- Боб вычисляет b(aP) = 83(213, 1529) = (734, 475) и вычитает результат из (4790, 5356), что дает сообщение (10, 12).

Отметим, что вычисления, выполненные Алисой и Бобом, совершенно другого порядка сложности, чем те, что предстоят выполнить Еве. Оба могут использовать «удвоение»; например, чтобы вычислить 83P, они могли бы сгенерировать последовательность

$$P + P = 2P \rightarrow 2P + 2P = 4P \rightarrow 4P + 4P = 8P \rightarrow 8P + 8P = 16P$$

$$\rightarrow 16P + 16P = 32P \rightarrow 32P + 32P = 64P,$$

а затем построить составную сумму:

$$83P = ((64P + 16P) + 2P) + P.$$

В итоге нужно 6 + 3 = 9, а не 82 операций сложения, но это уменьшение работы Еве недоступно. Впрочем, Еве все же доступны способы значительно сократить работу, однако за счет тщательного выбора компонентов системы их можно сделать относительно безвредными. А Алисе и Бобу нужно всего лишь представить коэффициент в двоичной форме, а затем удваивать и складывать; такие операции компьютеру выполнять – сплошное удовольствие.

ECC и RSA можно объединить, применяя первый метод для передачи общего ключа, который будет использоваться во втором. Этот обмен ключами – прямолинейный вариант описанного выше метода, в котором:

[1] Из многих альтернатив см. работы Steef et al. (2017) и King (2009).

- Алиса вычисляет точку aP и отправляет ее Бобу;
- Боб вычисляет точку bP и отправляет ее Алисе;
- зная bP, Алиса вычисляет $a(b$P);
- зная aP, Боб вычисляет $b(a$P) и получает ту же точку abP на кривой;
- Алиса и Боб извлекают ключ из abP, применяя ранее согласованный способ, например используя последние 256 бит координаты x.

Снова предположим, что $a = 27$ и $b = 83$:

- Алиса вычисляет aP = 27(5, 8) = (213, 1529) и отправляет ее Бобу;
- Боб вычисляет bP = 83(5, 8) = (6602, 7629) и отправляет ее Алисе;
- Алиса вычисляет $a(b$P) = 27(6602, 7629) = (734, 475);
- Боб вычисляет $b(a$P) = 83(213, 1529) = (734, 475);
- согласованным способом извлекается ключ.

Мы надеемся, что эти простые примеры помогли вам понять принцип и его реализацию, но, чтобы дать представление о сложностях, возникающих при организации связи в реальном мире, приведем сведения о двух распространенных эллиптических кривых.

Кривая для управления цифровыми правами Microsoft

Используемая для защиты прав интеллектуальной собственности на программное обеспечение и медиа, эта кривая имеет уравнение $y^2 = x^3 + ax + b$, где

$$a = 31768908125132550347631747641382769327274695592 7,$$

$$b = 79052896607878758718120572025718535432100651934,$$

и определена над конечным полем простого порядка

$$p = 78596310237942882237669478944689739620749856895 1.$$

Точка на кривой имеет координаты

$$P_x = 77150721626264982617064826856579889907769254176,$$

$$P_y = 39015751024655662852527946926651499556253319665 5.$$

Кривая содержит

$$78596310237942882237669302488171495761268615742 9$$

точек.

Все числа записаны в десятичной системе, хотя обычно числа (особенно длинные) записывают в шестнадцатеричной системе. Если мы дадим себе труд преобразовать указанное выше простое число в шестнадцатеричную систему, то обнаружим, что отвечавшие за кривую сотрудники Microsoft были не лишены чувства юмора; ведь выглядит это число так:

$$p = 89|ABCDEF|01234567|27182818|31415926|14142|4F7,$$

где вертикальные черточки разбивают среднюю его часть на дополнительные шестнадцатеричные цифры, первые восемь десятичных цифр (те и другие по порядку), а затем первые цифры десятичных представлений e, π и $\sqrt{2}$ соответственно.

Кривая P-384

Вторая эллиптическая кривая называется P-384, где P – первая буква слова «prime» (простое), но 384 – не порядковый номер простого числа (оно было бы слишком маленьким для наших целей, всего-то 2657). Это обозначение говорит, что мы имеем дело с обобщенным простым числом Мерсенна, в котором наивысшая степень 2 равна 384. Это число равно

$$p = 2^{384} - 2^{128} - 2^{96} + 2^{32} - 1$$

$$= 39402006196394479212279040100143613805079737270465744$$

$$66794829340424572171496870329047266088258938001861606$$

$$6973112319$$

в десятичном виде. Над полем такого порядка определена кривая с уравнением $y^2 = x^3 - 3x + b$, где

$$b = b3312fa7e23ee7e4988e056be3f82d19181d9c6ef\ e814112$$

$$0314088f5013875ac656398d8a2ed19d2a85c8edd3ec2aef.$$

Точка P на кривой имеет координаты

$$P_x = aa87ca22be8b05378eb1c71ef320ad746e1d3b628ba79b98$$

$$59f741e082542a385502f25bf55296c3a545e3872760ab7,$$

$$P_y = 3617de4a96262c6f5d9e98bf9292dc29f8f41dbd289a147c$$

$$e9da3113b5f0b8c00a60b1ce1d7e819d7a431d7c90ea0e5f.$$

А поскольку мы также знаем, что

$$\#E(F_p) = 39402006196394479212279040100143613805079$$

$$7392704679466679466079469052796276593991 1$$

$$3263569398956308152294913554433653942643,$$

то выбирать точки есть из чего.

Эта информация была взята с сайта АНБ (Агентства национальной безопасности США), где мы с чувством глубокого удовлетворения узнаем, что она защищает данные вплоть до грифа СОВЕРШЕННО СЕКРЕТНО.

Если нам не хочется производить столько вычислений и мы согласны на гриф СЕКРЕТНО, то можем использовать кривую $y^2 = x^3 + x + 61$ с $p = 10^{77} + 21$, для которой

$$\#E(F_p) = 10000000000000000000000000006131967$$

$$22711727187812910225277780387603.$$

Оставляем читателю в качестве упражнения найти хотя бы одну точку на ней!

Почему правительства, банки, крупные корпорации и все остальные довольствуются шифрованием с помощью эллиптических кривых? Вообще-то, есть у нас подозрение, что претензии у них имеются, но без электронных коммуни-

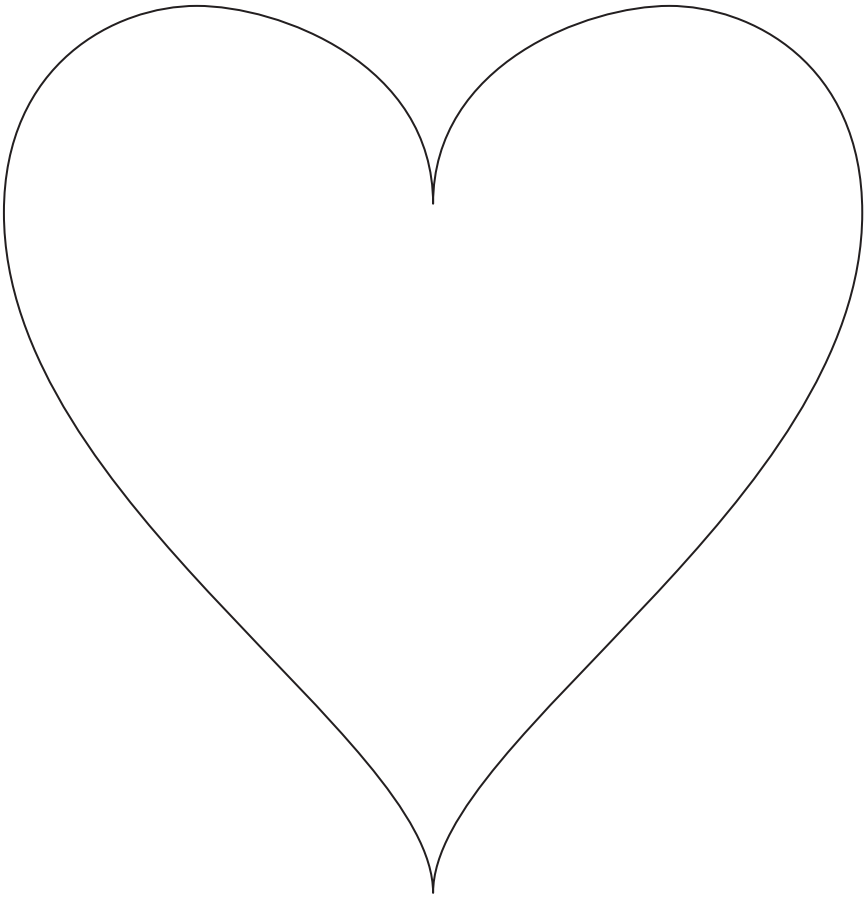
каций не обойтись, а в настоящее время эта защита считается безопасной, потому что самый быстрый из известных алгоритмов вскрытия ECDLP (*ро-алгоритм Полларда*) работает экспоненциальное время, ожидаемое время его выполнения пропорционально $\sqrt{p}$; при быстродействии современных компьютеров это означает, что решить задачу ECDLP практически невозможно при $p > 2^{160}$.

10.8. Апология

Мы подошли к концу обсуждения эллиптической кривой, но лишь едва коснулись ее истории, потому что она гораздо богаче, чем мы можем здесь рассказать. Для кривых над полем рациональных чисел мы совсем не обсудили важный показатель – число точек на эллиптической кривой (ее ранг) – и многие проблемы, связанные с измерением ее размера. Мы опустили обсуждение *теоремы Таннелла*, которая дает вычислимое необходимое условие конгруэнтности числа, свободного от квадратов, которое было бы также достаточным, если бы оказалась истинной (слабая) *гипотеза Бёрча и Свиннертон-Дайера*. И это одна из проблем тысячелетия, за решение которой объявлен приз в 1 000 000 долл. А еще есть *L-функции* и дзета-функции эллиптических кривых, которые связывают их с *гипотезой Римана* – еще одной проблемой тысячелетия. Для развития этих идей нам понадобилось бы гораздо больше страниц и пришлось бы обратиться к комплексным числам. Но размер книги ставит непреодолимый барьер, и мы решили ограничиться только декартовой плоскостью, поэтому избегали их всеми силами, а значит, вынуждены были отказаться от разных вещей, которые несет с собой синтез эллиптических кривых с этим конкретным полем. Как «выглядит» эллиптическая кривая над полем комплексных чисел? Как тор. И так же выглядит ассоциированная с ней группа $E(\mathbb{C})$. А еще надо бы рассмотреть комплексные интегралы и понятие о функции с двумя отдельными периодами, которая ведет к *℘-функции Вейерштрасса* (мы подозреваем, что так он писал букву «р» и произносится она как «р», но по какой-то причине осталась неизменной по прошествии времени). А со всем этим связаны *модулярные формы*, которые возвращают нас к эллиптическим кривым, потому что, в силу гипотезы Ютака Танияма и Горо Шимура, это одно и то же, каким бы удивительным это ни казалось: если бы Великая теорема Ферма была неверна и существовали бы целые числа, для которых $a^n + b^n = c^n$ для $n > 3$, то, как показал в 1986 г. Кен Рибет, эллиптическая кривая $y^2 = x(x - a^n)(x + b^n)$ не была бы модулярной, но в 1994 г. Эндрю Уайлз доказал, что всякая эллиптическая кривая модулярна.

Итак, мы приносим извинения за то, что включили в эту антологию кривую, о которой не смогли толком рассказать, но думаем, что опустить ее было бы куда большим прегрешением. Памятуя о неугрозе Сержа Лэнга, мы на этом закончим – главу и книгу.

Быть может, самая важная кривая



$$x^2 + (\frac{5}{4}y - \sqrt{|x|})^2 = 1$$

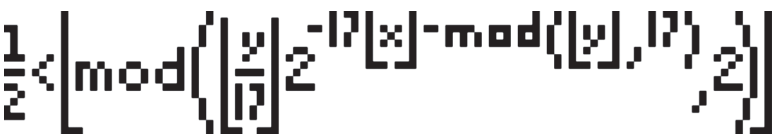
Приложение А. Титульный лист

Мы полагаем, что возбудили любопытство читателя двумя блоками цифр в начале книги. Они занимают первую печатную страницу после титульного листа, и это самое подходящее место, потому что представляемые ими слова и *есть* название книги.

В статье 2001 года Джефф Таппер с факультета информатики Торонтского университета опубликовал удивительный результат, согласно которому множество всех точек (x, y) на плоскости, для которых

$$\frac{1}{2} < \left\lfloor \mathrm{mod}\left(\left\lfloor \frac{1}{17} y \right\rfloor 2^{-17\lfloor x \rfloor - \mathrm{mod}(\lfloor y \rfloor, 17)}, 2\right)\right\rfloor,$$

где $0 \leqslant x < 106$ и $N \leqslant y < N + 17$, имеет вид:



То есть условие порождает само себя в этой прямоугольной области плоскости. Для этого, правда, прямоугольник должен находиться довольно высоко по оси y; точнее число N равно[1]

> 48584506361897134235820959624942020445814005879832445494
> 83093085061934704708809928450644769865524364849997247024
> 91511911041160573917740785691975432657185544205721044573
> 58836818298237541396343382251994521916512843483329051311
> 93199953502413758765239264874613394906870130562295813219
> 48111368533953556529085002387509285689269455597428154638
> 65107300491067230589335860525440966643512653493636439571
> 25565695936815184334857605266940161251266951421550539554
> 51915378545752575659074054015792900176596796548006442782
> 91314885482599147212485063526866304763 00.

Мы надеемся, что читателю, незнакомому с этим феноменом, факт покажется удивительным. Но скрыто в нем гораздо больше. Для прямоугольника размером 106 × 17 = 1802 любой из возможных в нем пиксельных паттернов (иногда их называют Tupperware) также можно воспроизвести с помощью формулы, если расположить прямоугольник на подходящем расстоянии N над осью y. Два больших числа на титульной странице – это соответственно значения N для областей, занимаемых фразами «Curves for the Mathematically»

[1] Необязательно именно такое, но оно должно быть большим. Часто приводят альтернативу 960…719, хотя для математика изображение при этом, естественно, оказалось бы перевернутым.

и «Curious», а два числа взято, потому что ширины одной области недостаточно для размещения текста:

CURVES FOR THE MATHEMATICALLY CURIOUS

Учитывая, что определяющее условие связывает координаты точек на плоскости, мы позволили себе сомнительную роскошь считать генерируемые им изображения кривыми, но взамен чувствуем себя обязанными объяснить читателю математику, стоящую за этим кажущимся мистическим феноменом, эксперименты с которым дают отличную возможность скоротать несколько часов свободного времени. Разберем это условие по частям.

Во-первых, функция *целой части*, которая встречается несколько раз и обозначается $\lfloor\alpha\rfloor$, равна наибольшему целому числу, меньшему или равному α. На самом деле самые внешние скобки необязательны, потому что неравенство $\frac{1}{2} < \lfloor\alpha\rfloor$ эквивалентно $\alpha \geqslant 1$, и значит, условие принимает вид:

$$\mathrm{mod}\left(\left\lfloor \tfrac{1}{17} y \right\rfloor 2^{-17\lfloor x\rfloor - \mathrm{mod}\,(\lfloor y\rfloor,17)}, 2\right) \geqslant 1.$$

Далее (как отмечает Таппер в статье 2001 г.) $\lfloor \frac{1}{17} y\rfloor = \lfloor \frac{1}{17}\lfloor y\rfloor\rfloor$, и потому связь имеет место между $\lfloor x\rfloor$ и $\lfloor y\rfloor$, а не между $\lfloor x\rfloor$ и y. Это означает, что для каждой точки сетки (x, y) в изображении определен целый единичный пиксель, противоположными углами которого является эта точка и $(x + 1, y + 1)$, и потому мы можем ограничиться самими точками сетки (x, y), каждая из которых порождает соответствующий пиксель 1×1. Условие принимает вид:

$$\mathrm{mod}\left(\left\lfloor \tfrac{1}{17}\lfloor y\rfloor \right\rfloor 2^{-17\lfloor x\rfloor - \mathrm{mod}\,(\lfloor y\rfloor,17)}, 2\right) \geqslant 1,$$

и далее:

$$\mathrm{mod}\left(\left\lfloor \tfrac{1}{17} y \right\rfloor 2^{-17x - \mathrm{mod}\,(y,17)}, 2\right) \geqslant 1,$$

где теперь предполагается, что переменные принимают целые положительные значения.

Далее, выражение $\mathrm{mod}(y, 17)$ на жаргоне компьютерщиков то же самое, что математики записывают как r в равенстве $y = 17q + r$, означающем, что это остаток от деления на 17, т. е. целое число $0 \leqslant r < 17$. Раз так, то условие можно записать в виде:

$$\mathrm{mod}\left(q 2^{-17x-r}, 2\right) = \mathrm{mod}\left(\frac{q}{2^{17x+r}}, 2\right) \geqslant 1.$$

Если мы теперь рассмотрим двоичную форму целого числа q, то частное $q/2^{17x+r}$ определяет десятичную точку справа от цифры с номером $17x + r$ (считая справа налево и сопоставляя самой правой цифре номер 0), и неравенство раз-

личает цифры 0 и 1. Точка сетки, а значит, и пиксель, рисуется только при условии, что эта цифра q равна 1. Если мы хотим, чтобы условие изобразило конкретный растр, то представим себе, что мы рисуем сетку 106×17 и заполняем ее соответствующими пикселями. Теперь заменим нарисованные пиксели единицами, а остальные – нулями; из получившихся цифр сформируем большое двоичное число некоторым систематическим образом (например, сверху вниз, слева направо, сверху вниз, справа налево и т. д.), преобразуем его в десятичную систему – и вот вам q; умножим на 17 – и вот оно, N.

Как с любым волшебным фокусом, после объяснения волшебство пропадает, а остается только фокус. Ведь математик – это фокусник, который раскрывает свои секреты[1].

[1] Слова Джона Хортона Конвея.

Приложение B. Все конические сечения в одном флаконе

Мы не включили конические сечения отдельной главой, но, по крайней мере, отдадим им дань уважения, охарактеризовав одним дифференциальным уравнением.

Эта идея, кажется, восходит к французскому математику Гаспару Монжу (Monge 1810).

В следующих строчках мы не будем каждый раз заводить новые буквы для постоянных, а будем повторно использовать уже имеющиеся.

Начнем с общего уравнения конического сечения (*c* считается положительным):

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

и будем рассматривать его как квадратное уравнение относительно *y*:

$$cy^2 + (e + bx)y + (ax^2 + dx + f) = 0.$$

Корень со знаком плюс равен

$$\begin{aligned} y &= \frac{-(e + bx) + \sqrt{(e + bx)^2 - 4c(ax^2 + dx + f)}}{2c} \\ &= ax + b + \sqrt{Ax^2 + 2Bx + C}. \end{aligned}$$

Продифференцировав по *x*, получим

$$y' = a + \frac{Ax + B}{\sqrt{Ax^2 + 2Bx + C}},$$

а продифференцировав второй раз –

$$\begin{aligned} y'' &= \frac{A\sqrt{Ax^2 + 2Bx + C} - (Ax + B)^2/\sqrt{Ax^2 + 2Bx + C}}{Ax^2 + 2Bx + C} \\ &= \frac{A(Ax^2 + 2Bx + C) - (Ax + B)^2}{(Ax^2 + 2Bx + C)^{3/2}} \\ &= \frac{AC - B^2}{(Ax^2 + 2Bx + C)^{3/2}} = (AC - B^2)(Ax^2 + 2Bx + C)^{-3/2}. \end{aligned}$$

Теперь возведем обе части в степень –⅔:

$$(y'')^{-2/3} = (AC - B^2)^{-2/3}(Ax^2 + 2Bx + C).$$

И наконец продифференцируем трижды, получив в результате, что все конические сечения являются решениями дифференциального уравнения:

$$((y'')^{-2/3})''' = 0.$$

Дифференцируя и упрощая, приходим к замечательному результату:

$$9\left(\frac{d^2y}{dx^2}\right)^2\frac{d^5y}{dx^5} - 45\frac{d^2y}{dx^2}\frac{d^3y}{dx^3}\frac{d^4y}{dx^4} + 40\left(\frac{d^3y}{dx^3}\right)^3 = 0.$$

Это дифференциальное уравнение описывает все варианты всех конических сечений, хотя в полезности такого подхода мы не уверены.

Приложение C. Тригонометрический вариант кривой Безье

Для варианта

$$f_{3,i}(t) = \alpha_i \sin \tfrac{1}{2}\pi t + \beta_i \cos \tfrac{1}{2}\pi t + \gamma_i \sin^2 \tfrac{1}{2}\pi t + \delta_i \cos^2 \tfrac{1}{2}\pi t \qquad \text{(C.1)}$$

условия в концевых точках дают

$$\beta_1 + \delta_1 = 0,\ \alpha_1 + \gamma_1 = 1,$$
$$\beta_2 + \delta_2 = 0,\ \alpha_2 + \gamma_2 = 1,$$
$$\beta_3 + \delta_3 = 0,\ \alpha_3 + \gamma_3 = 1.$$

Первая производная равна

$$\begin{aligned} f'_{3,i}(t) = {} & \tfrac{1}{2}\pi\alpha_i \cos \tfrac{1}{2}\pi t - \tfrac{1}{2}\pi\beta_i \sin \tfrac{1}{2}\pi t \\ & + 2\gamma_i \sin \tfrac{1}{2}\pi t \times \tfrac{1}{2}\pi \cos \tfrac{1}{2}\pi t \\ & - 2\delta_i \cos \tfrac{1}{2}\pi t \times \tfrac{1}{2}\pi \sin \tfrac{1}{2}\pi t \\ = {} & \tfrac{1}{2}\pi\alpha_i \cos \tfrac{1}{2}\pi t - \tfrac{1}{2}\pi\beta_i \sin \tfrac{1}{2}\pi t \\ & + \tfrac{1}{2}\pi\gamma_i \sin \pi t - \tfrac{1}{2}\pi\delta_i \sin \pi t. \end{aligned}$$

Соответствующие условия приводят к

$$\alpha_2 = 0 \to \gamma_2 = 1,$$
$$\alpha_3 = 0 \to \gamma_3 = 1.$$

Вторая производная равна

$$\begin{aligned} f''_{3,i}(t) = {} & -\tfrac{1}{4}\pi^2\alpha_i \sin \tfrac{1}{2}\pi t - \tfrac{1}{4}\pi^2\beta_i \cos \tfrac{1}{2}\pi t \\ & + \tfrac{1}{2}\pi^2\gamma_i \cos \pi t - \tfrac{1}{2}\pi^2\delta_i \cos \pi t. \end{aligned}$$

Соответствующее условие упрощается и принимает вид:

$$\beta_3 - 2\gamma_3 + 2\delta_3 = 0.$$

Решение этой системы уравнений имеет вид:

$$\begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \gamma_1 & \gamma_2 & \gamma_3 \\ \delta_1 & \delta_2 & \delta_3 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & -2 \\ -1 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

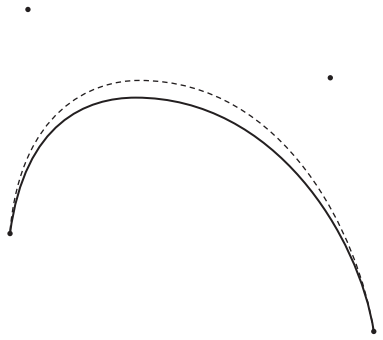
Подставляя эти значения в уравнение (C.1), получаем все три функции коэффициентов:

$$f_{3,1}(t) = 2\sin\tfrac{1}{2}\pi t - \sin^2\tfrac{1}{2}\pi t,$$
$$f_{3,2}(t) = \sin^2\tfrac{1}{2}\pi t,$$
$$f_{3,3}(t) = -2\cos\tfrac{1}{2}\pi t + \sin^2\tfrac{1}{2}\pi t + 2\cos^2\tfrac{1}{2}\pi t,$$

что после преобразования к форме Бернштейна дает уравнение кривой:

$$\begin{aligned}\boldsymbol{r}(t) = {} & [1 - 2\sin\tfrac{1}{2}\pi t + \sin^2\tfrac{1}{2}\pi t]\boldsymbol{p}_0 \\ & + [2\sin\tfrac{1}{2}\pi t - 2\sin^2\tfrac{1}{2}\pi t]\boldsymbol{p}_1 \\ & + [2\cos\tfrac{1}{2}\pi t - 2\cos^2\tfrac{1}{2}\pi t]\boldsymbol{p}_2 \\ & + [-2\cos\tfrac{1}{2}\pi t + \sin^2\tfrac{1}{2}\pi t + 2\cos^2\tfrac{1}{2}\pi t]\boldsymbol{p}_3.\end{aligned}$$

На рисунке ниже показаны графики кубической кривой Безье (сплошная линия) и этого тригонометрического варианта (пунктирная линия) для одного произвольно выбранного множества из четырех точек: различие в данном и общем случае едва ли стоит затраченных усилий!



Приложение D. Огибающие

Приведенное в этой книге посвящение заимствовано из 67-страничной книжечки Эдит Сомервелл под названием «Ритмический подход к математике» (Somervell 1906). Она была переиздана в 1975 г. в виде пятого тома серии «Classics in Mathematical Education»[1] под патронажем Национального совета преподавателей математики. Посвящение выражает чувства, которые мы разделяем, а сама работа имеет прямое отношение к этой книге и, должно быть, предлагала прекрасный материал для мотивации некоторых везучих детей того времени. В ней есть и вступительное слово, последний абзац которого мы приводим ниже:

> «Описанный здесь метод имеет важное преимущество перед многими другими видами реформы образования; его могут практиковать женщины, не вызывая никаких волнений и публичных обсуждений. Нам не нужно ждать постановлений парламента или разрешения школьных инспекторов. Любая дама, которая захочет потратить несколько часов на изучение пути, рекомендуемого миссис Сомервелл, сможет затем обучать деревенских детишек в процессе игры на каникулах. Материалы дешевые, аппарат простой, а работа будет интересна почти всем детям. К тому же уроки можно проводить на открытом воздухе, без столов и стульев. Маленькие учителя и маленькие ученики, усевшиеся в круг на земле, у их ног на траве разбросаны цветные тряпочки, юные глаза горят нетерпеливым любопытством увидеть, какой узор получится в следующий раз, – вот картина, приятная для созерцания».

Это написала Мэри Эверест Буль. Второе имя совпадает с фамилией ее дяди, Джорджа Эвереста, в честь которого названа гора Эверест, а фамилия принадлежит мужу, Джорджу Булю, которого мы помним куда лучше, в основном по ассоциациям с булевой алгеброй. Так что же это за образовательная нирвана, приведшая в такой восторг миссис Буль? Миссис Сомервелл называла ее *вышиванием по кривой*, и обе дамы восхваляют использование кривого стежка как средства получить как эстетическое удовлетворение, так и подсознательное ощущение узора, которое тем самым создает гармонию и демонстрирует взаимосвязи между объектами: *ритмический* подход к математике, пропагандируемый и, быть может, изобретенный миссис Буль.

Вместе с книгой продавался набор карточек с дырочками в нужных местах, которые дети могли соединять прямыми отрезками нити, образуя различные кривые, в том числе кривые преследования и пятиугольник Рёло.

Математически такое вышивание по кривой связано с *огибающими*, нужно только заменить пробитые карточки и красивые цветные ниточки компьютерными распечатками – и математикой, слишком сложной для детей, о которых пеклась Сомервелл.

Мы начнем с обратного процесса: дана дифференцируемая кривая, а требуется провести касательные к ней во многих точках, построив сеть из прямых

[1] Классики математического образования. – *Прим. перев.*

линий, в которых кривая покоится как в колыбельке; это показано на рис. D.1 для кривой с уравнением $y = x^2$. Каждая касательная описывается уравнением вида $y - a^2 = 2a(x - a)$, поэтому семейство касательных можно параметризовать: $y = 2ax - a^2$, а поскольку касательные огибают кривую, допустимо сказать, что кривая является их *огибающей*.

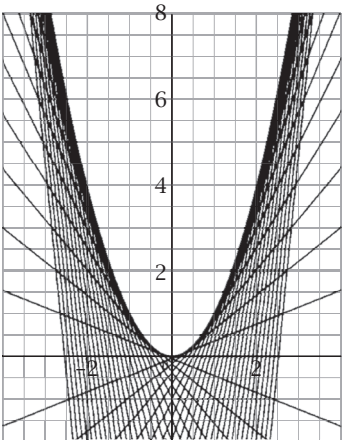


Рис. D.1. Огибающая $y = x^2$

Но наша цель противоположна: зная однопараметрическое семейство прямых, вывести уравнение их огибающей. Для этого еще раз рассмотрим наше семейство касательных и выберем две близких друг к другу; первой пусть будет прямая с написанным выше уравнением, а второй – с уравнением $y = 2(a + h)x - (a + h)^2$ для малого h. Эти кривые пересекутся в точке, близкой к точке на искомой кривой. Полагая $2(a + h)x - (a + h)^2 = 2ax - a^2$, после упрощения получаем $h(2x - 2a - h) = 0$, так что это уравнение имеет единственное решение $x = a + \frac{1}{2}h$. Подставляя его в уравнение каждой касательной, находим $y = a^2 + ah$. Итак, точка с координатами $(a + \frac{1}{2}h, a^2 + ah)$ находится сколь угодно близко к нашей кривой, а в пределе при $h \to 0$ будет лежать на ней, т. е. точка (a, a^2) лежит на кривой, и, стало быть, кривая имеет уравнение $y = x^2$.

Искушению поэкспериментировать с другими однопараметрическими семействами прямых противиться, конечно, очень трудно: если, например, мы хотим сгенерировать кривую $y^2 = x$, то должны рассмотреть семейство $y = ax + 1/(4a)$, так что в результате получится рис. D.2.

Обобщая, можно сказать, что само однопараметрическое семейство могло бы состоять из кривых. Так, меняя роли на противоположные, мы получаем, что семейство парабол $y = ax^2 + 1/(4a)$ порождает пару прямых линий $y = \pm x$.

Если описанная процедура напоминает вам определение дифференцирования по параметру, то удивляться нечему – ведь это оно и есть. Мы можем обобщить семейство прямых, заменив его семейством кривых $F(x, y, a) = 0$ и взяв в качестве ближнего члена семейства $F(x, y, a + h) = 0$, так что $(F(x, y, a + h) - F(x, y, a))/h = 0$. Устремив $h \to 0$, получим $\partial F(x, y, a) / \partial a = 0$. Систему уравнений

$$F(x, y, a) = 0 \quad \text{и} \quad \frac{\partial F(x, y, a)}{\partial a} = 0$$

можно использовать для определения огибающей кривых из семейства $F(x, y, a) = 0$, уравнения которой находятся путем исключения (при возможности) параметра из двух уравнений.

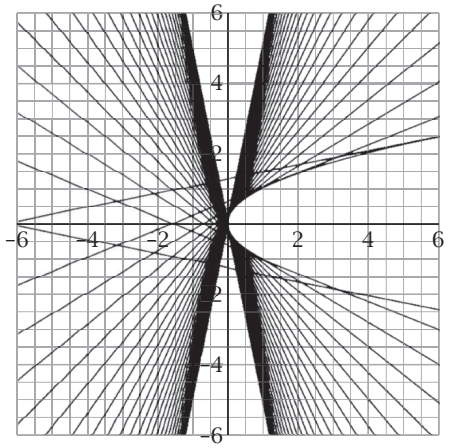


Рис. D.2. Огибающая $y^2 = x$

Простейшее построение миссис Сомервелл для детей – то, что известно также под названием *задачи о падающей лестнице*: какая кривая порождается, когда лестница единичной длины соскальзывает вдоль вертикальной стены?

На рис. D.3 лестница представлена лежащими в первом квадранте отрезками прямых из семейства $x/a + y/(1 - a) = 1$, проходящих через точки $(a, 0)$ и $(0, 1 - a)$, которое можно переписать в виде $ay = (a - 1)(x - a) = ax - a^2 - x + a$. Его производная по a равна $y = x - 2a + 1$, и после подстановки $a = \frac{1}{2}(x + 1 - y)$ в исходное уравнение и упрощений приходим к окончательной форме $x^2 + y^2 - 2xy - 2x - 2y + 1 = 0$, уравнению от x и y степени 2, представляющему коническое сечение. Его дискриминант равен $(-2)^2 - 4 \times 1 \times 1 = 0$, а значит, это парабола.

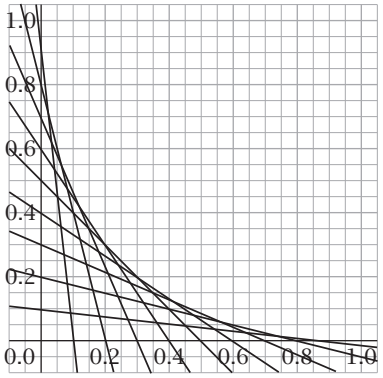


Рис. D.3. Падающая лестница

Попытавшись использовать эту идею в главе 1, мы не сумеем исключить параметр, так что остаемся с параметрическими уравнениями огибающей.

Приложение E. Математика арки

Мы докажем, что если двумерная конструкция, образующая арку, подвергается сжатию только в вертикальном направлении, то арка принимает форму цепной линии.

На рис. E.1a показано вертикальное сечение арки, сделанной из материала весом w на единицу длины. Система координат выбрана в соответствии с задачей и так, что граничные условия имеют вид $x = 0$, $y = 0$, $\mathrm{d}y/\mathrm{d}x = 0$. На рис. E.1b увеличен элемент длины $\mathrm{d}s$, на который действуют три силы: его вес, $w\,\mathrm{d}s$, верхняя сила P, направленная вниз, и компенсирующая сила $P + \mathrm{d}P$. Мы требуем, чтобы форма была такой, чтобы сечение не подвергалось горизонтальному растяжению, т. е. конструкция оставалась устойчивой под собственным весом. Проецируя силы на горизонтальную ось, мы можем записать это условие в виде:

$$P\cos\psi\,\mathrm{d}(P\cos\psi) = P\cos\psi \rightarrow \mathrm{d}(P\cos\psi) = 0 \rightarrow P\cos\psi = k.$$

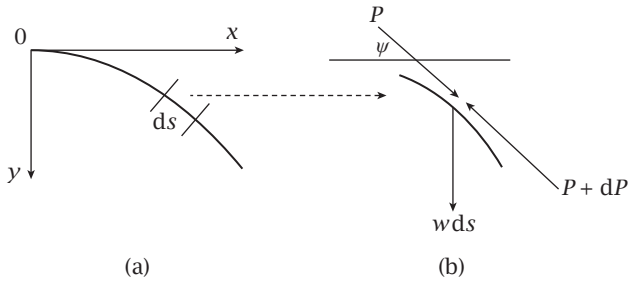


Рис. E.1. Внутренние силы, действующие на арку

В верхней точке арки $\psi = 0$ и $P = P_0$, откуда следует, что $P_0 \cos 0 = P_0 = k$, и значит, $P\cos\psi = P_0$. Кроме того, по правилу дифференцирования произведения $\mathrm{d}P\cos\psi - P\sin\psi\,\mathrm{d}\psi = 0$.

Теперь спроектируем силы на ось y и запишем условие равновесия:

$$w\,\mathrm{d}s + P\sin\psi = P\sin\psi + \mathrm{d}(P\sin\psi) \rightarrow w\,\mathrm{d}s = \mathrm{d}(P\sin\psi).$$

По правилу дифференцирования произведения

$$\begin{aligned} w\,\mathrm{d}s &= \mathrm{d}P\sin\psi + P\cos\psi\,\mathrm{d}\psi \\ &\rightarrow w\,\mathrm{d}s = \frac{P\sin\psi}{\cos\psi}\,\mathrm{d}\psi + P\cos\psi\,\mathrm{d}\psi \\ &\rightarrow w\,\mathrm{d}s = \frac{P_0\sin^2\psi}{\cos^2\psi}\,\mathrm{d}\psi + P_0\,\mathrm{d}\psi = \frac{P_0}{\cos^2\psi}\,\mathrm{d}\psi \end{aligned}$$

$$\rightarrow \frac{\mathrm{d}s}{\mathrm{d}\psi} = \frac{P_0}{w}\sec^2\psi$$
$$\rightarrow s = \frac{P_0}{w}\tan\psi = a\ \mathrm{tg}\ \psi.$$

И мы получили внутреннее уравнение продольной кривой арки. Следуя стандартной процедуре, перепишем его в декартовых координатах:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \mathrm{tg}\,\psi = \frac{s}{a} \rightarrow \frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{1}{a}\frac{\mathrm{d}s}{\mathrm{d}x}$$

и

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 \rightarrow \frac{\mathrm{d}s}{\mathrm{d}x} = \sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2},$$

откуда

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{1}{a}\sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2} \rightarrow \frac{\mathrm{d}y'}{\mathrm{d}x} = \frac{1}{a}\sqrt{1 + y'^2}, \qquad \text{где} \quad y' = \frac{\mathrm{d}y}{\mathrm{d}x}.$$

Разделяя переменные, получаем

$$\int \frac{1}{\sqrt{1 + y'^2}}\,\mathrm{d}y' = \frac{1}{a}\int 1\,\mathrm{d}x \rightarrow \mathrm{sh}^{-1}\,y' = \frac{x}{a} + c.$$

Граничное условие $x = 0$, $y' = 0$ дает $c = 0$, поэтому $\mathrm{sh}^{-1}\,y' = x/a$. Отсюда $y' = \mathrm{sh}(x/a)$ и потому $y = a\ \mathrm{ch}(x/a) + c$, где постоянная интегрирования вычисляется в терминах высоты арки. Итак, цепная линия.

Приложение F. Простой маятник

В качестве примера использования эллиптических интегралов и самой хитроумной подстановки в мире мы упоминали, что в движении простого маятника участвуют эллиптические кривые, а теперь увидим почему.

Во всех учебниках приводится элементарное доказательство того, что движение маятника описывается дифференциальным уравнением $\mathrm{d}^2\theta/\mathrm{d}t^2 = -k^2 \sin\theta$, где θ – угол, который маятник составляет с вертикалью в момент t. Обычно предполагают, что θ мал, так что $\sin\theta \sim \theta$ (в радианах), поэтому дифференциальное уравнение можно упростить, записав в виде $\mathrm{d}^2\theta/\mathrm{d}t^2 = -k^2\theta$, которое легко решается в отличие от анализа общего случая, когда мы имеем

$$\begin{aligned}
\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} = -k^2 \sin\theta \rightarrow \int \frac{\mathrm{d}^2\theta}{\mathrm{d}t^2}\,\mathrm{d}\theta &= \int \frac{\mathrm{d}^2\theta}{\mathrm{d}t^2}\frac{\mathrm{d}\theta}{\mathrm{d}t}\,\mathrm{d}t = \int -k^2 \sin\theta\,\mathrm{d}\theta \\
&\rightarrow \frac{1}{2}\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2 = k^2\cos\theta \\
&\rightarrow \frac{\mathrm{d}\theta}{\mathrm{d}t} = \sqrt{2}k\sqrt{\cos\theta} \\
&\rightarrow \int \frac{1}{\sqrt{\cos\theta}}\,\mathrm{d}\theta = \int \sqrt{2}k\,\mathrm{d}t.
\end{aligned}$$

И у нас есть возможность использовать вышеупомянутую подстановку $x = \mathrm{tg}\,\tfrac{1}{2}\theta$, так что

$$\cos\theta = \frac{1-x^2}{1+x^2} \quad \text{и} \quad \frac{\mathrm{d}\theta}{\mathrm{d}x} = \frac{2}{1+x^2}.$$

После упрощения остается

$$2\int \frac{1}{\sqrt{1-x^4}}\,\mathrm{d}x = \sqrt{2}kt,$$

эллиптическая кривая в знаменателе и эллиптический интеграл, с которым нужно разбираться.

Приложение G. Метод Фибоначчи

Ниже описана современная интерпретация метода Фибоначчи порождения трехчленных арифметических прогрессий рациональных квадратов. Для удобства мы разбили описание на несколько частей.

Сумма $2k$ положительных нечетных чисел с центром в четном числе $2a$.
Такие числа можно записать в виде последовательности $2a \pm 1$, $2a \pm 3$, $2a \pm 5$, …, $(2a \pm (2k - 1)$, сумма которой равна $4a \times k = 4ak$.

Потребуем, чтобы нижняя последовательность содержала только положительные нечетные числа.
Это означает, что ее наименьшее число должно быть больше или равно 1: $2a - (2k - 1) \geqslant 1 \Rightarrow a \geqslant k$.

Повторить дважды и получить ту же сумму.
Рассмотрим нижнюю последовательность слева с центром в $2a$, имеющую K членов, и верхнюю последовательность справа с центром в $2A$, имеющую k членов. Тогда $a < A$, $K > k$ и $4aK = 4Ak \Rightarrow aK = Ak$.

Потребуем, чтобы обе последовательности нечетных чисел соседствовали.
Это означает, что наибольший член нижней последовательности является нечетным числом, непосредственно предшествующим наименьшему члену верхней последовательности. В символьном виде $- 2a + (2K - 1) + 2 = 2A - (2k - 1)$, и потому $a + K = A - k$.

Применить результат о сумме последовательности нечетных чисел.

$$\sum_{r=1}^{n} (2r - 1) = n^2.$$

Эти наблюдения позволяют поместить заданное четное число вида $4aK = 4Ak$ между соседними нечетными числами, которые являются наибольшим членом нижней последовательности и наименьшим членом верхней. И именно этим фактом мы воспользуемся.

Примеры

1. Сначала возьмем в качестве разности арифметической прогрессии число 840, имеющее вид $pq(p + q)(p - q)$, где $p = 7$, $q = 3$ (см. стр. 183); в этом случае $4aK = 4Ak = 840 \Rightarrow aK = Ak = 210$. Значит, мы должны рассмотреть множители числа 210, решая, о какой букве – заглавной или строчной – идет речь, в зависимости от контекста. В табл. G.1 показан итоговый результат.

Таблица G.1. Разность арифметической прогрессии 840

a	210	105	70	42	35	30	21	15
k	1	2	3	5	6	7	10	14
$a + k$	211	107	73	47	41	**37**	31	*29*
$a - k$	209	103	67	**37**	*29*	23	11	1

Приведенное выше предпоследнее условие заставляет нас искать в третьей строке число, повторяющееся в четвертой строке, и таких случаев два: один, когда центр нижней последовательности $a = 15$, а $K = 14$, а для верхней последовательности центр $A = 35$ и $k = 6$; второй, когда центр нижней последовательности $a = 30$, а $K = 7$, а для верхней последовательности центр $A = 42$ и $k = 5$. В явном виде сгенерированные последовательности таковы:

$$3 + \ldots + 27 + 29 + (30) + 31 + 33 + \ldots + 57$$
$$= 840$$
$$= 59 + 61 + \ldots + 67 + 69 + (70) + 71 + 73 + \ldots + 79 + 81$$

и

$$47 + \ldots + 57 + 59 + (60) + 61 + 63 + \ldots + 73$$
$$= 840$$
$$= 75 + 77 + \ldots + 81 + 83 + (84) + 85 + 87 + \ldots + 91 + 93.$$

В обозначениях с сигмой эти равенства принимают вид:

$$\sum_{r=1}^{29}(2r-1) - \sum_{r=1}^{1}(2r-1) = 840 = \sum_{r=1}^{41}(2r-1) - \sum_{r=1}^{29}(2r-1)$$
$$\rightarrow 29^2 - 1^2 = 840 = 41^2 - 29^2,$$

а это означает, что $29^2 - 840 = 1^2$ и $29^2 + 840 = 41^2$, и мы имеем арифметическую прогрессию $\{1^2, 29^2, 41^2\}$ с разностью 840

и

$$\sum_{r=1}^{37}(2r-1) - \sum_{r=1}^{23}(2r-1) = 840 = \sum_{r=1}^{47}(2r-1) - \sum_{r=1}^{37}(2r-1)$$
$$\rightarrow 37^2 - 23^2 = 840 = 47^2 - 37^2,$$

а это означает, что $37^2 - 840 = 23^2$ и $37^2 + 840 = 47^2$, и мы имеем арифметическую прогрессию $\{23^2, 37^2, 4^2\}$ также с разностью 840, которую мы уже упоминали на стр. 183.

2. Рассмотрим более простой случай с разностью 96, которая тоже имеет вид $pq(p + q)(p - q)$, но $p = 4$, $q = 2$; в этом случае $4aK = 4Ak = 96 \Rightarrow aK = Ak = 24$. Значит, нам нужно рассмотреть множители, снова интерпретируя в зависи-

мости от контекста, какая буква заглавная, а какая – строчная. Результат показан в табл. G.2.

Таблица G.2. Разность арифметической прогрессии 96

a	24	12	8	6
k	1	2	3	4
$a + k$	25	14	11	**10**
$a - k$	23	**10**	5	2

Единственный случай, когда число в третьей строке повторяется также в четвертой, – когда центр нижней последовательности $a = 6$ и $K = 4$, а центр верхней последовательности $A = 12$ и $k = 2$. В явном виде сгенерированная последовательность такова:

$$5 + \ldots + 9 + 11 + (12) + 13 + 15 + \ldots + 19$$
$$= 96 + 21 + 23 + (24) + 25 + 27.$$

В обозначениях с сигмой это равенство принимает вид:

$$\sum_{r=1}^{10}(2r-1) - \sum_{r=1}^{2}(2r-1) = 96 = \sum_{r=1}^{14}(2r-1) - \sum_{r=1}^{10}(2r-1),$$

а это означает, что $10^2 - 2^2 = 96 = 1^2$ и $14^2 - 10^2$, и мы имеем арифметическую прогрессию $\{2^2, 10^2, 14^2\}$ с разностью 96. Мы могли бы, впрочем, записать эту разность в виде 6×4^2, тогда имели бы

$$10^2 - 2^2 = 6 \times 4^2 = 14^2 - 10^2 \rightarrow (\tfrac{10}{4})^2 - (\tfrac{2}{4})^2 = 6 = (\tfrac{14}{4})^2 - (\tfrac{10}{4})^2.$$

И мы получили трехчленную арифметическую прогрессию квадратов рациональных чисел:

$$\{(\tfrac{1}{2})^2, (\tfrac{5}{2})^2, (\tfrac{7}{2})^2\}$$

с разностью 6.

3. Фибоначчи было предложено рассмотреть случай прогрессии с разностью 5, поэтому сумма должна иметь вид $5 \times 24 \times N$, где $24N$ – полный квадрат. В простейшем случае $N = 6$, и мы должны довести сумму до числа 720, которое имеет вид $4pq(p + q)(p - q)$, где $p = 5$, $q = 4$. Читатель может проверить, что $a = 36$, $K = 5$ и $A = 45$, $k = 4$ дают одну и ту же комбинацию, и мы приходим к равенству

$$\sum_{r=1}^{41}(2r-1) - \sum_{r=1}^{31}(2r-1) = 720 = \sum_{r=1}^{49}(2r-1) - \sum_{r=1}^{41}(2r-1),$$

а это и есть решение Фибоначчи.

Наконец, что до конгруэнтности числа 7: процесс порождает

$$227 + \ldots + 447 + 449 + (450) + 451 + 453 + \ldots + 673$$

$$= 100, 800$$

$$= 675 + \ldots + 797 + 799 + (800) + 801 + 803 + \ldots + 925.$$

А это приводит к последовательности:

$$\left\{\left(\frac{113}{120}\right)^2, \left(\frac{337}{120}\right)^2, \left(\frac{463}{120}\right)^2\right\}.$$

Литература

Abdulle, A. and Wanner, G. 2002 200 Years of least squares method. *Elementary Mathematics* 57: 45–60.

Ames, J. S. 1902 *The Astrophysical J.* XV(5): 299–301.

Ampère, A.-M. 1806 Elaboration of certain issues in differential calculus which enable a new demonstration of Taylor expansions and expressions thereof in closed form if the summation is limited. *Journal de L'école Polytechnique* VI: 148–81.

Ash, A. and Gross, R. 2014 *Elliptic Tales: Curves, Counting, and Number Theory*. Princeton University Press.

Bardet, M. and Bayen, T. 2013 On the degree of the polynomial defining a planar algebraic curve of constant width, 16 December, arXiv: 1312.4358v1.

Bardet, M. and Bayen, T. 2018 On the degree of the polynomial defining planar algebraic curves of constant width, 8 February, arXiv: 1312.4358v1.

Barnett, J. H. 2004 Enter stage center: the early drama of the hyperbolic functions. *Math. Mag.* 77 (1): 15–30.

Bellhops, D. R. and Genest, C. 2007 Maty's biography of Abraham de Moivre (translated, annotated and augmented), 29 August, arXiv: 0708.3965v1.

Bernoulli, J. 1690 *Acta Eruditorum Leipzig*, pp. 217–19.

Bernoulli, J. 1694 The curvature of an elastic band. *Acta Eruditorum*, pp. 262–76.

Bernoulli, J. 2004 Lecture on the integral calculus, transl. W. A. Ferguson Jr. *21st Century Sci. Technol.* 17 (1): 34–42.

Bézier, P. 1990 Interview in *Science et Vie Micro*, February.

Block, P., DeJong, M. and Ochsendorf, J. 2006 As hangs the flexible line: equilibrium of masonry arches. *Nexus Network J.* 8 (2): 9–19.

Boas, M. 1962 *The Scientific Renaissance: 1450–1630*. Harper.

Bos, H. J. M. 1986 The concept of construction and representation of curves in seventeenth century mathematics. *Proc. Int. Congress of Math*, Berkeley.

Bos, H. J. M. 1996 Johann Bernoulli on exponential curves…implicit functions. *Vierde serie Deel* 14 (1): 1–19.

Boyer, C. B. and Merzbach, U. C. 2011 *A History of Mathematics*, 3rd edn. Wiley.

Bradley, R. E. and Sandifer, C. E. (eds) 2007 *Leonhard Euler: Life, Work and Legacy*. Elsevier.

Brown, E. 2000 Three Fermat trails to elliptic curves. *College Math. J.* 31 (3): 162–72.

Brown, E. and Myers, B. T. 2002 Elliptic curves from Mordell to Diophantus and back. *Am. Math. Monthly* 109 (7): 639–49.

Bryant, J. and Sangwin, C. 2011 *How Round Is Your Circle? Where Engineering and Mathematics Meet*. Princeton University Press.

Bukowski, J. F. 2006 *Huygens, Holland and Hanging Chains*. Bookend Seminar.

Bukowski, J. F. 2008 Christiaan Huygens and the problem of the hanging chain. *College Math. J.* 39 (1): 2–11.

Burn, B. 2000 Gregory St. Vincent and the rectangular hyperbola. *Math. Gazette* 84 (501): 480–85.

Butz, A. R. 1969 Convergence with Hilbert's space filling curve. *J. Comput. Syst. Sci.* 3: 128–46.

Cantor, G. 1878 Ein Beitrag zur Mannigfaltigkeitslehre. *Crelle's J.* 84: 242–58.

Cajori, F. 1913 History of the exponential and logarithmic concepts. *Am. Math. Monthly* 20 (1): 5–14.

Carpenter, F. B. 1995 *The Inner Life of Abraham Lincoln*. University of Nebraska Press.

Casteljau, P. de Faget de 1999 De Casteljau's autobiography: my time at Citroën. *Computer Aided Geometric Design* 16: 583–86.

Chatterjee, N. and Nita, B. G. 2010 The hanging cable problem for practical applications. *Atlantic Electron. J. Math.* 4 (1): 70–77.

Coolidge, J. L. 1963 *The Mathematics of Great Amateurs*. Dover.

Cook, T. A. 1979 *Curves of Life*. Dover.

Cotes, J. H. 2005 Congruent number problem. *Pure Appl. Math. Q.* 1 (1): 14–27.

Cundy, H. M. and Rollett, A. P. 1961 *Mathematical Models*. Oxford University Press.

Dafner, R., Cohen-Or, D. and Matias, Y. 2000 Context-based space-filling curves. *Eurographics* 19 (3): 209–18.

Dauben, J. W. 1975 The importance of dimension: problems in the early development of set theory. *Historia Mathematica* 2: 273–88.

Davis, P. J. 2001 *Spirals from Theodorus to Chaos*. A. K. Peters.

Davis, P. J. 2014 *Interpolation and Approximation*. Dover.

de Moivre, A. 1730 *Miscellanea Analytica de Seribus et Quadraturis*. London.

de Moivre, A. 1756 *The Doctrine of Chances*, reprinted for A. Millar, London.

Diaconis, P. and Zabell, S. 1991 Closed form summation for classical distributions: variations on a theme of de Moivre. *Statist. Sci.* 6 (3): 284–302.

Dickson, L. E. 1920 *History of the Theory of Numbers*, vol. II. Carnegie Institute of Washington.

Dirichlet, L. 1829 Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données. *Crelle's J.* 4: 157–69.

Dorodnov, A. W. 1947 On circular lunes quadrable with the use of ruler and compass. *Dokl. Akad. Nauk SSSR* 58: 965–68.

du Bois-Reymond, P. 1875 An attempt to classify arbitrary functions of real arguments according to their changes in the smallest intervals (Versuch einer Classification der willkürlichen Funktionen reeller Argumente nach ihren Äenderungen in den kleinsten Intervallen). *Borchardt's J.* 79: 21–37.

Duncan, D. D. 2006 *Lump: The Dog Who Ate a Picasso*, 2nd edn. Thames and Hudson.

Eknoyan, G. 2008 Adolphe Quetelet 1796–1874 – the average man and indices of obesity. *Nephrol Dial Transplant* 23: 47–51.

Euler, L. 1744 *De Curvis Elasticis*, Additamentum I. (Translated in 1933 as *A Method for Finding Curved Lines…Accepted Sense* (ed. W. A. Oldfather, C. A. Ellis and D. M. Brown). *Isis* 20 (1): 72–160.)

Euler, L. 2007 On the values of integrals extended from the variable term $x = 0$ up to x = infinity (transl.), 31 May, arXiv: 0705.4640v1.

Eves, H. 1983 *Great Moments in Mathematics: Before 1650*. Mathematical Association of America.

Eves, H. 1983 *Great Moments in Mathematics: After 1650*. Mathematical Association of America.

Eves, H. 1990 *An Introduction to the History of Mathematics*. Brook/Cole.

Fauvel, J. and Gray, J. (eds) 1987 *The History of Mathematics: A Reader*. Palgrave Macmillan.

Feynman, R. P. 1998 *What Do You Care What Other People Think? Further Adventures of a Curious Character*. W. W. Norton.

Forrest, A. R. 1972 Interactive interpolation and approximation by Bézier polynomials. *Computer J.* 15 (1): 71–79.

Forrest, A. R. 1990 Interactive interpolation and approximation by Bézier polynomials. *Computer Aided Design* 22 (9): 527–37.

Galilei, G. 1914 *Two New Sciences*. Macmillan.

Gardner, M. 1991 *The Unexpected Hanging and Other Mathematical Diversions*. University of Chicago Press. (Republished as *Knote and Borromean Rings etc.* by the MMA in 2014.)

Gerver, J. 1970 The differentiability of the Riemann function at certain rational multiples of _. *Am. J. Math.* 92: 33–55.

Gil, J. B. 2005 The catenary (almost) everywhere. *Boletìn de la Asociaciòn Matemàtica Venezolana* XII (2): 251–58.

Gillispie, C. C. 2000 *Pierre-Simon Laplace, 1749–1827: A Life in Exact Science*. Princeton University Press.

Gouvea, F. Q. 2011 Was Cantor surprised? *Am. Math. Monthly* 118: 198–209.

Gray, J. J. 2000 *The Hilbert Challenge: A Perspective on Twentieth Century Mathematics*. Oxford University Press.

Gregory, D. 1695–1697 The properties of the hanging catenaria. *Philosophical Transactions of the Royal Society of London* 19: 637–52.

Hadamard, J. 1921 L'oeuvre mathématique de Henri Poincaré. *Acta* 38: 203–87.

Hahn, A. J. 2012 *Mathematical Excursions to the World's Great Buildings*. Princeton University Press.

Hald, A. 1990 *History of Probability and Statistics and Their Applications before 1750*. Wiley.

Hardy, G. H. 1916 Weierstrass's non-differentiable function. *Trans. Am. Math. Soc.* 17 (3): 301–25.

Havil, J. 2014 *John Napier: Life, Logarithms, and Legacy*. Princeton University Press.

Heath, T. 1910 *Diophantus of Alexandria*. Cambridge University Press.

Heath, T. 1981 *A History of Greek Mathematics*, Vol. 1: *From Thales to Euclid*. Dover.

Heilberg, J. L. 2007 *Euclid's Elements of Geometry*, Greek text edited and translated by R. Fitzpatrick (https://archive.org/stream/JL_Heiberg__EUCLIDS_ELEMENTS_OF_ GEOMETRY/Elements_djvu.txt).

Herschel, J. 1850 *Edin. Rev.* XCII: 1–57.

Hermite, C. 1905 *Correspondance d'Hermite et de Stieltjes*, vol. II. Paris: Gauthier-Villars.

Heyman, J. 1982 *The Masonry Arch*. Wiley.

Hilbert, D. 1891 Über die stetige Abbildung einer Linie auf ein Flächenstück. *Math. Annln*, pp. 459–60.

Huygens, C. 1638–1656 *Oeuvres Complétes Correspondence of Christiaan Huygens* tome 1, no. 14.

Jesseph, D. M. 2000 *Squaring the Circle: The War between Hobbes and Wallis*. Chicago University Press.

Johnson, D. M. 1979 The problems of invariance of dimension in the growth of modern topology, Part I (communicated by H. Freudenthal). *Archive for History of Exact Sciences* 20 (2): 97–188.

Johnson, D. M. 1981 The problems of invariance of dimension in the growth of modern topology, Part II (communicated by H. Freudenthal). *Archive for History of Exact Sciences* 25 (2–3): 85–266.

Jones, W. 1706 *Synopsis Palmariorum Mathesios*.

Joy, K. I. 2000 Bernstein polynomials. Visualization and Graphics Research Group, Department of Computer Science, University of California, Davis.

King, B. 2009 Mapping an arbitrary message. *Int. J. Net Security* 8 (2): 169–76.

Klein, F. 1958 *On Mathematics* (ed. R. Moritz). Dover.

Kline, M. 1990 *Mathematical Thought from Ancient to Modern Times*, vols 1–3. Oxford University Press.

Kolwankar, K. M. and Gangal, A. D. 1996 Fractional differentiability of nowhere differentiable functions and dimensions, 21 November 1996 (arXiv: chaodyn/9609016v2).

Lambert, J. 1761 Mémoire sur quelques propriétés remarquable des quantitiés transcendantes circulaires et logarithmique. *Mémoires de l'Académie royale des sciences de Berlin* 17: 265–322.

Lambert, J. 1768–1770 Observations trigonométrique. *Mémoires de l'Académie royale des sciences de Berlin* 24: 327–54.

Landau, E. 2001 *Differential and Integral Calculus*. American Mathematical Society.

Lang, S. 1978 *Elliptic Curves, Diophantine Analysis*. Springer.

Laplace, P. S. 1986 Memoir on the probability of the causes of events (transl. S. M. Stigler). *Statist. Sci.* 1 (3): 364–78.

Lazarov, B. 2001 Teaching envelopes in secondary school. *The Teaching of Mathematics* XIII (1): 45–55.

Leibniz, W. W. 1691 Concerning the curve formed by a heavy flexible chord … mean proportionals and logarithms. *Actis Erudit. Lips.*, June.

Levien, R. 2009 From spiral to spline: optimal techniques in interactive curve design. PhD thesis.

Liouville, J. 1833 Mémoire sur les transcendantes elliptiques de premiere et de seconde espace considérées comme fonctions de leur amplitude. *Journal de l'Ecole Polytechnique* 23: 37–83.

Lockwood, E. H. 1961 *A Book of Curves*. Cambridge University Press.

Lopez, G. M. 1998 Poleni's manuscripts about the dome of St. Peter's. PhD thesis.

Mahoney, M. S. 1994 *The Mathematical Career of Pierre de Fermat: 1601–1665*. Princeton University Press.

Mainstone, R. 2003 Saving the dome of St. Peter's. *Construction History* 19: 3–18.

Mascheroni, L. 1785 *Nuove Ricerche Sull' Equilibrio Delle Volte*. Bergamo: Francesco Locatelli.

Mazur, J. 2010 *What's Luck Got to Do with It? The History, Mathematics, and Psychology of the Gambler's Illusion*. Princeton University Press.

Mellish, A. P. 1931 Notes on differential geometry. *Ann. Math.* 32 (1): 181–90.

Milne, J. S. 2006 *Elliptic Curves*. Booksurge.

Monge, G. 1810 Sur les équations différentielles des courbes du second degré. *Bull. Soc. Philom. Paris*, pp. 87–88.

Moore, E. H. 1900 On certain crinkly curves. *Trans. Am. Math. Soc.* 1: 72–90.

Morain, F. 2006 *Proc. LIX Colloq. École Polytechnique*, Paris, 26 November.

Newman, J. (ed.) 2003 *The World of Mathematics*, vols 1–4. Dover.

Newton, I. 1667 *Enumeratio Curvarum Trium Dimensionum* («The Enumeration of Cubics»). Mathematical Papers 12, pp. 10–89.

Ng, C. H. B. and Fan, W. Y. 2014 Reuleaux triangle disks: new shape on the block. *J. Am. Chem. Soc.* 136 (37): 12, 840–43.

O'Grady, P. 2003 Hippias of Elis. *Proc. Biennial Int. Conf. of Greek Studies*, Flinders University.

Osserman, R. 2010 How the gateway arch got its shape. *Nexus Network J.* 12: 167–89.

Osserman, R. 2010 Mathematics of the gateway arch. *Notices AMS* 57 (2): 220–29.

Pardies, I. G. 1725 *La Statique, ou la Science des Forces Mouvantes*. Paris.

Peano, G. 1890 Sur une courbe, qui rempli toute une aire plane. *Math. Annln* 36 (1): 157–60.

Poincaré, H. 1899 La Logique et l'intuition dans la science mathématique et dans l'enseignement. *L'enseignement mathématique* 1: 157–61.

Rabut, C. 2002 On Pierre Bézier's life and motivations. *Computer-Aided Design* 34: 493–510.

Ramanujan, S. 1962 *Modular Equations and Approximations to π. Ramanujan's Collected Works*. New York: Chelsea.

Reissmann, N. and Meyer, J. C. 2016 A study of energy and locality effects using space-filling curves, 20 June, arXiv: 1606.06133v1.

Resnikoff, H. L. 2015 On curves and surfaces of constant width, 25 April, arXiv: 1504.06733v1.

Reuleaux, F. 1963 *The Kinematics of Machinery*. Dover.

Rice, R. and Brown, E. 2012 Why are ellipses not elliptic curves. *Mathematics Magazine* 85: 163–76.

Robertson, S. A. 1984 Smooth curves of constant width and transnormality. *Bull. Lond. Math. Soc.* 16: 264–74.

Rouse Ball, W. W. 2003 *A Short Account of the History of Mathematics*. Dover.

Sagan, H. 1994 *Space Filling Curves*. Springer.

Sagan, H. 1991 Some reflections on the emergence of space-filling curves: the way it could have happened and should have happened, but did not happen. *J. Franklin Inst.* 328 (4): 419–30.

Scott, J. F. 1960 *A History of Mathematics; From Antiquity to the Beginning of the Nineteenth Century*. Barnes & Noble.

Server, J. 1971 More on the differentiability of the Riemann function. *Am. J. Math.* 93: 33–41.

Shoesmith, E. 1985 Thomas Simpson and the arithmetic mean. *Historia Mathematica* 12: 352–55.

Simpson, T. 1755 A letter to the Rt. Hon. George Macclesfield. On the advantage of taking the mean of a number of observations in practical astronomy. *Phil. Trans. R. Soc. Lond.* 49: 82–93.

Smith, D. E. 1985 *A Source Book in Mathematics*. Dover.

Somervell, E. L. 1906 *A Rhythmic Approach to Mathematics*. London: George Philip and Son.

Stahl, S. 2006 The evolution of the normal distribution. *Mathematics Magazine* 79 (2): 96–113.

Steef, A., Shamma, M. N. and Alkhatib, A. 2017 A secure approach for embedding message text on an elliptic curve defined over prime fields, and building «EC-RSA-ELGamal» Cryptographic System. *Int. J. Comput. Sci. Inform. Security* 15 (6).

Stigler, S. M. 1990 *The History of Statistics: The Measurement of Uncertainty before 1900*. The Belknap Press of Harvard University Press.

Stigler, S. M. 2002 *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard University Press.

Stillwell, J. 1995 Elliptic curves. *Am. Math. Monthly* 102 (9): 831–37.

Swetz, F. J. 2013 *The European Mathematical Awakening*. Dover.

Talbot, A. N. 1890–91 *The Railway Transition Spiral*, vol. 5, p. 96. University of Illinois Technograph.

Teets, D. and Whitehead, K. 1999 The discovery of Ceres: how Gauss became famous. *Mathematics Magazine* 72 (2): 83–93.

Thomas, I. 2018 *Selections Illustrating the History of Greek Mathematics. I. From Thales to Euclid*. Harvard University Press.

Timmerman, G. 2014 Approximating continuous functions and curves using Bernstein polynomials. *Math* 336, 2 June.

Todhunter, I. 1865 *History of the Theory of Probability from the Time of Pascal to that of Laplace*. Macmillan.

Tupper, J. 2001 Reliable two-dimensional graphing methods for mathematical formulae with two free variables. *Proc. 28th Ann. Conf. Computer Graphics and Interactive Techniques*, pp. 77–86. ACM Press.

Vilenkin, N. Ya. 1995 *In Search of Infinity*. Birkhäuser.

Villain, M. B. 2008 Ramanujan's perimeter of an ellipse, 1 February, arXiv. math/0506384v1.

Waller, R. 1705 *The Posthumous Works of Dr. Robert Hooke*. The Royal Society.

Washington, L. C. 2008 *Elliptic Curves Number Theory and Cryptography*. Chapman and Hall.

Watson, G. N. 1918 The problem of the square pyramid. *Messenger of Mathematics* 48: 1–22.

Wells Jr, R. O. 2015 Geometry in the age of Enlightenment, 30 June, arXiv: 1507.00060v1.

Yaglom, I. M. and Boltyanskii, V. G. 1961 *Convex Figures*, transl. P. J. Kelly and L. F. Walto.

Предметный указатель

Книги издательства «ДМК Пресс» можно купить оптом и в розницу
на складе издательства по адресу:
Москва, ул. Электродная, д. 2, стр. 12, офис 7,
тел. **+7 (499) 322–19–38**,
а также заказать на сайте **www.dmkpress.com**
с доставкой в любой регион РФ.

Джулиан Хэйвил

Замечательные математические кривые

Антология непредсказуемого, исторического,
чарующего и романтического

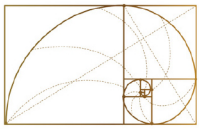
Главный редактор *Мовчан Д. А.*
Зам. главного редактора *Яценков В. С.*
editor@dmkpress.com
Перевод *Слинкин А. А.*
Корректор *Абросимова Л. А.*
Верстка *Луценко С. В.*
Дизайн обложки *Мовчан А. Г.*

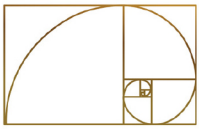
Формат 70×100 1/16.
Гарнитура «PT Serif». Печать цифровая.
Усл. печ. л. 19,83. Тираж 200 экз.

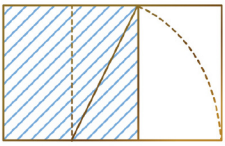
Веб-сайт издательства: *www.dmkpress.com*

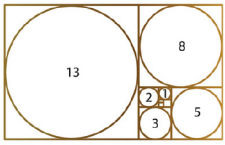
«Что такое кривая? Всякий знает, что такое кривая, пока не выучится математике настолько, что вконец запутается в бесконечных исключениях».

Феликс Клейн (1958)









В этой книге собраны описания десяти математических кривых, тщательно отобранных за их значимость, интересность и красоту. В каждой главе читатель найдет историю и определение кривой, а также узнает о красивой и часто неожиданной математической основе, связанной с ее созданием и эволюцией. Автор знакомит нас с математиками и другими первопроходцами научной мысли – слава одних пережила века, имена других преданы забвению. Книга построена так, что все желающие могут превратиться в исследователей, просто вооружившись карандашом и бумагой.

Основные темы книги:

- знакомство с Пьером Безье, Дольфом Кетле и другими мыслителями, имена которых носят кривые;
- описание парадоксов, проблем и озарений;
- категории кривых;
- рассказы об удивительных свойствах различных кривых.

Издание адресовано широкому кругу любителей математики и может быть полезно преподавателям и руководителям математических кружков.

У каждой кривой есть история, достойная рассказа!

ISBN 978-5-93700-241-9

9 785937 002419 >

ДМК издательство

www.дмк.рф