

Летняя школа «Современная математика»

А. А. Кириллов

Повесть о двух фракталах

Издание второе, исправленное

Москва
Издательство МЦНМО
2010

УДК 517.938
ББК 22.151
К43

Проведение летних школ «Современная математика» и издание ее материалов поддержано Московской городской Думой и Департаментом образования г. Москвы, а также фондом «Династия», фирмой «НИКС» и корпорацией Boeing.

Кириллов А. А.

К43 Повесть о двух фракталах. — 2-е изд., исправленное. — М.: МЦНМО, 2010. — 180 с.

ISBN 978-5-94057-670-9

Эта брошюра, написанная по материалам лекций, прочитанных автором для школьников и студентов на летней школе «Современная математика», представляет собой введение в теорию фракталов — новый, актуальный раздел математики. Начинаясь с основных определений, книга доходит до свежих результатов и нерешенных проблем.

Для студентов младших курсов и школьников старших классов.

ББК 22.151

ISBN 978-5-94057-670-9

© А. А. Кириллов, 2009,
2010 (исправления).

© МЦНМО, 2009.

Посвящается Бену (Коле) и Лизе

ВВЕДЕНИЕ

Эта книга посвящена обсуждению фрактальных множеств, или просто *фракталов*. Такие множества известны уже больше ста лет и появлялись в разных областях науки. Но только недавно (около 30 лет тому назад) они стали предметом математического исследования.

Пионером теории фракталов был Б. Мандельброт. Его книга [Man] впервые появилась в 1977 году, а второе, расширенное издание вышло в 1982 году. После этого серьезные работы, обзоры, популярные статьи и книги о фракталах стали появляться десятками, если не сотнями. С 1993 года в издательстве World Scientific выходит специальный периодический журнал «Фракталы». Так зачем писать еще одну книгу о фракталах?

Во-первых, несмотря на обширную литературу, многие люди, включая студентов, аспирантов и значительную часть работающих математиков, имеют довольно смутное представление о фракталах.

Во-вторых, во многих популярных книгах читатель увидит массу цветных картинок и любопытных примеров, но не найдет ни точных определений, ни строго доказанных результатов. С другой стороны, работы профессиональных математиков, как правило, слишком трудны для начинающих. Они обычно посвящены довольно специальным вопросам и часто предполагают заранее известными все связи и мотивировки.

Последняя и, может быть, самая важная причина состоит в том, что самостоятельное изучение геометрии, анализа и арифметики фракталов, на мой взгляд, является одним из лучших способов для молодого математика активно и прочно овладеть основными математическими знаниями.

Мне кажется также, что это — прекрасная возможность проверить свою способность к творческой работе в математике¹. Я имею в виду не только решение точно сформулированных задач, но и распознавание скрытых закономерностей и постановку новых плодотворных вопросов.

Мой личный интерес к фракталам возник, когда я читал специальный курс о фракталах в 1995 году по просьбе нескольких студентов разных специальностей. Я повторял этот курс в 1999, 2003 и 2005 г.

1 По определению Ю. И. Манина, творить в математике — это вычислять, волнуясь.

В 2004 и в 2007 году я имел возможность изложить часть этого материала в лекциях для участников летней математической школы в Дубне под Москвой, организованной для старшеклассников и младшекурсников. Я был приятно удивлен активностью аудитории и тем, как быстро слушатели воспринимали новую для них информацию.

В этой книге я намеренно ограничиваюсь только двумя примерами фрактальных множеств: коврами Серпинского и Аполлония. Мы рассматриваем и точно формулируем серию задач, возникающих при изучении этих фракталов. Большинство из них можно ставить и решать независимо от остальных, но только вся их совокупность дает реальное представление о мире фракталов.

Некоторые из этих задач являются просто упражнениями на понимание терминов и логики изложения, другие представляют сравнительно недавние результаты, а несколько наиболее интересных являются нерешенными проблемами неизвестной степени трудности. Решение (и даже понимание формулировки) этих задач требует некоторых предварительных знаний. В частности, мы предполагаем известными:

- элементы анализа: функции одной вещественной переменной, дифференциальное и интегральное исчисление, числовые ряды и ряды функций;
- элементы линейной алгебры: вещественные и комплексные линейные пространства, размерность, линейные операторы, квадратичные формы, собственные значения и собственные векторы, координаты и скалярное произведение;
- элементы геометрии: прямые линии, плоскости, окружности, круги и сферы в \mathbb{R}^3 , основные тригонометрические формулы, начала сферической и гиперболической геометрии;
- элементы арифметики: простые числа, взаимно простые числа, НОД (наибольший общий делитель), рациональные числа, понятие об алгебраических числах;
- элементы теории групп: подгруппы, нормальные подгруппы, однородные пространства, классы смежности, матричные группы.

Все это обычно входит в программу первых двух или трех лет университета. Разнообразие этих сведений и их взаимосвязь я рассматриваю как большое преимущество теории фракталов и как характерную черту современной математики.

Несколько слов о стиле изложения. Я старался избежать двух главных опасностей: сделать книгу скучной, объясняя слишком подробно

простые детали, и сделать ее непонятной, используя наиболее эффективную современную технику, которая порой слишком абстрактна. Читателю судить, насколько это мне удалось.

Я также старался довести до читателей неформальное понимание математических методов, которое отличает (почти любого) профессионала от начинающего любителя. Иногда одна фраза объясняет больше, чем длинная статья или толстая книга. В моей практике это случилось, когда я пытался понять, что такое индуцирование в теории представлений, спектральная последовательность в алгебраической топологии, язык схем в алгебраической геометрии. Поэтому я иногда использую «высоконаучные» термины и понятия, объясняя всякий раз, что они значат, если «отбросить незабудки»¹.

Дополнительная информация включена в текст в виде кратких вставок, именуемых «схолиями»². Конец каждой схолии отмечен знаком \diamond .

Иногда я также привожу дополнительную информацию в замечаниях. Конец замечания отмечается знаком \heartsuit .

Конец доказательства или его отсутствие отмечено знаком \square .

Автор благодарен Институту Эрвина Шрёдингера (ESI) в Вене, где эта работа была начата, Институту Макса Планка (MPI) в Бонне и Институту высших научных исследований (IHES) во Франции, где она была завершена.

Я также благодарен моим студентам и аспирантам, настоящим и бывшим, за многочисленные замечания и \TeX ническую помощь.

Department of Mathematics, The University of Pennsylvania, Philadelphia, PA 19104–6395, USA

Институт проблем передачи информации РАН, Б. Каретный, д. 19, Москва 101477, ГСП–4, Россия.

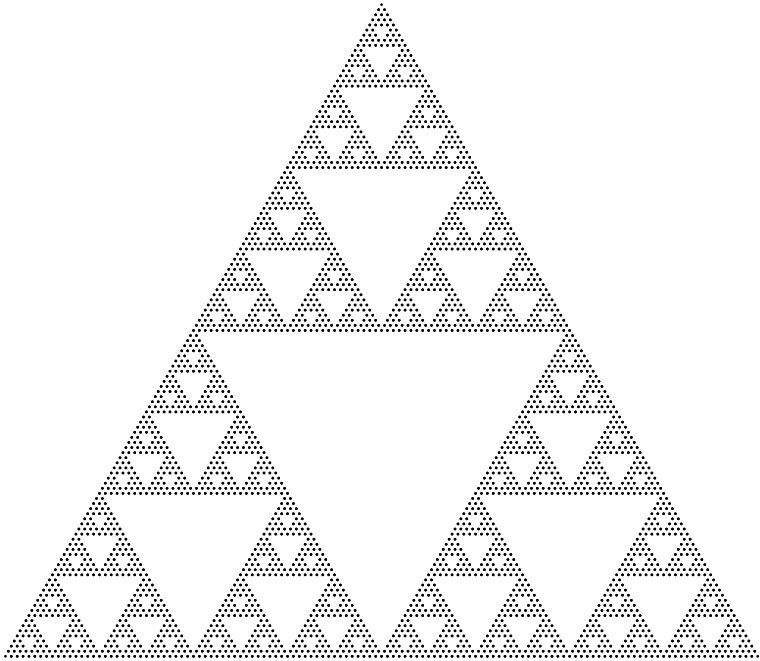
E-mail address: kirillov@math.upenn.edu

- 1 См. Козьма Прутков «Незабудки и запятки» (басня), а также «Избранные анекдоты и притчи семинара И. М. Гельфанда» (планируемая статья).
- 2 Что такое схолия, читатель может узнать, прочитав книгу С. Боброва «Волшебный двурог» (см. ссылку на с. 152) или спросив у своих знакомых.

Часть I



КОВЕР СЕРПИНСКОГО



Теперь давайте заменим эти числа их вычетами по модулю 2. Другими словами, поставим вместо каждого четного числа 0, а вместо нечетного числа 1. Мы получим следующую таблицу:

				1				
				1	1			
			1	0	1			
		1	1	1	1			
		1	0	0	0	1		
	1	1	0	0	1	1		
	1	0	1	0	1	0	1	
1	1	1	1	1	1	1	1	1
.....								

Как можно описать эту картину? Заметим, что весь наш треугольник со стороной 8 содержит три одинаковых треугольника со стороной 4 (левый, правый и верхний); каждый из них содержит три одинаковых треугольника со стороной 2, состоящих из трех единиц. Все остальные места заняты нулями.

Попробуем вообразить, что получится, если мы продолжим наш треугольник до 2^N -го ряда, где N — большое число. Если мы сожмем наш треугольник до размера книжной страницы и заменим единицы черными точками, а нули — белыми, то мы получим такую картину:

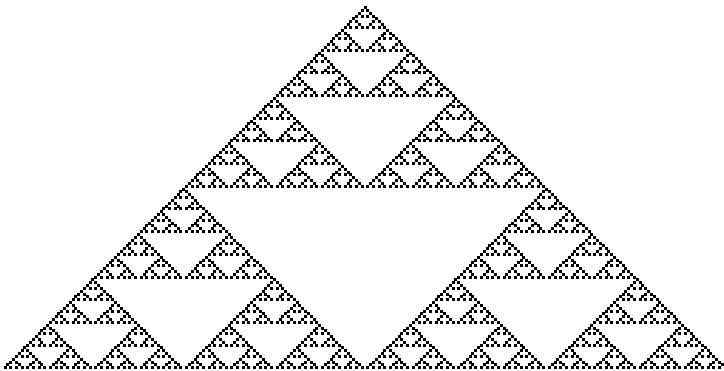


Рис. 1.1. Треугольник Паскаля mod 2

Здесь целый треугольник состоит из трех треугольников половинного размера, которые выглядят подобно целой картине. Пространство, ограниченное этими треугольниками, заполнено белыми точками.

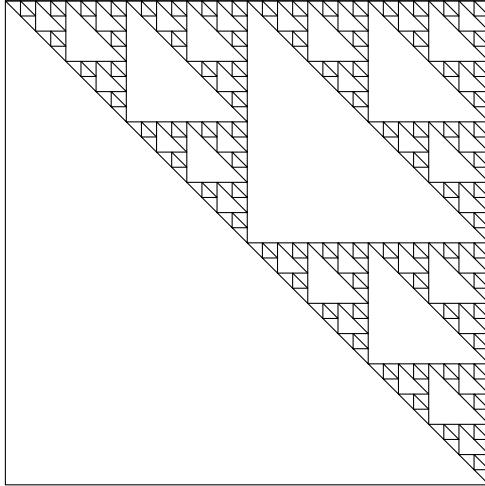


Рис. 1.2. Треугольная матрица Паскаля

Довольно ясно, что когда N стремится к бесконечности, наша картина стремится к некоторому пределу¹. Этот предел — так называемый ковер Серпинского, открытый в 1916 году польским математиком Вацлавом Серпинским.

Другой пример появления того же множества связан со следующей задачей линейной алгебры. Пусть E_N — матрица размером $N \times N$ с элементами из простейшего конечного поля $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$, заданная условиями:

$$(E_N)_{ij} = \begin{cases} 1, & \text{если } i < j, \\ 0 & \text{в противном случае.} \end{cases}$$

Согласно общей теории, эта матрица приводится к жордановой нормальной форме, представляющей собой один блок J_N , где

$$(J_N)_{ij} = \begin{cases} 1, & \text{если } j = i + 1, \\ 0 & \text{в противном случае.} \end{cases}$$

Попробуем найти матрицу A_N , которая устанавливает эквивалентность $E_N = A_N J_N A_N^{-1}$. Оказывается, что такая матрица может быть выбрана в виде, показанном на рис. 1.2.

¹ См. ниже схиолю А по поводу строгого определения предела в этой ситуации.

Мы оставляем читателю объяснить этот факт и найти связь между треугольником Паскаля и матрицей A_N .

Прежде чем идти дальше, мы должны обобщить понятие предела — основное понятие анализа — так, чтобы оно было применимо не только к числовым последовательностям, но к последовательностям объектов другой природы. В частности, мы хотим придать точный смысл выражению «последовательность множеств $\{X_n\}$ стремится к предельному множеству X ».

Соответствующий раздел математики называется теорией метрических пространств. Используя эту теорию, мы можем определить фракталы (которые являются довольно сложными множествами) как пределы последовательностей более простых множеств.

Схолия А. МЕТРИЧЕСКИЕ ПРОСТРАНСТВА

Мы начнем с довольно общих и абстрактных определений, которые позже будут проиллюстрированы и объяснены на многих конкретных примерах. Возможно, некоторым читателям наше изложение покажется слишком абстрактным и трудным для понимания и запоминания. Но вы вскоре убедитесь, что новые понятия очень полезны во многих случаях. Они позволяют рассматривать с единой точки зрения много задач, которые выглядят совершенно различно.

А.1. Определение метрического пространства

ОПРЕДЕЛЕНИЕ А.1. Метрическое пространство — это пара (M, d) , где M — множество, а $d: M \times M \rightarrow \mathbb{R}$ — функция, которая каждой паре точек x и y из M ставит в соответствие число $d(x, y)$ — *расстояние* между x и y . При этом требуется, чтобы следующие аксиомы были выполнены:

- **положительность:** для всех $x, y \in M$ величина $d(x, y)$ — неотрицательное вещественное число, которое равно нулю тогда и только тогда, когда $x = y$;
- **симметричность:** $d(x, y) = d(y, x)$ для всех $x, y \in M$;
- **неравенство треугольника:** $d(x, y) \leq d(x, z) + d(z, y)$ для всех $x, y, z \in M$.

Модельными примерами метрических пространств являются:

1. Вещественная прямая (\mathbb{R}, d) , где расстояние определено обычной формулой

$$d(x, y) = |x - y|. \quad (1.1.1)$$

2. Плоскость (\mathbb{R}^2, d) с обычным расстоянием между $x = (x_1, x_2)$ и $y = (y_1, y_2)$:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}. \quad (1.1.2)$$

3. Трехмерное пространство с обычным расстоянием

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}. \quad (1.1.3)$$

ОПРЕДЕЛЕНИЕ А.2. Мы говорим, что последовательность $\{x_n\}$ точек M сходится к точке $a \in M$, или имеет предел a , если $d(x_n, a) \rightarrow 0$, когда $n \rightarrow \infty$.

ОПРЕДЕЛЕНИЕ А.3. Последовательность $\{x_n\}$ называется *фундаментальной*, или *последовательностью Коши*, если она обладает свойством:

$$\lim_{m, n \rightarrow \infty} d(x_m, x_n) = 0. \quad (1.1.4)$$

Легко показать (попробуйте сами), что любая сходящаяся последовательность является фундаментальной. Обратное не всегда верно. Например, если наше метрическое пространство — это луч $\mathbb{R}_{>0}$, состоящий из всех положительных чисел, с обычным расстоянием (А.1), то последовательность $x_n = \frac{1}{n}$ фундаментальна, но не имеет предела.

ОПРЕДЕЛЕНИЕ А.4. Метрическое пространство (M, d) называется *полным*, если каждая фундаментальная последовательность в нем имеет предел.

В нашей книге мы рассматриваем, как правило, полные метрические пространства. Например, пространства (А.1–3) полны согласно известной теореме анализа.

ОПРЕДЕЛЕНИЕ А.5. Подмножество X в метрическом пространстве (M, d) называется *замкнутым в M* , если оно содержит все свои предельные точки (то есть пределы последовательностей $\{x_n\} \subset X$).

УПРАЖНЕНИЕ 1. Пусть (M, d) — полное метрическое пространство и X — подмножество в M . Тогда (X, d) само является метрическим пространством.

Докажите, что это пространство полно тогда и только тогда, когда X замкнуто в M .

Подсказка. Это — просто упражнение на знание и понимание определений. Сформулируйте аккуратно, что дано и что требуется доказать, и вы получите доказательство.

ПРЕДУПРЕЖДЕНИЕ. Если это упражнение кажется вам трудным, попробуйте еще раз и посоветуйтесь со своим преподавателем.

Следующие определение и упражнение полезны во многих случаях.

ОПРЕДЕЛЕНИЕ А.6. Говорят, что последовательность $\{x_n\}$ в метрическом пространстве (M, d) имеет конечную длину, если ряд

$$\sum_{n=1}^{\infty} d(x_n, x_{n+1})$$

сходится.

УПРАЖНЕНИЕ 2. Докажите, что

- а) каждая последовательность конечной длины фундаментальна;
- б) каждая фундаментальная последовательность содержит подпоследовательность конечной длины.

А.2. Сжимающие отображения

ОПРЕДЕЛЕНИЕ А.7. Отображение f из метрического пространства (M, d) в себя называется *сжимающим*, если для некоторого числа $\lambda \in (0, 1)$ выполнено условие

$$d(f(x), f(y)) \leq \lambda \cdot d(x, y) \quad \text{для всех } x, y \in M. \quad (1.1.5)$$

Нам понадобится следующая теорема.

ТЕОРЕМА О СЖИМАЮЩИХ ОТОБРАЖЕНИЯХ. Пусть M — полное метрическое пространство и f — сжимающее отображение M в себя. Тогда в M существует единственная неподвижная точка для отображения f , то есть такая точка, что $f(x) = x$.

Доказательство этой теоремы очень коротко и поучительно. Кроме того, оно дает простой метод для нахождения неподвижной точки. Поэтому мы приведем доказательство здесь.

Доказательство. Пусть x_0 — произвольная точка из M . Рассмотрим последовательность $\{x_n\}_{n \geq 0}$, определяемую индуктивно формулой $x_n = f(x_{n-1})$ для $n \geq 1$.

Оказывается, эта последовательность всегда сходится. А именно, мы покажем, что она является последовательностью Коши. В самом деле, пусть $d(x_0, x_1) = d$. Тогда из (1.1.5) мы заключаем, что

$$d(x_1, x_2) \leq \lambda \cdot d, \quad d(x_2, x_3) \leq \lambda^2 \cdot d, \quad \dots, \quad d(x_n, x_{n+1}) \leq \lambda^n \cdot d.$$

Поэтому для любых $m < n$ мы имеем $d(x_m, x_n) \leq \sum_{k=m}^{n-1} \lambda^k \cdot d \leq \frac{\lambda^m}{1-\lambda} \cdot d$. Значит,

$$\lim_{m, n \rightarrow \infty} d(x_m, x_n) \rightarrow 0$$

и фундаментальность $\{x_n\}$ доказана.

Поскольку M полно, наша последовательность Коши имеет предел, который мы обозначим x_∞ . Далее, функция f , как всякое сжимающее отображение, непрерывна.

Поэтому $f(x_\infty) = \lim_{n \rightarrow \infty} f(x_n) = \lim_{n \rightarrow \infty} x_{n+1} = x_\infty$, то есть x_∞ является неподвижной точкой.

Наконец, если бы существовали две неподвижные точки x и y , то было бы справедливо неравенство $d(x, y) = d(f(x), f(y)) \leq \lambda \cdot d(x, y)$. Но это возможно лишь при $d(x, y) = 0$, следовательно, $x = y$. \square

Доказанная теорема, решает, в частности, следующую шуточную задачу, предлагавшуюся на некоторых математических олимпиадах для школьников.

Задача 1. Мальчик Петя вышел из своего дома и пошел в школу. На полпути к школе он решил прогулять школу и пойти на каток. На полпути к катку он подумал, что лучше пойти в кино. Однако на полпути к кинотеатру он снова передумал и свернул к школе. Куда придет мальчик Петя, если он будет продолжать двигаться таким образом?

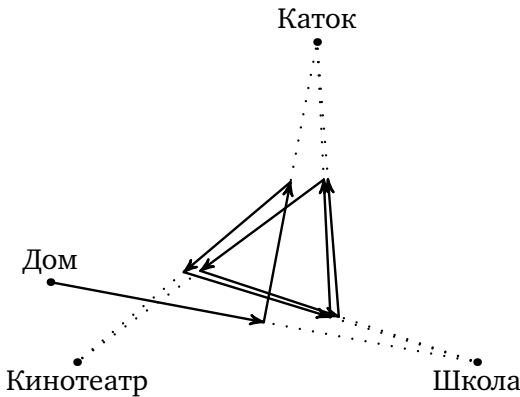


Рис. 1.3. Ленивый Петя

А.3. Компактные пространства

ОПРЕДЕЛЕНИЕ А.8. Метрическое пространство (M, d) называется *компактным*, если каждая последовательность $\{x_n\}$ точек в M содержит сходящуюся подпоследовательность.

ОПРЕДЕЛЕНИЕ А.9. Подмножество $S \subseteq M$ называется ε -сетью в M , если для любой точки $t \in M$ найдется точка $s \in S$, для которой $d(t, s) < \varepsilon$.

ТЕОРЕМА (об ε -сети). *Метрическое пространство (M, d) компактно тогда и только тогда, когда оно полно и для каждого $\varepsilon > 0$ в M существует конечная ε -сеть.*

УПРАЖНЕНИЕ 3. Докажите, что подмножество X в \mathbb{R} , в \mathbb{R}^2 или в \mathbb{R}^3 компактно тогда и только тогда, когда оно замкнуто и ограничено.

Подсказка. Если подмножество X не замкнуто или не ограничено, легко построить последовательность $\{x_n\}$ точек в X , которая не содержит сходящихся подпоследовательностей.

Если X ограничено, то оно содержится в каком-нибудь отрезке (квадрате, кубе) размера R для достаточно большого R . Из теоремы об ε -сети можно вывести, что каждое замкнутое подмножество отрезка (квадрата, куба) компактно. \diamond

1.2. САМОПОДОВНЫЕ ФРАКТАЛЫ

Теперь мы можем ввести главное техническое средство для работы с фрактальными множествами.

Пусть M — метрическое пространство. Обозначим через $\mathbb{K}(M)$ совокупность всех непустых компактных подмножеств в M . Мы хотим определить расстояние между двумя подмножествами так, чтобы $\mathbb{K}(M)$ было также метрическим пространством. Для этого мы сначала определим¹ расстояние $d(x, Y)$ между точкой x и компактным непустым множеством Y :

$$d(x, Y) := \min_{y \in Y} d(x, y). \quad (1.2.1)$$

Расстояние между двумя компактными непустыми множествами X и Y определяется формулой

$$d(X, Y) := \max_{x \in X} d(x, Y) + \max_{y \in Y} d(y, X). \quad (1.2.2)$$

Более прямое определение, не использующее промежуточных понятий, выглядит более громоздко:

$$d(X, Y) := \max_{x \in X} \min_{y \in Y} d(x, y) + \max_{y \in Y} \min_{x \in X} d(x, y) \quad (1.2.3)$$

Однако если подумать немного, как определить расстояние между двумя подмножествами так, чтобы выполнялись все три аксиомы, вы увидите, что (1.2.2), или (1.2.3), является простейшим выбором.

¹ Знак $:=$, который мы используем здесь, читается «равно по определению». Он означает, что правая часть уравнения является определением левой части. Аналогичный смысл имеет знак $=:$.

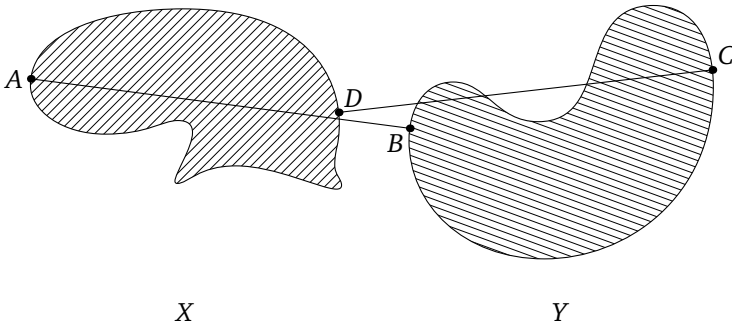


Рис. 1.4. Хаусдорфово расстояние

На рисунке 1.4 первое и второе слагаемые в (1.2.3) — это длины отрезков AB и CD соответственно.

УПРАЖНЕНИЕ 4. Докажите, что минимум в (1.2.1) и максимум в (1.2.2) действительно достигаются.

Подсказка. Используйте компактность X и Y .

УПРАЖНЕНИЕ 5. Подсчитайте расстояния: а) между границей квадрата со стороной 1 и его диагональю; б) между единичной окружностью и единичным кругом, ограниченным этой окружностью.

Ответ. а) $\frac{1+\sqrt{2}}{2}$; б) 1.

ТЕОРЕМА 1.1. Если метрическое пространство M полно (соответственно компактно), то пространство $\mathbb{K}(M)$ тоже полно (соответственно компактно).

Подсказка. Пусть $\{X_n\}$ — последовательность компактных непустых подмножеств в M , которая образует фундаментальную последовательность точек в $\mathbb{K}(M)$. Рассмотрим совокупность X таких точек $x \in M$, для которых существует последовательность $\{x_n\}$, обладающая свойствами: $x_n \in X_n$ и $\lim_{n \rightarrow \infty} x_n = x$.

Докажите последовательно, что X непусто, компактно и является пределом последовательности $\{X_n\}$ в $\mathbb{K}(M)$.

В доказательстве первого и третьего утверждений полезна задача 2 б), а в доказательстве второго — теорема об ε -сети.

Предположим, что задано семейство сжимающих отображений $\{f_1, f_2, \dots, f_k\}$ из M в себя. Определим отображение $F: \mathbb{K}(M) \rightarrow \mathbb{K}(M)$ с помощью формулы

$$F(X) = f_1(X) \cup f_2(X) \cup \dots \cup f_k(X). \quad (1.2.4)$$

ТЕОРЕМА 1.2. *Отображение F — сжимающее, и, следовательно, существует единственное непустое компактное множество $X \subset M$, обладающее свойством $F(X) = X$.* \square

ОПРЕДЕЛЕНИЕ 1.10. Множество X из теоремы 1.2 называется *однородным самоподобным фрактальным множеством* или, короче, *самоподобным фракталом*. Семейство функций f_1, \dots, f_k обычно называют *порождающей системой функций*, определяющей фрактал X .

Иногда используется более общее определение. А именно, вместо (1.2.4) зададим отображение F формулой

$$F(X) = f_1(X) \cup f_2(X) \cup \dots \cup f_k(X) \cup Y, \quad (1.2.5)$$

где Y — фиксированное компактное подмножество в M . Новое отображение F будет также сжимающим. Это легко вывести из следующего факта.

УПРАЖНЕНИЕ 6. Покажите, что «постоянное» отображение f_Y , которое переводит любое множество $X \in \mathbb{K}(M)$ в $Y \in \mathbb{K}(M)$, является сжимающим.

Значит, для любого $X \in \mathbb{K}(M)$ последовательность $\{X_n\}_{n \geq 1}$, где

$$X_n := F^n(X) := F(F(\dots F(X_0)\dots)),$$

сходится в $\mathbb{K}(M)$ и ее предел X_∞ является неподвижной точкой для преобразования F в $\mathbb{K}(M)$.

ОПРЕДЕЛЕНИЕ 1.11. Множество X_∞ , определенное выше, называется *неоднородным самоподобным фракталом*.

Теперь пора от общих слов перейти к конкретным примерам.

1. *Канторово множество C* . В этом случае $M = [0, 1]$, $f_1(x) = \frac{1}{3}x$, $f_2(x) = \frac{x+2}{3}$. Поучительно, как множество C , которое является неподвижной точкой для отображения F , приближается последовательностью множеств $\{C_n\}$, определенных рекуррентным соотношением $C_{n+1} = F(C_n)$.

Положим сначала $C_1 = [0, 1]$; тогда

$$C_2 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right],$$

$$C_3 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right], \quad \dots$$

Последовательность $\{C_n\}$ — убывающая: $C_{n+1} \subset C_n$ и предельное множество в этом случае — это просто пересечение $C = \bigcap_{n \geq 1} C_n$. (Напомним, что пересечение любого семейства непустых компактных множеств компактно и непусто.)

Теперь положим $C'_1 = \{0, 1\}$. Тогда

$$C'_2 = \left\{0, \frac{1}{3}, \frac{2}{3}, 1\right\}, \quad C'_3 = \left\{0, \frac{1}{9}, \frac{2}{9}, \frac{1}{3}, \frac{2}{3}, \frac{7}{9}, \frac{8}{9}, 1\right\}, \quad \dots$$

Последовательность $\{C'_n\}$ — возрастающая: $C'_{n+1} \supset C'_n$. В этом случае мы могли бы ожидать, что предел — это объединение $C'_\infty := \bigcup_{n \geq 1} C'_n$. Однако это объединение не замкнуто, следовательно, не является точкой $\mathbb{K}(M)$ и не может быть пределом $\{C'_n\}$. Но это легко поправить: оказывается, что искомым предел является замыканием C'_∞ .

На этом примере хорошо видно основное свойство самоподобных фракталов: если мы рассматриваем кусок канторова множества под микроскопом, увеличивающим в 3^n раз, мы видим в точности ту же картину, что и невооруженным глазом.

2. I_α -фрактал. Пусть Y — отрезок на плоскости \mathbb{R}^2 , заданный условиями $x = 0$, $-1 \leq y \leq 1$. Выберем вещественное число $\alpha \in \left(0, \frac{1}{\sqrt{2}}\right)$ и определим отображения

$$f_1(x, y) = (-\alpha y, \alpha x + 1); \quad f_2(x, y) = (-\alpha y, \alpha x - 1). \quad (1.2.6)$$

Неоднородный самоподобный фрактал, определяемый этими отображениями и подмножеством Y , показан на рис. 1.5.

Название этого фрактала происходит из того, что его первое приближение $Y \cup f_1(Y) \cup f_2(Y)$ для малых α выглядит как заглавная латинская буква I.

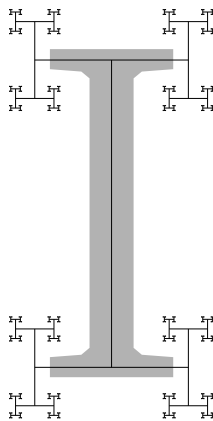


Рис. 1.5. I_α -фрактал для $\alpha = 0,5$

УПРАЖНЕНИЕ 7. Вычислите:

а) диаметр D множества I_α как подмножества метрического пространства \mathbb{R}^2 (то есть наибольшее возможное расстояние между точками I_α);

б) максимальную возможную длину L несамопересекающегося пути в I_α .

ОТВЕТ. а) $D = 2 \frac{\sqrt{1+\alpha^2}}{1-\alpha^2}$; б) $L = \frac{2}{1-\alpha}$.

3. *Ковер Серпинского* \mathcal{S} . Пусть $M = \mathbb{C}$ — комплексная плоскость с обычным расстоянием $d(z_1, z_2) = |z_1 - z_2|$.

Обозначим через $\omega := e^{\pi i/3}$ корень шестой степени из 1. Определим

$$f_1(z) = \frac{z}{2}, \quad f_2(z) = \frac{z+\omega}{2}, \quad f_3(z) = \frac{z+1}{2}.$$

ОПРЕДЕЛЕНИЕ 1.12. Самоподобный фрактал, определенный с помощью $\{f_1, f_2, f_3\}$, называется *ковром Серпинского*.

Это — один из двух главных примеров, изучаемых в нашей книге. Есть три наиболее естественных выбора начального множества, которые мы обозначим соответственно $\mathcal{S}_1, \mathcal{S}'_1, \mathcal{S}''_1$. Как следует из общей теории, предел аппроксимирующей последовательности не зависит от выбора начального множества. Но сами последовательности выглядят по-разному.

Сначала выберем в качестве начального множества сплошной треугольник с вершинами $0, \omega, 1$, который мы обозначим \mathcal{S}''_1 . Тогда последовательность $\mathcal{S}''_n = F^{n-1}(\mathcal{S}''_1)$ — убывающая и ее пределом является пересечение $\mathcal{S} = \lim_{n \rightarrow \infty} \mathcal{S}''_n = \bigcap \mathcal{S}''_n$, см. рис. 1.6.



Рис. 1.6. Приближение \mathcal{S}''_6

Следующий кандидат в начальное множество — обычный (пустой) треугольник \mathcal{S}'_1 с теми же вершинами $0, \omega, 1$. Тогда последовательность $\mathcal{S}'_n = F^{n-1}(\mathcal{S}'_1)$ — возрастающая и ее пределом является замыкание \mathcal{S} объединения $\mathcal{S}'_\infty = \bigcup_{n \geq 0} \mathcal{S}'_n$.

УПРАЖНЕНИЕ 8. Пусть V_n, E_n, F_n означают соответственно число вершин, ребер и пустых треугольников в \mathcal{S}'_n . Подсчитайте V_n, E_n, F_n , доказав предварительно рекуррентные соотношения

$$V_{n+1} = 3V_n - 3, \quad E_{n+1} = 3E_n, \quad F_{n+1} = 3F_n + 1.$$

Наконец, примем в качестве начального множества \mathcal{S}_1 три точки $0, \omega, 1$. Тогда последовательность $\mathcal{S}_n = F^{n-1}(\mathcal{S}_1)$ — возрастающая и состоит из конечных множеств. В этом случае, как и в предыдущем, предел \mathcal{S} является замыканием объединения $\mathcal{S}_\infty = \bigcup_{n \geq 0} \mathcal{S}_n$. Заметим, что последнее множество счетно и состоит из всех вершин во всех множествах $\mathcal{S}'_n, n \geq 1$.

Мы будем называть аппроксимации $\{\mathcal{S}''_n\}, \{\mathcal{S}'_n\}$ и $\{\mathcal{S}_n\}$ соответственно двумерными, одномерными и нульмерными. Двумерные аппроксимации приближают наше множество сверху, а одномерные и нульмерные — снизу.

Замечание. Глядя на рис. 1.8, можно подумать, что некоторые точки \mathcal{S}_6 имеют 6 соседей. Такова, например, точка P — средняя в третьем ряду. Однако сравнение с рис. 1.7 показывает, что из этих 6 точек только 4 соединены ребрами с P . Поэтому только они являются настоящими соседями. Отметим также, что при включении \mathcal{S}_6 в \mathcal{S}_7 и

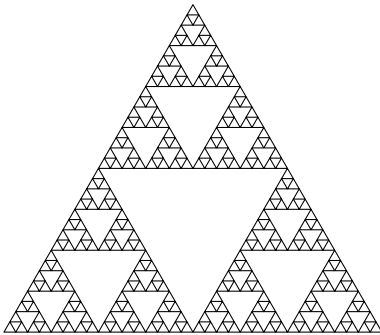


Рис. 1.7. Приближение \mathcal{S}'_6

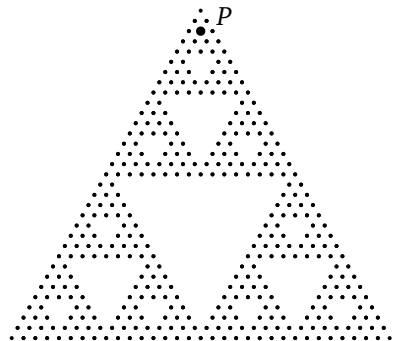


Рис. 1.8. Приближение \mathcal{S}_6

в последующие приближения точка P будет иметь ровно 4 ближайших соседа.

Таким образом, свойство «быть соседом» сильнее свойства «находиться на кратчайшем расстоянии». ♡

Схолия В. ХАУСДОРФОВА МЕРА И ХАУСДОРФОВА РАЗМЕРНОСТЬ

Мы оцениваем размер кривой с помощью ее длины, размер поверхности с помощью ее площади, а размер трехмерного тела с помощью его объема. А как измерить величину фрактала?

Решение этой проблемы было найдено в 1915 году немецким математиком Ф. Хаусдорфом. Он предложил для каждого вещественного числа $p > 0$ следующее определение p -мерной меры.

Пусть X — компактное подмножество в \mathbb{R}^n с обычным расстоянием. Тогда для каждого $\varepsilon > 0$ существует конечная ε -сеть для X . Другими словами, мы можем покрыть X конечным числом шаров радиуса ε . Обозначим через $N(\varepsilon)$ наименьшее число шаров радиуса ε , которыми можно покрыть X . Ясно, что $N(\varepsilon)$ растет при уменьшении ε и, как правило, стремится к бесконечности, когда ε стремится к нулю.

Вопрос: какова скорость этого стремления? Предположим, что рост $N(\varepsilon)$ имеет степенной характер, а именно что существует предел

$$c = \lim_{\varepsilon \rightarrow 0} N(\varepsilon) \cdot \varepsilon^p. \quad (1.2.1)$$

В этом случае мы пишем $N(\varepsilon) \cong c \cdot \varepsilon^{-p}$. Константа c , возникающая здесь, называется p -мерной мерой Хаусдорфа множества X и обозначается $\mu_p(X)$.

Мы не будем обсуждать здесь общее понятие меры. Для наших целей достаточно знать, что мера Хаусдорфа имеет следующие свойства:

1) монотонность: если $X \subset Y$, то $\mu_p(X) \leq \mu_p(Y)$.

2) полуаддитивность: если $X \subset \bigcup_{k=1}^{\infty} Y_k$, то

$$\mu_p(X) \leq \sum_{k=1}^{\infty} \mu_p(Y_k); \quad (1.2.2)$$

3) аддитивность: если X_i , $1 \leq i \leq m$, — компактные множества и $\mu_p(X_i \cap X_j) = 0$ для $i \neq j$, то

$$\mu_p\left(\bigcup_{i=1}^m X_i\right) = \sum_{i=1}^m \mu_p(X_i); \quad (1.2.3)$$

4) однородность: если $X \subset \mathbb{R}^n$, то для любого $\lambda \in \mathbb{R}$ справедливо равенство

$$\mu_p(\lambda \cdot X) = |\lambda|^p \cdot \mu_p(X). \quad (1.2.4)$$

Здесь $\lambda \cdot X$ означает множество, полученное из X умножением всех его точек на λ .

На самом деле первое свойство является частным случаем второго, но мы предпочли сформулировать его отдельно ввиду его простоты и важности.

Если p -мерная мера множества X конечна, мы говорим, что X имеет хаусдорфову размерность $\leq p$.

Если эта мера положительна, мы говорим, что X имеет хаусдорфову размерность p .

УПРАЖНЕНИЕ 9. Покажите, что если X имеет хаусдорфову размерность d , то предел (1.2.1) равен 0 для $p > d$ и бесконечности для $p < d$.

ЗАМЕЧАНИЕ. Имеется много вариантов определения p -мерной меры и соответствующей размерности. Например, вместо шаров радиуса ε можно использовать любые множества диаметра ε , или, в случае $M = \mathbb{R}^n$, n -мерные кубы со стороной ε .

Можно также рассматривать покрытия X подмножествами X_k разных размеров $\varepsilon_k \leq \varepsilon$ и вместо $N(\varepsilon)$ исследовать величину $\sqrt[p]{\sum_k \varepsilon_k^p}$.

Эти подходы могут на практике приводить к различным значениям p -мерной меры, но для «приличных» множеств, включая самоподобные фракталы, дают эквивалентные определения хаусдорфовой размерности. ♡

Во многих случаях нелегко доказать, что предел (1.2.1) существует, и еще труднее вычислить этот предел. Но часто выполняется более слабое условие, которое легче проверить:

$$N(\varepsilon) = O^*(\varepsilon^{-p}), \quad (1.2.5)$$

то есть $0 < c \cdot \varepsilon^{-p} \leq N(\varepsilon) \leq C \cdot \varepsilon^{-p} < \infty$ для достаточно малого ε .

В этом случае мы также будем считать, что множество X имеет хаусдорфову размерность p . Константы c и C в (1.2.5) дают нижнюю и верхнюю оценку для меры $\mu_p(X)$, если эта мера определена.

УПРАЖНЕНИЕ 10. Покажите, что хаусдорфова размерность множества X , если она существует, может быть задана формулой

$$d_H(X) = - \lim_{\varepsilon \searrow 0} \frac{\ln N(\varepsilon)}{\ln \varepsilon}. \quad (1.2.6)$$

УПРАЖНЕНИЕ 11. Покажите, что хаусдорфова размерность отрезка вещественной оси равна 1, размерность квадрата на плоскости равна 2, а размерность куба в пространстве равна 3.

Таким образом, во всех этих случаях хаусдорфова размерность совпадает с обычной¹.

ПРИМЕРЫ. Вычислим хаусдорфову размерность самоподобных фракталов, определенных выше. Во всех этих случаях мы предположим, что не только хаусдорфова размерность существует, но и соответствующая хаусдорфова мера определена. Этот факт не очевиден, но настойчивый читатель может попытаться доказать его самостоятельно.

После этого мы используем простое соображение, которое проиллюстрируем на примерах.

1. *Канторово множество* C . Предположим, что C имеет хаусдорфову размерность d и конечную ненулевую d -меру $\mu_d(C)$.

Множество C состоит из двух частей, подобных самому C с коэффициентом $\frac{1}{3}$. Из однородности d -меры следует, что каждая из этих частей имеет d -меру $\frac{1}{3^d}\mu_d(C)$. Теперь из аддитивности меры мы получаем уравнение $2 \cdot \left(\frac{1}{3}\right)^d = 1$, из которого следует, что $3^d = 2$, или

$$d = \log_3 2 = \frac{\ln 2}{\ln 3} \approx 0,63093\dots$$

Заметьте, что в этом рассуждении точное значение хаусдорфовой меры не участвует. Вы можете попытаться найти это значение, но его геометрическая интерпретация не так наглядна (и не так важна), как обычная длина, площадь или объем, потому что нет естественного эталона d -мерного множества, такого как единичный отрезок, квадрат или куб.

2. *I_α -фрактал*. Чтобы вычислить хаусдорфову размерность I_α , мы используем ту же схему. Предположим, что для некоторого числа d мы имеем $0 < \mu_d(I_\alpha) < \infty$. Напомним разложение

$$I_\alpha = f_1(I_\alpha) \cup f_2(I_\alpha) \cup Y.$$

¹ Чтобы читатель оценил это с виду простое замечание по достоинству, я замечу, что так называемая «обычная размерность» строго определяется только в специальных курсах топологии или (для очень специальных множеств) в курсах линейной алгебры. Так что подход Хаусдорфа полезен не только для изучения фракталов, но и в других разделах математики.

Поскольку множества $f_1(I_\alpha)$ и $f_2(I_\alpha)$ подобны I_α с коэффициентом α , мы приходим к уравнению $\mu_d(I_\alpha) = 2\alpha^d \mu_d(I_\alpha) + \mu_d(Y)$.

Заметим, что $1 \leq d \leq 2$, поскольку I_α содержит отрезок Y хаусдорфовой размерности 1 и содержится в квадрате хаусдорфовой размерности 2. Предположим сначала, что $d > 1$. Тогда, согласно упражнению 9, $\mu_d(Y) = 0$; поэтому мы приходим к уравнению $2 \cdot \alpha^d = 1$ и получаем

$$d = \log_\alpha \frac{1}{2} = -\frac{\ln 2}{\ln \alpha}. \quad (1.2.7)$$

Если $d = 1$, то хаусдорфова мера — обычная длина, и из рекуррентной формулы вытекает, что $\mu(I) = 2 + 2\alpha\mu(I)$, то есть $\mu(I) = \frac{2}{1-2\alpha}$, что согласуется с прямым определением длины как суммы отрезков. При $\alpha = \frac{1}{2}$ длина бесконечна, но размерность по-прежнему равна 1.

Правая часть (1.2.7) удовлетворяет неравенству $1 \leq d \leq 2$ в точности тогда, когда $\alpha \in \left[\frac{1}{2}, \frac{1}{\sqrt{2}} \right]$ (см. определение I_α).

УПРАЖНЕНИЕ 12. Докажите, что (1.2.7) дает правильное значение хаусдорфовой размерности I_α , когда $\alpha \in \left(\frac{1}{2}, \frac{1}{\sqrt{2}} \right)$.

Мы оставляем читателю разобрать самостоятельно случаи $\alpha = \frac{1}{2}$, $\alpha = \frac{1}{\sqrt{2}}$ и $\alpha \notin \left[\frac{1}{2}, \frac{1}{\sqrt{2}} \right]$. \diamond

Глава 2

ОПЕРАТОР ЛАПЛАСА НА КОВРЕ СЕРПИНСКОГО

Мощным математическим методом изучения данного множества X является рассмотрение пространств функций на этом множестве.

Например, если X — конечное множество, полезно ввести линейное пространство всех вещественных или комплексных функций на X ; если X — топологическое пространство, можно изучать пространство непрерывных функций на X ; если X — область в \mathbb{R}^n или, более общо, гладкое многообразие, то уместно рассмотрение всех гладких функций на X ; если X — многообразие, на котором действует группа G , то особый интерес представляют функции, которые инвариантны относительно G или преобразуются известным образом. Этот последний случай подробно изучается в теории представлений.

Если M — гладкое многообразие, то на нем часто определены естественные дифференциальные операторы, и тогда представляет интерес изучение собственных функций таких операторов. Результаты такого изучения широко используются в прикладных задачах.

В последние полстолетия возникло новое обширное математическое направление — так называемая *спектральная геометрия*. Главный объект этого направления — изучение спектров естественных (то есть геометрически определенных) дифференциальных операторов.

Самой известной публикацией спектральной геометрии стала статья Марка Каца «Можно ли услышать форму барабана?». Этот вопрос на строгий математический язык переводится так:

Можно ли восстановить форму плоской области, если известен спектр оператора Лапласа на ней?

В последние 30 лет спектральная геометрия включает в себя анализ на фрактальных множествах. Мы отсылаем читателя к прекрасным обзорам [Str1, Tep] и оригинальным статьям [Str2, MT, Ram] по поводу деталей. В книге мы только кратко коснемся этой теории и сосредоточимся на изучении гармонических функций, то есть собственных функций оператора Лапласа, отвечающих нулевому собственному значению.

Схолия С. ОПЕРАТОР ЛАПЛАСА И ГАРМОНИЧЕСКИЕ ФУНКЦИИ

С.1. Анализ на римановом многообразии. В этом разделе мы предполагаем, что читатель знаком с элементами дифференциальной геометрии на римановых многообразиях (достаточно, например, знания теории двумерных поверхностей в трехмерном пространстве). Формально содержание этого раздела не используется в дальнейшем, но оно дает мотивировку для изучения оператора Лапласа и гармонических функций на фракталах.

Один из самых известных дифференциальных операторов, если не самый известный, — это оператор Лапласа Δ на \mathbb{R}^n , определенный формулой

$$\Delta f = \left(\sum_{k=1}^n \left(\frac{\partial}{\partial x^k} \right)^2 \right) f. \quad (2.0.1)$$

Замечательным свойством этого оператора является его инвариантность относительно группы E_n всех движений \mathbb{R}^n (и даже относительно группы всех изометрий \mathbb{R}^n , включающей движения и отражения).

Оператор второго порядка с такими свойствами можно определить на любом римановом многообразии M . Пусть x^1, x^2, \dots, x^n — локальная система координат на M ; для краткости через ∂_k мы обозначим оператор $\frac{\partial}{\partial x^k}$ взятия частной производной¹ по переменной x^k .

Касательный вектор к многообразию M в какой-нибудь точке $x \in M$, покрываемой выбранной системой координат, имеет вид

$$\vec{v} = \sum_{k=1}^n v^k \partial_k, \quad (2.0.2)$$

где $\{v^k\}$ — координаты вектора \vec{v} .

Векторное поле в области, покрываемой системой координат, имеет тот же вид (2.0.2), но теперь $\{v_k\}$ — уже не числа, а функции от координат x^1, x^2, \dots, x^n точки x .

Пусть $g = g_{ij}(x)$ — координаты метрического тензора на M , образующие вещественную симметрическую положительно определенную матрицу порядка $n = \dim M$. Длина касательного вектора $v = \{v^k\}$ в точке $x_0 \in M$ с координатами $x_0^1, x_0^2, \dots, x_0^n$ определяется формулой

$$|v|^2 = \sum_{i,j} g_{ij}(x_0) v^i v^j.$$

¹ По традиции в дифференциальной геометрии координаты обычно снабжаются верхними, а дифференцирования — нижними индексами. Подробнее о том, почему это удобно и важно, будет сказано ниже.

Как известно, это выражение дает одно и то же значение для $|v|$ при любом выборе локальной системы координат.

Рассмотрим теперь матрицу, обратную к $\|g_{ij}\|$, и обозначим ее $\|g^{ij}\|$. Геометрический смысл этой матрицы — это квадратичная форма в так называемом *кокасательном пространстве*, двойственном к касательному пространству. Можно также интерпретировать $\|g^{ij}\|$ как матрицу симметричного¹ линейного оператора \check{g} из кокасательного в касательное пространство.

Дифференциал гладкой функции f на M является гладким ковекторным полем. В каждой локальной системе координат он имеет вид

$$df(x) = \sum_{k=1}^n \partial_k f dx^k. \quad (2.0.3)$$

Используя оператор \check{g} , мы можем «поднять индекс» и преобразовать ковекторное поле df в векторное поле v с координатами

$$v^k(x) := \sum_{j=1}^n g^{k,j}(x) (\partial_j f)(x).$$

Это векторное поле называется *градиентом* функции f и обозначается $\text{grad } f$. Таким образом,

$$\text{grad } f = \sum_{k=1}^n (\text{grad } f)^k \partial_k = \sum_{j=1}^n g^{k,j} \partial_j f \partial_k.$$

С другой стороны, на пространстве векторных полей на римановом многообразии M определена естественная операция *дивергенция*, которая ставит в соответствие векторному полю v функцию $\text{div } v$. Эта операция выглядит особенно просто, если выбрать локальную систему координат так, что $\det \|g_{ij}(x)\| \equiv 1$ (такие системы называют *унимодулярными*; известно, что любую систему можно сделать унимодулярной, изменив одну координату). В унимодулярной системе координат дивергенция определяется формулой

$$\text{div } v = \sum_k \partial_k v^k.$$

Наконец, мы формулируем главное в этом разделе

¹ То есть совпадающего со своим двойственным; для понимания этого утверждения освежите свои знания линейной алгебры или просто читайте дальше.

ОПРЕДЕЛЕНИЕ С.1. Оператор Лапласа—Бельтрами Δ на римановом многообразии M задается формулой

$$\Delta f = \operatorname{div} \operatorname{grad} f.$$

В каждой точке многообразия оператор Лапласа—Бельтрами может быть записан в простейшем виде (2.0.1) за счет выбора подходящей локальной системы координат. Однако такое выражение в целой окрестности данной точки, вообще говоря, невозможно. Препятствием служит кривизна метрики на M .

Существует другое, более геометрическое определение оператора Лапласа—Бельтрами. Рассмотрим ε -окрестность $U_\varepsilon(x_0)$ точки x_0 . Интеграл функции f по этой окрестности при $\varepsilon \rightarrow 0$ допускает следующее асимптотическое выражение:

$$\int_{U_\varepsilon(x_0)} f(x) d^n x = a_n \varepsilon^n \cdot f(x_0) + b_n \varepsilon^{n+2} \cdot (\Delta f)(x_0) + o(\varepsilon^{n+2}), \quad (2.0.4)$$

где n — размерность многообразия, $a_n = \frac{\pi^{n/2}}{\Gamma(1+n/2)}$ — объем единичного шара в \mathbb{R}^n , а $b_n = \frac{n}{n+2} a_n$. Таким образом, мы можем определить значение $(\Delta f)(x_0)$ как предел

$$(\Delta f)(x_0) = \lim_{\varepsilon \rightarrow 0} \frac{1}{b_n \varepsilon^{n+2}} \int_{U_\varepsilon(x_0)} (f(x) - f(x_0)) d^n x, \quad (2.0.5)$$

который заведомо существует для любой функции с непрерывными вторыми частными производными.

ОПРЕДЕЛЕНИЕ С.2. Функция f на римановом многообразии, удовлетворяющая уравнению $\Delta f = 0$, называется *гармонической*.

Известно, что на каждом римановом многообразии постоянной кривизны (например, на евклидовом пространстве \mathbb{R}^n , на сфере S^n или на пространстве Лобачевского H^n) гармонические функции характеризуются свойством

$$\frac{1}{\operatorname{vol}(U_\varepsilon(x_0))} \int_{U_\varepsilon(x_0)} f(x) d^n x = f(x_0),$$

то есть среднее по любой сферической окрестности равно значению в центре. Это свойство имеет очень важное следствие.

ТЕОРЕМА С.1 (принцип максимума). *Предположим, что M — связанное риманово многообразие с границей. Тогда любая непостоянная гармоническая функция на M может достигать максимума только на границе ∂M .*

Известно также, что если M компактно, имеет гладкую границу ∂M и φ — любая непрерывная функция на ∂M , то существует единственная гармоническая функция f на M , для которой $f|_{\partial M} = \varphi$.

Более того, для каждой точки $m \in M$ существует такая вероятностная мера μ_m на ∂M , что $f(m) = \int_{\partial M} \varphi(x) d\mu(x)$. Эта мера называется *мерой Пуассона* и задается гладкой плотностью на многообразии ∂M .

Имеется красивая и простая физическая интерпретация гармонических функций (как равновесных распределений тепла) и вероятностная интерпретация меры Пуассона $\mu_m(A)$ (как вероятности выхода на границу в данной области $A \subseteq \partial M$ при случайном блуждании по M с начальной точкой $m \in M$).

С.2. Оператор как отношение двух квадратичных форм. Возможен чисто алгебраический подход к определению оператора Лапласа.

Предположим, что на вещественном линейном пространстве V заданы две квадратичные формы Q_0 и Q_1 . Предположим также, что Q_0 положительна: $Q_0(v) > 0$ для всех $v \neq 0$, а Q_1 — неотрицательна: $Q_1(v) \geq 0$.

Тогда мы можем ввести в V скалярное произведение

$$(v_1, v_2) := \frac{Q_0(v_1 + v_2) - Q_0(v_1) - Q_0(v_2)}{2}. \quad (2.0.6)$$

Если V бесконечномерно, мы предположим дополнительно, что оно полно относительно нормы $\|v\|^2 := (v, v) = Q_0(v)$. Таким образом, V — вещественное гильбертово пространство. На самом деле условие полноты легко выполнить: достаточно заменить V его пополнением \overline{V} относительно введенной нормы.

Другая квадратичная форма Q_1 определена только на плотном подпространстве $V \subset \overline{V}$. Из теории операторов в гильбертовых пространствах мы знаем, что в рассматриваемом случае справедливо

ПРЕДЛОЖЕНИЕ С.1. *Существует такой симметричный неотрицательный оператор A в V с областью определения $Dom(A)$, что*

$$Q_1(v) = (Av, v) \quad \text{для всех } v \in Dom(A) \supset V.$$

ЗАМЕЧАНИЕ. Иногда A называют отношением квадратичных форм Q_1 и Q_0 . В самом деле, любая квадратичная форма Q определяет

симметричную билинейную форму $B_Q: V \times V \rightarrow V$ по формуле:

$$B_Q(v_1, v_2) := \frac{Q(v_1 + v_2) - Q(v_1) - Q(v_2)}{2}. \quad (2.0.7)$$

Всякая билинейная форма B , в свою очередь, может рассматриваться как отображение $\tilde{B}: V \rightarrow V^*$. А именно, вектору $v \in V$ ставится в соответствие линейный функционал $f \in V^*$ по формуле

$$f(w) = \tilde{Q}(v, w). \quad (2.0.8)$$

Таким образом, из двух квадратичных форм мы получаем два линейных оператора из V в V^* : \tilde{B}_{Q_0} и \tilde{B}_{Q_1} .

Мы предоставляем читателю убедиться, что рассматриваемый оператор A можно записать как отношение $A = \tilde{B}_{Q_0}^{-1} \circ \tilde{B}_{Q_1}$. ♡

В случае, когда V конечномерно, мы можем применить стандартную теорему анализа об условном экстремуме к задаче отыскания экстремумов формы Q_1 на множестве, определяемом условием $Q_0(v) = 1$. В качестве результата мы получаем

Следствие. Собственные значения и собственные векторы оператора A являются в точности критическими значениями и критическими точками функции $Q_1(v)$ на сфере¹ $Q_0(v) = 1$.

В случае бесконечномерного V аналогичный результат также имеет место при дополнительном ограничении: сфера $Q_0(v) = 1$ компактна в топологии, определяемой формой Q_1 .

С.3. Оператор Лапласа на римановом многообразии. Применим общую схему, описанную выше, в следующей ситуации. Пусть M — гладкое компактное риманово многообразие с границей. Возьмем в качестве V пространство гладких функций на M , удовлетворяющих некоторому граничному условию (см. ниже).

На V имеются две квадратичные формы:

$$Q_0(v) = \int_M v^2(m) dm \quad \text{и} \quad Q_1(v) = \int_M |\text{grad } v|^2 dm, \quad (2.0.9)$$

где мера t на M и скалярный квадрат $|\text{grad } v|^2$ определяются метрическим тензором.

¹ Эквивалентная формулировка: собственные значения и собственные векторы оператора A — это в точности критические значения и критические точки функции $Q(v) := \frac{Q_1(v)}{Q_0(v)}$ на $V \setminus \{0\}$.

Согласно общей схеме, существует оператор A на $\bar{V} = L^2(M, dm)$ такой, что

$$\int_M (\text{grad } v_1, \text{grad } v_2) dm = \int_M Av_1(m) \cdot v_2(m) dm. \quad (2.0.10)$$

С другой стороны, прямое вычисление с помощью формулы Стокса дает для интеграла в левой части выражение

$$\int_{\partial M} v_1 \partial_\nu v_2 dn - \int_M \Delta v_1(m) \cdot v_2(m) dm, \quad (2.0.11)$$

где ∂_ν — производная в направлении внутренней нормали к границе, а dn — мера на границе ∂M , рассматриваемой как риманово многообразие с метрическим тензором, наследуемым от M .

Предположим, что граничные условия выбраны так, что интеграл по границе обращается в нуль. Тогда оператор $A = \tilde{B}_{Q_0}^{-1} \circ \tilde{B}_{Q_1}$ совпадает с оператором $-\Delta$.

Два частных случая этой ситуации хорошо известны: *задача Дирихле*, когда граничное условие имеет вид

$$v|_{\partial M} = 0, \quad (2.0.12)$$

и *задача Неймана* с граничным условием

$$\partial_\nu v|_{\partial M} = 0. \quad (2.0.13)$$

В обоих случаях интеграл по границе обращается в нуль и оператор $-\Delta$ является неотрицательным самосопряженным оператором в $L^2(M, dm)$. Область определения $\text{Dom}(\Delta)$ состоит из непрерывно дифференцируемых функций v на M , удовлетворяющих граничному условию и таких, что $\Delta v \in L^2(M, dm)$ в смысле обобщенных функций.

Описанная здесь связь оператора Δ с вариационными задачами дает замечательное физическое истолкование собственных значений и собственных функций оператора Лапласа—Бельтрами. А именно, многообразие M рассматривается как упругая мембрана в \mathbb{R}^{n+1} . Собственные значения определяют частоту малых колебаний этой мембраны, а собственные функции — форму этих колебаний.

Вопрос, каким может быть спектр оператора Лапласа—Бельтрами на компактном гладком римановом многообразии, является одним из основных в спектральной геометрии. Поскольку фракталы играют все бóльшую роль в современных исследованиях, в последние годы стало популярным изучение аналогов оператора Лапласа—Бельтрами на фракталах. Мы отсылаем читателя к обзорам [Тер, Str1] и оригинальным статьям, которые там указаны. \diamond

2.1. ОПЕРАТОР ЛАПЛАСА НА \mathcal{S}_n

В первоначальном варианте этой книги я попытался привести точное определение операторов Лапласа на \mathcal{S}_n и \mathcal{S} , а также описать подробно их спектр. Потом я обнаружил, что эта программа уже реализована независимо несколькими физиками и математиками (см., например, [Ram, FS, HL]). По этой причине я решил не повторять еще раз одни и те же результаты, а сосредоточиться на других, менее известных задачах. Таким образом, здесь я ограничиваюсь кратким описанием довольно интересной техники, используемой при изучении спектра.

Чтобы определить аналог оператора Лапласа—Бельтрами на ковре Серпинского \mathcal{S} , мы рассмотрим сначала его конечное нульмерное приближение \mathcal{S}_n . Попробуем следовать описанной выше схеме. Обозначим через V_n линейное пространство всех вещественных функций на \mathcal{S}_n . Поскольку \mathcal{S}_n состоит из $\frac{3^n+3}{2}$ точек, размерность V_n равна $d_n = \frac{3^n+3}{2}$.

Определим на V_n две квадратичные формы:

$$Q_0(v) = \sum_{s \in \mathcal{S}_n} v(s)^2; \quad Q_1(v) = \sum_{s' \leftrightarrow s''} ((v(s') - v(s''))^2, \quad (2.1.1)$$

где первая сумма распространяется на все точки \mathcal{S}_n , а вторая — на все пары соседних точек.

Ясно, что эти квадратичные формы являются прямыми аналогами форм, определенных в (2.0.9). Как и в случае обычного оператора Лапласа, мы используем форму Q_0 для определения скалярного произведения в V_n :

$$(f_1, f_2) = \sum_{s \in \mathcal{S}_n} f_1(s) f_2(s).$$

Тогда вторая форма может быть записана как

$$Q_1(f) = (\Delta_n f, f), \quad \text{где } (\Delta_n f)(s) = k(s)f(s) - \sum_{s' \leftrightarrow s} f(s'). \quad (2.1.2)$$

Здесь $k(s)$ означает число точек, соседних с s , то есть $k(s) = 4$ для внутренних точек и $k(s) = 2$ для точек границы. Стоит отметить, что граница $\partial \mathcal{S}_n$ в этом случае состоит из трех точек, образующих первое приближение \mathcal{S}_1 .

Введем два типа граничных условий.

Условие Дирихле дается формулой

$$f(s) = 0 \quad \text{для } s \in \partial\mathcal{S}_n. \quad (2.1.3)$$

Пространство функций, удовлетворяющих этому условию, обозначается $V_n^{(D)}$. Его размерность равна $d_n - 3 = \frac{3^n - 3}{2}$. Оператор $\Delta_n^{(D)}$ действует в $V_n^{(D)}$ по формуле

$$(\Delta_n^{(D)} f)(s) = 4f(s) - \sum_{s' \leftrightarrow s} f(s'), \quad s \in \mathcal{S}_n \setminus \partial\mathcal{S}_n. \quad (2.1.4)$$

Условие Неймана дается формулой

$$2f(s) = f(s') + f(s''), \quad s \in \partial\mathcal{S}_n, \quad s', s'' \text{ — соседи } s. \quad (2.1.5)$$

Пространство функций, удовлетворяющих условию Неймана, обозначается $V_n^{(N)}$. Его размерность также равна $d_n - 3 = \frac{3^n - 3}{2}$. Оператор $\Delta_n^{(N)}$ действует в $V_n^{(N)}$ по той же формуле (2.1.4).

Оба оператора $\Delta_n^{(D)}$ и $\Delta_n^{(N)}$ самосопряжены и их спектры хорошо известны (см., например, [FS]).

Чтобы дать читателю представление об этих результатах, мы рассмотрим более подробно случай $n = 2$.

Пусть сначала $V = V_2^{(D)}$. Это — трехмерное пространство функций на \mathcal{S}_2 , значения которых показаны на рис. 2.1.

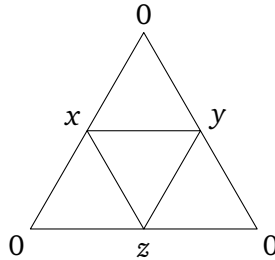


Рис. 2.1. Функции на \mathcal{S}_2 с условием Дирихле

Оператор $\Delta_2^{(D)}$ переводит тройку значений (x, y, z) в новую тройку $(4x - y - z, 4y - x - z, 4z - x - y)$. Относительно естественного

базиса этот оператор имеет матрицу $\begin{pmatrix} 4 & -1 & -1 \\ -1 & 4 & -1 \\ -1 & -1 & 4 \end{pmatrix}$. Собственные

значения этой матрицы легко сосчитать, используя следующий полезный факт.

ЛЕММА 2.1. Пусть матрица A размера $n \times n$ имеет элементы

$$a_{ij} = \begin{cases} a, & \text{если } i = j, \\ b, & \text{если } i \neq j. \end{cases}$$

Тогда спектр A состоит из собственного значения $a - b$ с кратностью $n - 1$ и еще одного собственного значения $a + (n - 1)b$. \square

В нашем случае мы получаем двукратное собственное значение 5 и простое собственное значение 2. Двумерное собственное подпространство состоит из векторов (x, y, z) с $x + y + z = 0$, а простое собственное подпространство — из векторов (x, y, z) с $x = y = z$.

Это значит, что соответствующая мембрана (с закрепленной границей) имеет две частоты колебаний с отношением $\sqrt{\frac{5}{2}} \approx 1,581$.

Пусть теперь $V = V_2^{(N)}$. Значения функций из этого пространства показаны на рис. 2.2.

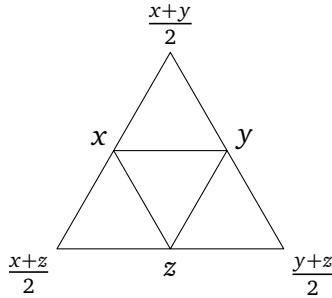


Рис. 2.2. Функции на \mathcal{S}_2 с граничным условием Неймана

Я оставляю читателю проверку того, что $\Delta_2^{(N)}$ переводит тройку (x, y, z) в тройку $\left(3x - \frac{3}{2}(y + z), 3y - \frac{3}{2}(y + z), 3z - \frac{3}{2}(y + z)\right)$. Матрица оператора тем самым имеет вид

$\begin{pmatrix} 3 & -3/2 & -3/2 \\ -3/2 & 3 & -3/2 \\ -3/2 & -3/2 & 3 \end{pmatrix}$. Спектр

состоит из двукратного собственного значения $4\frac{1}{2}$ и простого собственного значения 0.

Это значит, что мембрана с незакрепленной границей имеет одну частоту колебаний (несколько ниже, чем в первом случае) и одно равновесное состояние $x = y = z$.

2.2. СРАВНЕНИЕ СПЕКТРОВ Δ_n И Δ_{n-1}

Вычисления, которые мы проделываем в этом разделе, довольно скучны и громоздки, но они необходимы для получения в дальнейшем глубоких и красивых результатов о спектре оператора Лапласа.

Обозначим через V_n^λ пространство функций на \mathcal{S}_n , удовлетворяющих условию

$$(4 - \lambda)f(s) = \sum_{s \leftrightarrow s'} f(s') \quad (2.2.1)$$

во всех внутренних точках $s \in \mathcal{S}_n$.

Выберем функцию $f \in V_n^\lambda$ и предположим, что ее значение в некоторой точке $s \in \mathcal{S}_{n-1}$ равно $x \neq 0$. Рассмотрим подробнее часть \mathcal{S}_n , окружающую точку s . На прилагаемом рисунке надписаны значения f в точке s и в близких точках. Значения, не используемые в наших вычислениях, помечены знаком «?».».

$$\begin{array}{cccccc} & & & ? & & \\ & & & ? & & ? \\ & & y & ? & z & \\ u & & q & r & v & \\ b & p & x & t & c & \end{array}$$

Поскольку $f \in V_n^\lambda$, мы имеем систему уравнений

$$\begin{aligned} (4 - \lambda)x &= p + q + r + t; \\ (4 - \lambda)u &= b + y + p + q; & (4 - \lambda)v &= c + z + r + t; \\ (4 - \lambda)p &= b + u + q + x; & (4 - \lambda)q &= y + u + p + x; \\ (4 - \lambda)r &= z + v + t + x; & (4 - \lambda)t &= c + v + r + x. \end{aligned} \quad (2.2.2)$$

Сложив последние 4 равенства, мы получаем

$$(4 - \lambda)(p + q + r + t) = (p + q + r + t) + (b + y + z + c) + 2(u + v) + 4x, \quad (2.2.3)$$

а сложив два предыдущих, получаем

$$(4 - \lambda)(u + v) = (p + q + r + t) + (b + y + z + c). \quad (2.2.4)$$

Из (2.2.3), (2.2.4) мы можем выразить $(p + q + r + t)$ и $(u + v)$ через $(b + y + z + c)$ и x . Тогда первое уравнение (2.2.2) дает

$$(\lambda - 6)(b + y + z + c) = (\lambda - 6)(4 - \lambda)(1 - \lambda)x. \quad (2.2.5)$$

Мы приходим к альтернативе: или $\lambda = 6$, или ограничение функции

f на \mathcal{S}_{n-1} принадлежит подпространству V_{n-1}^μ , где μ находится из равенства

$$4 - \mu = (4 - \lambda)(1 - \lambda), \quad \text{или} \quad \mu = \lambda(5 - \lambda). \quad (2.2.6)$$

Первым важным следствием этой альтернативы является

ТЕОРЕМА 2.2. *Если f — гармоническая функция на \mathcal{S}_n , то ее ограничение на \mathcal{S}_{n-1} — тоже гармоническая функция.*

В самом деле, для гармонических функций $\lambda = 0$, следовательно, $\mu = \lambda(5 - \lambda)$ также равно нулю.

Этот факт приводит к естественному определению гармонических функций на \mathcal{S}_∞ .

ОПРЕДЕЛЕНИЕ 2.3. *Функция на \mathcal{S}_∞ называется гармонической, если таково ее ограничение на каждое подмножество $\mathcal{S}_n \subset \mathcal{S}_\infty$.*

2.3. СПЕКТР ОПЕРАТОРА ЛАПЛАСА НА \mathcal{S}_n

Здесь мы рассмотрим кратко природу спектров операторов $\Delta_n^{(D)}$, имея в виду построение оператора Лапласа $\Delta^{(D)}$ на \mathcal{S} .

Сначала исследуем динамику преобразования комплексной плоскости, заданного квадратичным многочленом $P(\lambda) = \lambda(5 - \lambda)$. Для каждого вещественного числа μ мы назовем μ -орбитой любую последовательность $\{\mu_k\}_{k \geq 0}$, для которой $\mu_0 = \mu$, а остальные члены связаны соотношениями: $P(\mu_k) = \mu_{k-1}$ для $k \geq 1$.

Если мы хотим продолжить ненулевую функцию $f \in V_n^{\mu_n}$ на \mathcal{S}_{n+1} так, чтобы продолженная функция принадлежала $V_{n+1}^{\mu_{n+1}}$, то из (2.2.6) мы видим, что это возможно, лишь если μ_n и μ_{n+1} лежат на одной μ -орбите.

Обратно, для любой μ -орбиты $\{\mu_k\}$ мы можем построить такую функцию f на \mathcal{S}_∞ , что ее ограничение на \mathcal{S}_n (которое может тождественно обращаться в нуль!) принадлежит $V_n^{\mu_n}$ для всех n .

Если мы хотим теперь продолжить построенную таким образом функцию f с \mathcal{S}_∞ до непрерывной функции на \mathcal{S} , мы должны убедиться, что f равномерно непрерывна на \mathcal{S}_∞ . В случае, когда это так, мы называем продолженную функцию \tilde{f} собственной функцией оператора Лапласа на всем \mathcal{S} . Соответствующее собственное значение является пределом ренормализованной подходящим образом последовательности $\{\mu_n\}$.

В этой книге мы рассматриваем только очень частный случай $\mu = 0$ (то есть гармонические функции на \mathcal{S}). Зато мы делаем это подробно, используя весь математический аппарат: геометрию, анализ, арифметику и теорию вероятностей.

Глава 3

ГАРМОНИЧЕСКИЕ ФУНКЦИИ НА КОВРЕ СЕРПИНСКОГО

В этой главе мы изучим подробнее гармонические функции на ковре Серпинского. Заметим, что гармонические функции, удовлетворяющие условию Дирихле, обращаются в нуль, а гармонические функции, удовлетворяющие условию Неймана, должны быть константами. Поэтому здесь мы рассматриваем гармонические функции, на которые не налагается никаких граничных условий.

Напомним, что граничными точками ковра \mathcal{S} являются три точки $0, 1, \omega = \frac{1+i\sqrt{3}}{2}$. Отрезок $[0, 1]$ вещественной оси принадлежит \mathcal{S} , и мы можем рассматривать ограничения гармонических функций на этот отрезок как обычные вещественные функции на $[0, 1]$. Оказывается, что эти функции имеют весьма нетривиальные аналитические и арифметические свойства.

3.1. Основные свойства гармонических функций

Начнем с рассмотрения гармонических функций на подмножествах \mathcal{S}_n и \mathcal{S}_∞ . Мы видели выше, что каждая гармоническая функция на \mathcal{S}_n продолжается единственным образом до гармонической функции на \mathcal{S}_∞ и удовлетворяет принципу максимума на каждой треугольной части ковра Серпинского, подобной самому ковру.

Лемма 3.1. Множество $\mathcal{H}(\mathcal{S}_\infty)$ всех гармонических функций на \mathcal{S}_∞ является трехмерным векторным пространством (вещественным или комплексным в зависимости от вида рассматриваемых функций).

Тремя естественными координатами функции $f \in \mathcal{H}(\mathcal{S}_\infty)$ являются ее значения в трех граничных точках.

Доказательство. Из линейной алгебры мы знаем, что если однородная система N линейных уравнений с N неизвестными имеет только тривиальное решение, то соответствующая неоднородная система разрешима при любой правой части и имеет единственное решение. В нашем случае условие гармоничности функции f на \mathcal{S}_m — это $N = \frac{3^{m+1} - 3}{2}$ линейных уравнений (по одному для каждой внут-

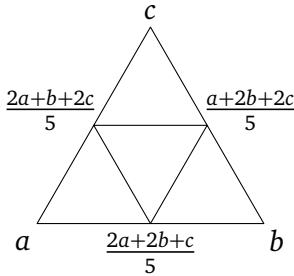
ренной точки) с N неизвестными (значениями f во внутренних точках). Из принципа максимума следует, что однородная система имеет только нулевое решение, поэтому для любых граничных значений существует единственная гармоническая функция на \mathcal{S}_m с такими граничными значениями. \square

Нам понадобится также следующее простое наблюдение.

ЛЕММА 3.2. Пусть x, y, z — три соседние точки на \mathcal{S}_m , образующие правильный треугольник, обращенный вершиной вверх. Положим

$$\alpha = \frac{y+z}{2}, \quad \beta = \frac{x+z}{2}, \quad \gamma = \frac{x+y}{2}. \text{ Тогда } \alpha, \beta, \gamma$$

также образуют правильный треугольник и являются соседними точками на \mathcal{S}_{m+1} (см. рис. 3.1). Для любой гармонической функции f на \mathcal{S}_{m+1} мы имеем:



$$f(\alpha) = \frac{f(x) + 2f(y) + 2f(z)}{5},$$

$$f(\beta) = \frac{2f(x) + f(y) + 2f(z)}{5}, \quad (3.1.1)$$

$$f(\gamma) = \frac{2f(z) + 2f(y) + f(x)}{5}.$$

Рис. 3.1. Правило 1 : 2 : 2

Неформальный смысл этого результата: соседняя точка оказывает вдвое большее влияние, чем противоположная.

Теперь мы можем доказать следующий важный факт.

ТЕОРЕМА 3.1. Каждая гармоническая функция на \mathcal{S}_∞ равномерно непрерывна и, следовательно, имеет единственное продолжение по непрерывности на \mathcal{S} .

Доказательство. Обозначим через $f_{a,b}^c$ гармоническую функцию на \mathcal{S}_∞ с граничными значениями

$$f(0) = a, \quad f(1) = b, \quad f(\omega) = c.$$

Назовем вариацией функции f на множестве X величину

$$\text{var}_X f = \sup_{x,y \in X} |f(x) - f(y)|.$$

Из принципа максимума мы заключаем, что

$$\text{var}_{\mathcal{S}_\infty} f_{a,b}^c = \max\{|a - b|, |b - c|, |c - a|\}.$$

Из леммы 3.2 с помощью индукции по n легко выводится, что для любых соседних точек x, y из \mathcal{S}_n мы имеем:

$$|f_{a_b}^c(x) - f_{a_b}^c(y)| \leq \text{var } f_{a_b}^c \cdot \left(\frac{3}{5}\right)^n \leq \text{const} \cdot d(x, y)^\beta, \quad \beta = \log_2 \frac{5}{3}.$$

Значит, функция $f_{a_b}^c$ принадлежит классу Гёльдера H_β . Следовательно, она равномерно непрерывна и продолжается по непрерывности на \mathcal{S} . Мы сохраним то же обозначение $f_{a_b}^c$ для продолженной функции. \square

Теперь мы обсудим вопрос: что значит вычислить гармоническую функцию? Для этого нужно каким-то образом параметризовать точки ковра Серпинского. Пример канторова множества подсказывает естественный способ нумерации точек \mathcal{S} бесконечными троичными дробями. А именно, пусть $x \in \mathcal{S}$; тогда соответствующая дробь строится таким образом. Первый знак этой дроби (0, 1 или 2) указывает, в какой трети ковра находится точка x (в левой, правой или верхней). Затем эта треть делится опять на три части и следующий знак дроби указывает, в какой из них лежит точка x , и т. д. Обратное, если задана бесконечная троичная дробь, мы можем последовательно выбрать одну из трех частей \mathcal{S} сообразно с первым знаком дроби, затем третью часть этой трети сообразно со вторым знаком и т. д. Мы получим последовательность вложенных компактных подмножеств \mathcal{S} , которые имеют единственную общую точку. Она и соответствует данной троичной дроби.

На самом деле построенное соответствие не будет взаимно однозначным. В то время как бесконечная троичная дробь определяет точку $x \in \mathcal{S}$ однозначно, обратное соответствие нарушает это свойство, когда наша точка лежит в \mathcal{S}_∞ . В этом случае мы имеем выбор: какому из двух треугольников с данной вершиной отнести нашу точку.

УПРАЖНЕНИЕ 13. Покажите, что две троичные дроби, соответствующие одной точке \mathcal{S} , получаются друг из друга заменой «хвоста» вида хууу... на «хвост» вида ухххх...

УПРАЖНЕНИЕ 14. Покажите, что точка x , соответствующая дроби $0, a_1 a_2 a_3 \dots$, вычисляется по формуле

$$x = \sum_{k=1}^{\infty} \frac{b_k}{2^k}, \quad \text{где } b_k = \begin{cases} 0, & \text{если } a_k = 0, \\ 1, & \text{если } a_k = 1, \\ \omega, & \text{если } a_k = 2. \end{cases}$$

Отметим специальный случай, когда троичная дробь содержит только цифры 0 и 1. Тогда соответствующая точка x лежит на отрез-

ке $[0, 1]$ и наша троичная дробь фактически является двоичной и совпадает с двоичным представлением x .

3.2. Базисные функции χ, φ, ψ, ξ

Обозначим через $u_{a^c b}$ ограничение гармонической функции $f_{a^c b}^c$ на отрезок $[0, 1]$, который составляет горизонтальную сторону \mathcal{S} .

Следующие соотношения довольно очевидны и выводятся из действия группы перестановок S_3 на \mathcal{S} и на $\mathcal{H}(\mathcal{S})$:

$$u_{a^c b}(t) = u_{b^c a}(1-t); \quad u_{a^c b}(t) + u_{b^c c}(t) + u_{c^c a}(t) \equiv a + b + c. \quad (3.2.1)$$

Отсюда следует, что значения любой гармонической функции в любой точке \mathcal{S}_m легко находятся, если известна единственная функция $\varphi := u_{0^c 1}$ на $[0, 1]$.

УПРАЖНЕНИЕ 15. Выведите из (3.2.1), что

$$u_{a^c b}(t) = c + (b-c)\varphi(t) + (a-c)\varphi(1-t). \quad (3.2.2)$$

Поэтому важно получить как можно больше информации о природе функции φ . Однако, как мы вскоре увидим, удобнее ввести наряду с φ три других функции:

$$\begin{aligned} \chi(t) &:= u_{0^c 1}(t) = -1 + 2\varphi(t) + \varphi(1-t), \\ \psi(t) &:= u_{0^c 1}(t) = 1 - \varphi(1-t), \\ \xi(t) &:= u_{0^c 1}(t) = 2 - \varphi(t) - 2\varphi(1-t). \end{aligned} \quad (3.2.3)$$

Мы будем называть эти функции *базисными*. Причина, по которой вводится четыре функции вместо одной, заключается в следующем. Пусть \mathcal{H} означает пространство всех вещественных функций на отрезке $[0, 1]$, состоящее из ограничений гармонических функций на \mathcal{S} . Оно имеет размерность 3; в качестве базиса можно взять константу и любые две функции из четверки χ, φ, ψ, ξ .

Рассмотрим три отображения отрезка $[0, 1]$ в себя: $\alpha_0(t) = \frac{t}{2}$, $\alpha_1(t) = \frac{1+t}{2}$ и $\tau(t) = 1-t$. Они порождают линейные операторы в пространстве $C([0, 1])$ непрерывных функций на отрезке:

$$(A_0 f)(t) = f\left(\frac{t}{2}\right), \quad (A_1 f)(t) = f\left(\frac{1+t}{2}\right) \quad \text{и} \quad (Tf)(t) = f(1-t). \quad (3.2.4)$$

Оказывается, все три оператора A_0, A_1 и T переводят в себя трехмерное подпространство $\mathcal{H} \subset C([0, 1])$. Более того, операторы A_0 и

A_1 имеют в \mathcal{H} одинаковый спектр, состоящий из трех собственных значений: $1, \frac{3}{5}, \frac{1}{5}$. Соответствующими собственными функциями являются $1, \psi, \chi$ для A_0 и $1, 1 - \xi, 1 - \varphi$ для A_1 .

Другими словами, если мы введем вектор-функции

$$\vec{f}(x) = \begin{pmatrix} \psi(x) \\ \chi(x) \\ 1 \end{pmatrix} \quad \text{и} \quad \vec{g}(x) = \begin{pmatrix} \varphi(x) \\ \xi(x) \\ 1 \end{pmatrix}, \quad (3.2.5)$$

то справедливы равенства

$$\vec{f}\left(\frac{t}{2}\right) = A_0 \vec{f}(t), \quad \vec{g}\left(\frac{1+t}{2}\right) = A_1 \vec{g}(t), \quad \vec{f}(1-t) = T \vec{g}(t), \quad (3.2.6)$$

где

$$A_0 = \begin{pmatrix} 3/5 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 3/5 & 0 & 2/5 \\ 0 & 1/5 & 4/5 \\ 0 & 0 & 1 \end{pmatrix}, \quad T = \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{pmatrix}. \quad (3.2.7)$$

УПРАЖНЕНИЕ 16. С помощью равенств (3.2.6), (3.2.7) заполните пустые места в таблице значений базисных функций χ, φ, ψ, ξ в точках $k/8, k = 0, 1, \dots, 7, 8$.

	0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{5}{8}$	$\frac{3}{4}$	$\frac{7}{8}$	1
χ	0	$\frac{1}{125}$	$\frac{1}{25}$		$\frac{1}{5}$				1
φ	0				$\frac{2}{5}$		$\frac{98}{125}$		1
ψ	0	$\frac{27}{125}$	$\frac{9}{25}$		$\frac{3}{5}$				1
ξ	0				$\frac{4}{5}$		$\frac{24}{25}$		1

Из (3.2.6) мы выведем несколько удивительных свойств введенных выше функций. Например, мы исследуем поведение этих функций в окрестности любой двоично-рациональной точки $r = \frac{k}{2^n}$.

ЛЕММА 3.3. Все четыре функции χ, φ, ψ и ξ строго монотонно возрастают от 0 до 1 на отрезке $[0, 1]$.

Доказательство. Достаточно проверить, что функции $\xi(t)$ и $\chi(t)$ строго монотонно возрастают, поскольку $\varphi(t) = \frac{\xi(t) + 2\chi(t)}{3}$, а

$\psi(t) = \frac{2\xi(t) + \chi(t)}{3}$. Пусть $0 \leq t < s \leq 1$. Мы должны показать, что $\xi(t) < \xi(s)$ и $\chi(t) < \chi(s)$.

Введем вектор-функцию $\vec{h}(t) := \begin{pmatrix} \xi(t) \\ \chi(t) \\ 1 \end{pmatrix}$. Из (3.2.6) вытекают следующие правила преобразования \vec{h} :

$$\vec{h}\left(\frac{t}{2}\right) = B_0 \vec{h}(t); \quad \vec{h}\left(\frac{1+t}{2}\right) = B_1 \vec{h}(t), \quad (3.2.8)$$

где

$$B_0 = \begin{pmatrix} 3/5 & 1/5 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad B_1 = \begin{pmatrix} 1/5 & 0 & 4/5 \\ 1/5 & 3/5 & 1/5 \\ 0 & 0 & 1 \end{pmatrix}. \quad (3.2.9)$$

Рассмотрим двоичную запись чисел t и s :

$$t = 0, t_1 t_2 \dots t_k \dots, \quad s = 0, s_1 s_2 \dots s_k \dots$$

Мы можем предположить, что для некоторого m справедливы соотношения: $t_i = s_i$ для $i < m$, $t_m = 0$, $s_m = 1$.

Применяя несколько раз (3.2.8), получаем:

$$\vec{h}(t) = B_{t_1} \dots B_{t_{m-1}} B_0 \vec{h}(z), \quad \vec{h}(s) = B_{t_1} \dots B_{t_{m-1}} B_1 \vec{h}(w)$$

для некоторых $z \in [0, 1)$, $w \in (0, 1]$. Поскольку матрица B_i имеет неотрицательные коэффициенты, достаточно проверить, что $B_1 \vec{h}(w) > B_0 \vec{h}(z)$. (Здесь мы придерживаемся соглашения: $\vec{a} > \vec{b}$, если первые две координаты вектора \vec{a} больше, чем соответствующие координаты вектора \vec{b} .)

Но

$$B_1 \vec{h}(w) = \begin{pmatrix} 1/5 & 0 & 4/5 \\ 1/5 & 3/5 & 1/5 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \xi(w) \\ \chi(w) \\ 1 \end{pmatrix} > \begin{pmatrix} 0,8 \\ 0,2 \\ 1 \end{pmatrix},$$

в то время как

$$B_0 \vec{h}(z) = \begin{pmatrix} 3/5 & 1/5 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \xi(z) \\ \chi(z) \\ 1 \end{pmatrix} < \begin{pmatrix} 0,8 \\ 0,2 \\ 1 \end{pmatrix}. \quad \square$$

ТЕОРЕМА 3.2. Для всех $x \in [0, 1]$ имеют место неравенства

$$A^{-1}x^\alpha \leq \psi(x) \leq Ax^\alpha, \quad B^{-1}x^\beta \leq \chi(x) \leq Bx^\beta, \quad (3.2.10)$$

где $A = \frac{5}{3}$, $\alpha = \log_2 \frac{5}{3}$, $B = 5$, $\beta = \log_2 5$.

Доказательство. Поскольку $\frac{3}{5} \leq \psi(x) \leq 1$ для $\frac{1}{2} \leq x \leq 1$, мы заключаем из первого равенства (3.2.6), что

$$\left(\frac{3}{5}\right)^{n+1} \leq \psi(x) \leq \left(\frac{3}{5}\right)^n \quad \text{для} \quad \frac{1}{2^{n+1}} \leq x \leq \frac{1}{2^n}.$$

Для данного значения α мы также имеем

$$\left(\frac{3}{5}\right)^{n+1} \leq x^\alpha \leq \left(\frac{3}{5}\right)^n \quad \text{для} \quad \frac{1}{2^{n+1}} \leq x \leq \frac{1}{2^n}.$$

Отсюда следует первое утверждение теоремы. Второе доказывается тем же образом. \square

В качестве следствия теоремы 3 мы получаем такое утверждение.

Следствие. Если u — одна из четырех базисных функций χ, φ, ψ, ξ и если $r = \frac{k}{2^n}$ — любое двоично-рациональное число из отрезка $[0, 1]$, то

$$u'(r) = +\infty \quad (3.2.11)$$

за исключением ровно двух случаев: $\chi'(0) = \xi'(1) = 0$ (см. рис. 3.2).

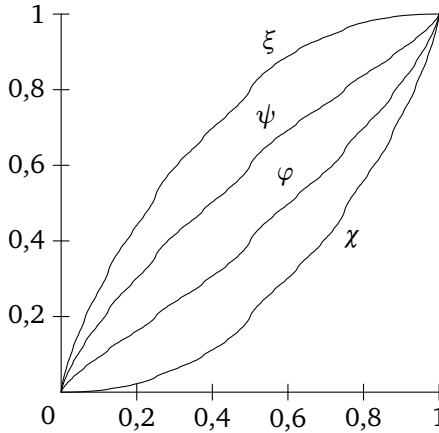


Рис. 3.2. Базисные функции χ, φ, ψ, ξ

Довольно трудно представить себе (и еще труднее нарисовать) график функции, которая во всех двоично-рациональных точках имеет бесконечную производную.

С другой стороны, из теории функций известно, что любая монотонная функция на отрезке $[0, 1]$ имеет почти всюду конечную производную. Как недавно показала студентка второго курса университета

Пенсильвании Таля Шамаш, справедливо равенство $\chi'\left(\frac{1}{3}\right) = 0$. Возможно, ее метод подсчета работает для всех рациональных точек.

Задача 2. Вычислить производную $u'(t)$ во всех случаях, когда это можно сделать (например, во всех рациональных точках).

Следующее интересное свойство функций $u(t)$ — это то, что можно явно вычислить интеграл от этих функций по отрезку $[0, 1]$.

Лемма 3.4.

$$\int_0^1 u_{a,b}^c(t) dt = \frac{3a+3b+c}{7}. \quad (3.2.12)$$

Я оставляю доказательство читателю в качестве нетривиального упражнения на понимание природы гармонических функций (в частности, соотношений (3.2.6), (3.2.7)).

3.3. Продолжение и вычисление функций $\chi(t)$ и $\psi(t)$

Существует метод быстрого подсчета значений функции $\chi(t)$ в двоично-рациональных точках. Он основан на соотношениях:

$$\chi(2t) = 5\chi(t), \quad \chi\left(\frac{1+t}{2}\right) + \chi\left(\frac{1-t}{2}\right) = \frac{2+3\chi(t)}{5}, \quad (3.3.1)$$

которые следуют из (3.2.6), (3.2.7). Еще более простой вывод этих соотношений состоит в проверке совпадения граничных условий для гармонических функций, стоящих в левой и правой частях искомого равенства. Например, для доказательства первого соотношения достаточно сравнить функцию $u_0^{-1}_1$ и ее ограничение на левую нижнюю часть ковра Серпинского с функцией $u_0^{-1/5}_{1/5}$.

Мы можем использовать первое равенство для продолжения функции $\chi(t)$ на всю положительную полуось, полагая

$$\chi(t) := 5^N \chi(2^{-N}|t|), \quad (3.3.2)$$

где N достаточно велико, чтобы $|t| \leq 2^N$.

Геометрически настроенный читатель легко интерпретирует продолженную функцию как граничное значение гармонической функции на «бесконечном ковре Серпинского».

Аналитическим следствием первого равенства (3.3.1) является тот факт, что отношение $R(t) := \frac{\chi(t)}{t^\beta}$, где $\beta = \log_2 5 = 2,3219281\dots$, обладает свойством $R(2t) = R(t)$ и поэтому достаточно знать его, скажем, на отрезке $\left[\frac{1}{2}, 1\right]$ или $[1, 2]$. После некоторых вычислений возникает

Гипотеза. Отношение $R(t)$ достигает максимального значения 1,044... в точке $t_{\max} \approx \frac{8}{15}$ и минимального значения 0,912... в точке $t_{\min} \approx \frac{93}{127}$.

Далее, второе равенство (3.3.1) для $t = \frac{k}{2^n}$ можно переписать в виде

$$\chi(2^n + k) + \chi(2^n - k) - 2\chi(2^n) = 3\chi(k) \quad \text{для } 0 \leq k \leq 2^n. \quad (3.3.3)$$

Напомним операцию второй разностной производной с шагом k для функций целого аргумента:

$$(\Delta_k^2 f)(n) = \frac{f(n+k) - 2f(n) + f(n-k)}{2}.$$

Тогда мы можем написать

$$(\Delta_k^2 \chi)(2^n) = 3\chi(k) \quad \text{для } 0 \leq k \leq 2^n. \quad (3.3.4)$$

Из (3.3.4) легко выводится следующее утверждение.

ТЕОРЕМА 3.3. *Для любого натурального k значение $\chi(k)$ является также натуральным числом. Более того, $\chi(k) \equiv k \pmod{3}$.*

Значение этого наблюдения состоит в том, что мы теперь имеем дело с целочисленной функцией целого аргумента: продолженная функция χ принимает целые значения во всех целых точках. Таким образом мы автоматически попадаем в царство теории чисел.

Аналогичный подход возможен и в случае других основных функций. Например, для продолжения и изучения функции ψ можно использовать равенства:

$$(\Delta_k^2 \psi)(2^n) = -\frac{1}{3}\chi(k) \quad \text{для } 0 \leq k \leq 2^n. \quad (3.3.5)$$

В таблице 3.1 мы указываем значения $\chi(k)$ и значения $\psi(k)$ (умноженные на $3^6 = 729$, чтобы сделать их целыми). Мы также приводим значения первой разностной производной $\Delta\psi(k) := \psi(k) - \psi(k-1)$ для функции $\psi(k)$ и второй разностной производной $\Delta_1^2\chi(k)$ для функции $\chi(k)$.

Заметим, что первая разностная производная $\Delta\psi(k)$ демонстрирует симметрию значений на отрезках $[2^l, 2^{l+1}]$. Аналитическое выражение этой симметрии таково:

$$\psi(3+t) + \psi(3-t) = 2\psi(3) = \frac{40}{3} \quad \text{для } |t| \leq 1. \quad (3.3.6)$$

Таблица 3.1

Значения функций $\chi(k)$, $3^6\psi(k)$ и их разностных производных

k	$\chi(k)$	$\frac{1}{3}\Delta^2\chi$	$3^6\psi(k)$	$3^6\Delta\psi$	k	$\chi(k)$	$\frac{1}{3}\Delta^2\chi$	$3^6\psi(k)$	$3^6\Delta\psi$
1	1	1	729	729	34	3745	-11	9985	245
2	5	1	1215	486	35	3965	5	10191	206
3	12	2	1620	405	36	4200	-2	10400	209
4	25	1	2025	405	37	4429	-11	10597	197
5	41	1	2403	378	38	4625	5	10755	158
6	60	2	2700	297	39	4836	26	10916	161
7	85	5	2997	297	40	5125	1	11125	209
8	125	1	3375	378	41	5417	-23	11331	206
9	168	-2	3744	369	42	5640	-2	11480	149
10	205	1	4005	261	43	5857	17	11617	137
11	245	5	4239	234	44	6125	5	11775	158
12	300	2	4500	261	45	6408	-2	11936	161
13	361	1	4761	261	46	6685	17	12085	149
14	425	5	4995	234	47	7013	53	12255	170
15	504	14	5256	261	48	7500	2	12500	245
16	625	1	5625	369	49	7993	-47	12745	245
17	749	-11	5991	366	50	8345	-11	12915	170
18	840	-2	6240	249	51	8664	14	13064	149
19	925	5	6453	213	52	9025	1	13225	161
20	1025	1	6675	222	53	9389	-11	13383	158
21	1128	-2	6888	213	54	9720	14	13520	137
22	1225	5	7065	177	55	10093	53	13669	149
23	1337	17	7251	186	56	10625	5	13875	206
24	1500	2	7500	249	57	11172	-33	14084	209
25	1669	-11	7749	249	58	11605	1	14245	161
26	1805	1	7935	186	59	12041	41	14403	158
27	1944	14	8112	177	60	12600	14	14600	197
28	2125	5	8325	213	61	13201	1	14809	209
29	2321	1	8547	222	62	13805	41	15015	206
30	2520	14	8760	213	63	14532	122	15260	245
31	2761	41	9009	249	64	15625	1	15625	365
32	3125	1	9375	366	65	16721	-119	$15989\frac{2}{3}$	$364\frac{2}{3}$
33	3492	-38	9740	365	66	17460	-38	$16233\frac{1}{3}$	$243\frac{1}{3}$

Например, полагая $t = \frac{k}{16}$, $0 \leq k \leq 16$, мы получаем:

$$\psi(48+k) + \psi(48-k) = \frac{25000}{729}.$$

Такая же симметрия наблюдается для φ :

$$\varphi\left(\frac{1}{4}+t\right) + \varphi\left(\frac{1}{4}-t\right) = 2\varphi\left(\frac{1}{4}\right) \quad \text{для } |t| \leq \frac{1}{4}. \quad (3.3.7)$$

Все это подводит к задаче о минимальных «волнушках»¹, из аффинных образов которых могут быть составлены графики всех основных функций.

Кандидатами в волнушки являются куски графика функции χ на $\left[\frac{1}{2}, 1\right]$ и графика ψ на $\left[\frac{3}{4}, 1\right]$.

Я горячо рекомендую читателю поискать другие закономерности в этой таблице и доказать соответствующие общие утверждения. Например, посмотрите на значения функции $\Delta\psi$ в точках 2^n , $2^n \pm 1$, $2^n + 2^{n-1}$ и $2^n + 2^{n-1} + 1$.

Очень интересно также исследовать p -адическое поведение функции $\chi(n)$ и возможное продолжение $\chi(n)$ до функции из \mathbb{Q}_2 в \mathbb{Q}_5 .

Наконец, я советую нарисовать график функции $k \mapsto \Delta\psi(k)$ на отрезке $[2^n + 1, 2^{n+1}]$ и подумать, что происходит при $n \rightarrow \infty$.

Схолия D. Производные и интегралы дробного порядка

Производная порядка n определяется обычно как n -я итерация первой производной. Иногда определенный интеграл $\int_0^x f(t) dt$ с переменным верхним пределом называют антипроизводной, или производной порядка -1 . Можно также определить производную порядка $-n$ как n -ю итерацию антипроизводной. Явная форма этой операции такова:

$$f^{(-n)}(x) = \int_0^x dt_1 \int_0^{t_1} dt_2 \dots \int_0^{t_{n-1}} f(t_n) dt_n.$$

Этот повторный интеграл может быть записан в виде n -мерного интеграла

$$\int_{\Delta_x} f(t_n) dt_1 dt_2 \dots dt_n,$$

где Δ_x — симплекс в \mathbb{R}^n с координатами t_1, t_2, \dots, t_n , подчиненными

¹ Перевод В. И. Арнольда термина «wavelets».

неравенствам

$$0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq x.$$

Изменяя порядок интегрирования, можно свести этот интеграл к другому одномерному интегралу

$$\int_{\Delta_x} f(t_n) dt_1 dt_2 \dots dt_n = \int_0^x \text{vol } \Delta_x(t) f(t) dt = \int_0^x \frac{(x-t)^{n-1}}{(n-1)!} f(t) dt. \quad (3.3.1)$$

Здесь $\Delta_x(t)$ — $(n-1)$ -мерный симплекс, который получается пересечением симплекса Δ_x с гиперплоскостью $t_n = t$.

Теперь заметим, что выражение $\frac{(x-t)^{n-1}}{(n-1)!}$ имеет смысл не только для $n \in \mathbb{N}$, но и для любого вещественного n . Поэтому мы заменим обозначение n на α и определим антипроизводную порядка α , или производную порядка $-\alpha$, формулой

$$f^{(-\alpha)}(x) = \int_0^x \frac{(x-t)^{\alpha-1}}{\Gamma(\alpha)} f(t) dt = \int_0^x \frac{t^{\alpha-1}}{\Gamma(\alpha)} f(x-t) dt. \quad (3.3.2)$$

Разумеется, мы должны уточнить, какой класс функций мы будем рассматривать и как понимать интеграл для этого класса функций. Для начала будет достаточно предположить, что наши функции определены и гладки на $(0, \infty)$, а также обращаются в нуль с достаточно большой кратностью в начале координат.

УПРАЖНЕНИЕ 17. Обозначим через $\Phi_\beta(x)$ функцию $\frac{x^{\beta-1}}{\Gamma(\beta)}$. Покажите, что

$$\Phi_\beta^{(-\alpha)}(x) = \Phi_{\beta-\alpha}(x). \quad (3.3.3)$$

Подсказка. Используйте В-функцию Эйлера, задаваемую формулой

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt,$$

и ее свойство:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Отметим связь дробного дифференцирования с операцией свертки на группе \mathbb{R}^+ для функций с носителем на положительной полуоси:

$$(f_1 * f_2)(x) = \int_0^x f_1(t) f_2(x-t) dt.$$

А именно, производная порядка α — это свертка с $\Phi_{-\alpha}$, а интеграл порядка α — это свертка с Φ_α . \diamond

3.4. НЕКОТОРЫЕ АРИФМЕТИЧЕСКИЕ СВОЙСТВА ОСНОВНЫХ ФУНКЦИЙ

Как показано в разделе 3.3, функция $\chi(t)$ принимает целые значения в целых точках. Такие функции часто имеют интересные арифметические свойства.

Остальные основные функции ψ , φ , ξ мы также продолжим на положительную полуось по формулам

$$\psi(2t) = \frac{5}{3}\psi(t), \quad \varphi(t) = \frac{\chi(t) + \psi(t)}{2}, \quad \xi(t) = \frac{3\psi(t) - \chi(t)}{2}. \quad (3.4.1)$$

Можно рассматривать эти функции как граничные значения гармонических функций на бесконечном ковре Серпинского, ограниченном лучами $x \geq 0$, $y = 0$ и $x \geq 0$, $y = \frac{x\sqrt{3}}{2}$.

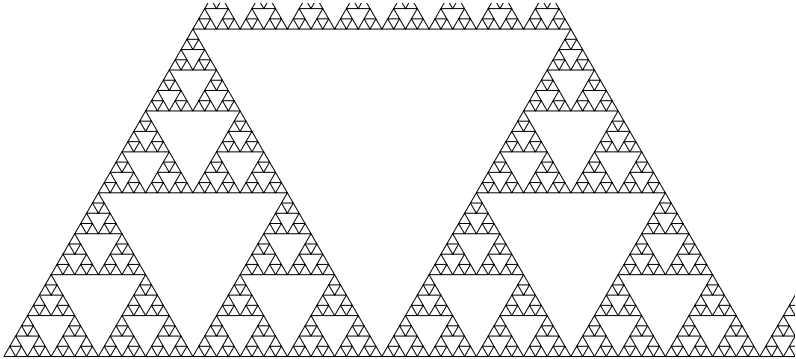


Рис. 3.3. Бесконечный ковер Серпинского

Изучим локальное поведение гармонических функций в окрестности некоторой двоично-рациональной точки $r = \frac{k}{2^n}$.

Мы начнем с функции χ . Ввиду соотношений (3.3.1) достаточно рассмотреть только случай $n = 0$ и нечетные положительные $k = 2m + 1$.

ТЕОРЕМА 3.4. Для любого нечетного положительного k и любого $\tau \in [0, 1]$ мы имеем

$$\chi(k \pm \tau) = \chi(k) + \Delta_2 \cdot \chi(\tau) \pm \Delta_1 \cdot (2\chi(\tau) + 3\psi(\tau)), \quad (3.4.2)$$

где $\Delta_2 = \frac{\chi(k-1) + \chi(k+1) - 2\chi(k)}{2}$, $\Delta_1 = \frac{\chi\left(\frac{k+1}{2}\right) - \chi\left(\frac{k-1}{2}\right)}{2}$.

СЛЕДСТВИЕ. Для любого n , любого k и любого нечетного положительного $l < 2^n$ мы имеем¹

$$\chi(2^n k + l) \equiv \chi(2^n k - l) \pmod{(2\chi(l) + 3^{n+1}\psi(l))} \quad (3.4.3)$$

и

$$\chi(2^n k + l) + \chi(2^n k - l) - 2\chi(2^n k) \equiv 0 \pmod{\chi(l)}. \quad (3.4.4)$$

Некоторые частные случаи полученных сравнений:

а) $n = 1, k = 2m + 1, l = 1: \chi(4m + 3) \equiv \chi(4m + 1) \pmod{11};$

б) $n = 2, k = 2m + 1, l = 3: \chi(8m + 7) \equiv \chi(8m + 1) \pmod{84};$

в) $k = 1: \chi(2^n + l) \equiv \chi(2^n - l) \pmod{(2\chi(l) + 3^{n+1}\psi(l))}$

(на самом деле это не только сравнение, но даже равенство, поскольку в этом случае $2\Delta_1 = 1$).

ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ. Рассмотрим треугольный кусок бесконечного ковра, опирающийся на отрезок $[k - 1, k + 1]$. Он показан на рис. 3.4.

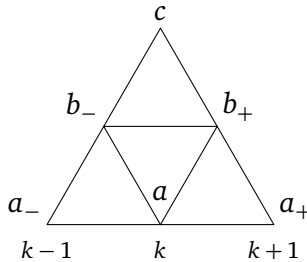


Рис. 3.4. Фрагмент бесконечного ковра Серпинского

Обозначим значения функции χ в точках $k - 1, k, k + 1$ через a_-, a_+ соответственно. Тогда значения b_+, b_-, c , принимаемые в остальных вершинах, как показано на рис. 3.4, могут быть вычислены из равенств:

$$5a = 2a_- + 2a_+ + c, \quad 5b_{\pm} = 2a_{\pm} + 2c + a_{\mp}.$$

В результате этого вычисления получаем:

$$c = 5a - 2a_- - 2a_+, \quad b_+ = 2a - \frac{3a_- + 2a_+}{5}, \quad b_- = 2a - \frac{2a_+ + 3a_-}{5}.$$

¹ Заметим, что $3^{n+1}\psi(l)$ — целое число, когда $l < 2^n$.

Введем в рассмотрение функции $g_{\pm}: \tau \rightarrow \chi(k \pm \tau)$. Зная граничные значения соответствующих гармонических функций на выбранном куске \mathcal{S} , мы можем написать:

$$g_{\pm}(\tau) = a + \frac{a_{\pm} + b_{\pm} - 2a}{2} \cdot \psi(\tau) + \frac{a_{\pm} - b_{\pm}}{2} \cdot \chi(\tau).$$

Для доказательства теоремы остается заметить, что

$$\frac{a_{\pm} + b_{\pm} - 2a}{2} = \pm \frac{3}{10}(a_+ - a_-) = \pm 3 \cdot \Delta_1$$

и

$$\frac{a_{\pm} - b_{\pm}}{2} = \frac{a_- + a_+ - 2a}{2} \pm \frac{1}{5}(a_+ - a_-) = \Delta_2 \pm 2\Delta_1. \quad \square$$

Доказательство следствия. Положим $\tau = \frac{l}{2^n}$ в (3.4.2). Тогда мы получаем

$$\begin{aligned} \chi(2^n k + l) - \chi(2^n k - l) &= 5^n \left(\chi\left(k + \frac{l}{2^n}\right) - \chi\left(k - \frac{l}{2^n}\right) \right) = \\ &= 2 \cdot 5^n \Delta_1 \left(2\chi\left(\frac{l}{2^n}\right) + 3\psi\left(\frac{l}{2^n}\right) \right) = 2 \cdot \Delta_1 \cdot (2\chi(l) + 3^{n+1}\psi(l)). \end{aligned}$$

Поскольку $2\Delta_1 \in \mathbb{Z}$, мы доказали (3.4.3). Сравнение (3.4.4) доказыва-
ется так же. □

В заключение этого раздела мы введем и начнем изучать еще одну целочисленную функцию натурального аргумента. А именно, по-
ложим

$$D(k) := \frac{\chi(k-1) - 2\chi(k) + \chi(k+1)}{3}. \quad (3.4.5)$$

По существу, это вторая разностная производная функции χ , по-
деленная на 3. Значения $D(k)$ для малых k приведены в таблице 3.2.

Эта таблица дает еще больше поводов для наблюдений и откры-
тий, чем таблица 3.1. Например, уже имеющиеся данные позволяют
предположить, что функция D обладает свойством

$$D(2k) = D(k). \quad (3.4.6)$$

Это заметно сокращает ее вычисление: достаточно вычислить зна-
чения в нечетных точках.

Далее, более тонкое наблюдение позволяет решить и эту задачу:
функция D удовлетворяет соотношению

$$D(2k-1) + D(2k+1) = 3D(k). \quad (3.4.7)$$

Таблица 3.2

Значения $D(k)$				
k	$\chi(k)$	$\Delta\chi(k)$	$\Delta^2\chi(k)$	$D(k)$
0	0			
1	1	1	3	1
2	5	4	3	1
3	12	7	6	2
4	25	13	3	1
5	41	16	3	1
6	60	19	6	2
7	85	25	15	5
8	125	40	3	1
9	168	43	-6	-2
10	205	37	3	1
11	245	40	15	5
12	300	55	6	2
13	361	61		

Мы оставляем читателю доказать свойства (3.4.6) и (3.4.7) функции $D(k)$. Пользуясь этими свойствами, легко продолжить таблицу значений $D(k)$ как угодно далеко.

Оказывается, в отличие от функции $\chi(k)$, функция $D(k)$ растет заметно медленнее, примерно как k^γ , $\gamma = \log_2 3$, и большие значения сосредоточены «кластерами» вблизи степеней двойки. Я очень советую читателю поразмышлять и поэкспериментировать на эту тему.

Например, интересно описать прообраз $D^{-1}(n)$ для разных n , начиная с $n = 1$.

3.5. Функции $x(t)$, $y(t)$ и $y(x)$

Теорема 3.4 и приведенное выше ее следствие заставляют задуматься над тем, что, возможно, координата t является не лучшим выбором аргумента для функций $u_{a^c b}$. Более естественный выбор аргумента x и функции $y(x)$ таков:

$$x = \varphi + \psi - 1 = \chi + \xi - 1; \quad y = \xi - \psi = \psi - \varphi = \varphi - \chi. \quad (3.5.1)$$

Когда t меняется от 0 до 1, x возрастает от -1 до 1, в то время как y растёт от значения 0 до максимума $\frac{1}{5}$ при $x = 0$, а потом убывает опять до 0.

Альтернативное (эквивалентное) определение: $x = u_{-1}^0 1$, $y = u_0^1 0$.

ТЕОРЕМА 3.5. *Величина u является дифференцируемой функцией от x . Более точно, производная $y' = \frac{dy}{dx}$ существует и является непрерывной строго убывающей функцией x .*

Мы вернемся к этому вопросу в части II.

Следующие три проблемы открыты.

ЗАДАЧА 3. Подсчитать моменты

$$m_n := \int_{-1}^1 x^n y \, dx. \quad (3.5.2)$$

ЗАДАЧА 4. Подсчитать коэффициенты Фурье

$$c_n := \int_{-1}^1 e^{-\pi i n x} y \, dx. \quad (3.5.3)$$

ЗАДАЧА 5. Исследовать дробные производные основных функций.

Например, интересно найти производную порядка $\alpha = \log_2 5$ от функции $\chi(t)$ и производную порядка $\beta = \log_2 \frac{5}{3}$ от функции $\psi(t)$ в окрестности нуля.

Все базисные функции легко выражаются в терминах x и y :

$$\chi = \frac{x+1-3y}{2}, \quad \varphi = \frac{x+1-y}{2}, \quad \psi = \frac{x+1+y}{2}, \quad \xi = \frac{x+1+3y}{2}. \quad (3.5.4)$$

Другое преимущество величин x и y заключается в их простом поведении относительно оператора T : $Tx = -x$, $Ty = y$.

Недостатком является более сложное поведение относительно операторов A_0 и A_1 . Именно, если мы введем вектор-функцию $\vec{h}(t) = (x(t), y(t), 1)^t$, то законы преобразования базисных функций (3.2.7) переходят в несколько более сложные формулы:

$$\vec{h}\left(\frac{t}{2}\right) = C_0 \vec{h}(t), \quad \vec{h}\left(\frac{1+t}{2}\right) = C_1 \vec{h}(t), \quad (3.5.5)$$

где

$$C_0 = \frac{1}{10} \begin{pmatrix} 5 & 3 & -5 \\ 1 & 3 & 1 \\ 0 & 0 & 10 \end{pmatrix}, \quad C_1 = \frac{1}{10} \begin{pmatrix} 5 & -3 & 5 \\ -1 & 3 & 1 \\ 0 & 0 & 10 \end{pmatrix}. \quad (3.5.6)$$

Обе величины x и y первоначально были функциями от $t \in [0, 1]$. Поскольку x задает биекцию $[0, 1] \rightarrow [-1, 1]$, мы можем рассматривать отображение

$$\tilde{y} := y \circ x^{-1}: [-1, 1] \rightarrow [0, 1].$$

Часто мы не будем делать различия между y и \tilde{y} и писать просто $y(x)$. Утверждение, что x — лучший параметр, чем t , подтверждается сформулированной выше теоремой 3.5. На мой взгляд, лучший способ доказательства этой теоремы — показать, что y является вогнутой функцией от x , то есть обладает свойством

$$y\left(\frac{x_1+x_2}{2}\right) > \frac{y(x_1)+y(x_2)}{2}. \quad (3.5.7)$$

УПРАЖНЕНИЕ 18. Докажите, что $y'(x)$ удовлетворяет уравнениям

$$y'\left(x\left(\frac{t}{2}\right)\right) = \frac{3y'(x(t))+1}{3y'(x(t))+5}, \quad y'\left(x\left(\frac{1+t}{2}\right)\right) = \frac{3y'(x(t))-1}{5-3y'(x(t))}. \quad (3.5.8)$$

Подсказка. Докажите и используйте равенства

$$\begin{aligned} x\left(\frac{t}{2}\right) &= \frac{1}{2}x(t) + \frac{3}{10}y(t) - \frac{1}{2}; & y\left(\frac{t}{2}\right) &= \frac{1}{10}x(t) + \frac{3}{10}y(t) + \frac{1}{10}; \\ x\left(\frac{1+t}{2}\right) &= \frac{1}{2}x(t) - \frac{3}{10}y(t) + \frac{1}{2}; & y\left(\frac{1+t}{2}\right) &= -\frac{1}{10}x(t) + \frac{3}{10}y(t) - \frac{1}{10}, \end{aligned} \quad (3.5.9)$$

которые следуют из (3.5.5) и (3.5.6).

Соотношения (3.5.8) позволяют явно найти производную $y'(x)$ в некоторых точках (если заранее известно, что производная существует).

Например, если мы положим $t=0$ в первом соотношении (3.5.9), мы получим уравнение для производной в точке -1 :

$$y'(-1) = \frac{3y'(-1)+1}{3y'(-1)+5}, \quad \text{или} \quad 3y'(-1)^2 + 2y'(-1) - 1 = 0.$$

Это квадратное уравнение имеет два корня: $\frac{1}{3}$ и -1 . Поскольку $y(-1) = 0$ и $y(-1 + \varepsilon) > 0$, только первый корень дает правильное значение для производной. Таким образом, мы получаем $y'(-1) = \frac{1}{3}$.

Аналогично, полагая $t=1$ во втором соотношении (3.5.9), мы получаем $y'(1) = -\frac{1}{3}$.

Графики функций $y(x)$ и $y'(x)$ показаны на рис. 3.5.

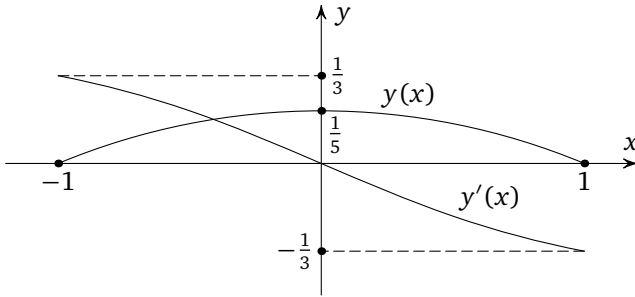


Рис. 3.5. Графики функций $y(x)$ и $y'(x)$

Использованный здесь метод нахождения производной применим для вычисления $y'(x)$ для любой точки x вида $x(t)$ с рациональным t . В самом деле, любое рациональное число r может быть записано в виде смешанно-периодической двоичной дроби. Тогда $r = \frac{k}{2^m(2^n - 1)}$, где n — длина периода двоичной дроби, представляющей r , а m — число цифр до начала периода.

Например, для $\frac{5}{6} = 0,11010101\dots = 0,1(10) = \frac{5}{2(2^2 - 1)}$.

Число $r' = \frac{k}{2^n - 1}$ является неподвижной точкой одного из преобразований вида $\alpha := \alpha_{i_1} \alpha_{i_2} \dots \alpha_{i_n}$ (см. раздел 3.2). А само число r является образом r' относительно другого преобразования $\alpha' := \alpha_{j_1} \alpha_{j_2} \dots \alpha_{j_m}$ того же вида.

Геометрически преобразование α — это сжатие в 2^n раз к точке 2^{-n} . Отсюда следует, что функции $x - x(r')$ и $y - y(r')$ преобразуются линейно с помощью некоторой матрицы формата 2×2 с рациональными коэффициентами. Это дает квадратное уравнение для производной $y'(x)$ в точке $x(r')$. Значение $y'(x(r))$ может быть найдено из (3.5.8).

УПРАЖНЕНИЕ 19. Найдите $x\left(\frac{5}{6}\right)$, $y\left(\frac{5}{6}\right)$ и значение $y'(x)$ в точке $x\left(\frac{5}{6}\right)$.

Следующая проблема открыта.

ЗАДАЧА 6. Пусть $\Gamma \subset \mathbb{R}^2$ — график функции $y(x)$. Он содержит большое множество X точек с рациональными координатами. Например, таковы все точки, соответствующие рациональным значениям параметра t . Очень интересно найти замыкание множества X в p -адической топологии.

3.6. ГАРМОНИЧЕСКИЙ ОБРАЗ КОВРА \mathcal{S}

Заканчивая первую часть книги, я хочу показать, что ковер Серпинского связан с другим замечательным фракталом — ковром Аполлония, который станет главным сюжетом второй части.

Рассмотрим комплексную гармоническую функцию $f_{a,b}^c$ на \mathcal{S} . Она задает отображение \mathcal{S} в комплексную плоскость, при котором три граничные точки переходят в точки a, b, c . Легко видеть, что для разных троек a, b, c , не лежащих на одной прямой, образы \mathcal{S} отличаются лишь аффинным преобразованием. Рассмотрим простейшее из таких отображений, при котором граничные точки остаются на месте, то есть $a = 0, b = 1, c = \omega$. Образ \mathcal{S} в этом случае показан на рис. 3.6. Мы обозначим его $\tilde{\mathcal{S}}$.

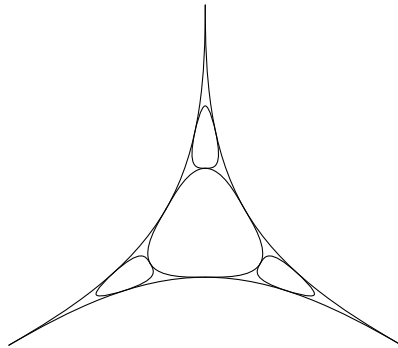


Рис. 3.6. Гармонический образ $\tilde{\mathcal{S}}$ ковра Серпинского

Мы видим, что этот образ похож на кусок так называемого ковра Аполлония — см. часть II. Правда, кривые, составляющие $\tilde{\mathcal{S}}$, не являются окружностями. Например, нижняя граница — это кривая, заданная параметрическими уравнениями

$$x = u_0^{1/2} u_1 = \frac{\varphi(t) + \psi(t)}{2}, \quad y = u_0^{\sqrt{3}/2} u_0 = \sqrt{3} \cdot \frac{\psi(t) - \varphi(t)}{2}. \quad (3.6.1)$$

Вспоминая определение функций $x(t)$ и $y(t)$, мы видим, что эта кривая — аффинный образ графика функции $y(x)$. Как мы видели выше, функция $y(x)$ имеет лишь две производные, так что наша кривая заведомо не является окружностью. Все остальные дуги кривых, образующие $\tilde{\mathcal{S}}$, являются аффинными образами граничной кривой и, следовательно, тоже не являются окружностями.

Тем не менее, $\tilde{\mathcal{S}}$ очень близок к ковру Аполлония \mathcal{A} и, возможно, имеет ту же хаусдорфову размерность. Мы рассмотрим ковер Аполлония более подробно во второй части. Здесь я только сформулирую общую проблему.

Задача 7. Использовать близость фракталов $\tilde{\mathcal{S}}$ и \mathcal{A} , чтобы лучше понять каждый из них.

3.7. МНОГОМЕРНЫЕ АНАЛОГИ КОВРА \mathcal{S}

Ковер Серпинского имеет естественные аналоги в высших размерностях. Это — самоподобные фракталы $\mathcal{S}^{(n)}$ в \mathbb{R}^n , задаваемые системой сжатий

$$f_i(x) = \frac{x + p_i}{2}, \quad (3.7.1)$$

где точки $p_i \in \mathbb{R}^n$, $1 \leq i \leq n + 1$, не лежат в одной гиперплоскости.

Контрольный вопрос: что такое одномерный ковер Серпинского?

Нетрудно вычислить, что n -мерный ковер Серпинского имеет хаусдорфову размерность $\log_2(n + 1)$.

Упражнение 20. Постройте проекцию $(2^n - 1)$ -мерного ковra Серпинского на n -мерную плоскость так, чтобы образом являлся n -мерный выпуклый многогранник и почти все его точки имели единственный прообраз в $\mathcal{S}^{(2^n - 1)}$.

Теория гармонических функций на многомерных коврах Серпинского вполне аналогична теории, описанной выше. Обсудим некото-

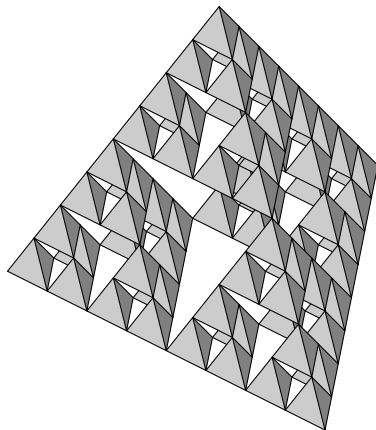


Рис. 3.7. Трехмерный ковер Серпинского

рые факты этой теории. Мы выбираем одно из одномерных ребер первоначального n -симплекса $\{p_1, p_2, \dots, p_{n+1}\}$, скажем $p_1 p_2$, отождествляем его с отрезком $[0, 1]$ и ограничиваем все гармонические функции на это ребро.

ЛЕММА 3.5. *Ограничение гармонической функции f на ребро $p_1 p_2$ зависит только от значений $f(p_1), f(p_2)$ и от суммы $\sum_{k=3}^{n+1} f(p_k)$.*

Подсказка. Используйте инвариантность ограничения относительно группы перестановок вершин p_3, \dots, p_{n+1} .

СЛЕДСТВИЕ. *Ограничения гармонических функций на $\mathcal{S}^{(n)}$ на любое ребро $p_i p_j$ образуют 3-мерное векторное пространство.*

Пусть $f_{a,b}^c$ означает любую гармоническую функцию на $\mathcal{S}^{(n)}$ с граничными значениями $f(p_1) = a$, $f(p_2) = b$ и $\sum_{k=3}^{n+1} f(p_k) = c$. Тогда ограничение этой функции на отрезок $[p_1, p_2] \simeq [0, 1]$ однозначно определено числами a, b, c . Мы обозначим это ограничение $u_{a,b}^c(t)$.

Определим базисные функции формулами

$$\chi(t) = u_0^{-1}{}_1(t), \quad \varphi(t) = u_0^0{}_1(t), \quad \psi(t) = u_0^{n-1}{}_1(t), \quad \xi(t) = u_0^n{}_1(t), \quad (3.7.2)$$

а также функции x, y соотношениями

$$x(t) = u_{-1}^0{}_1(t), \quad y(t) = u_0^1{}_0(t). \quad (3.7.3)$$

Тогда

$$x = \chi + \xi - 1 = \varphi + \psi - 1, \quad y = \varphi - \chi = \xi - \psi = \frac{\psi - \varphi}{n-1}.$$

Отметим, в частности, что $u_1^{n-1}{}_1(t) \equiv 1$.

Основные соотношения между базисными функциями выводятся как в двумерном случае. Они имеют вид

$$\chi(2t) = (n+3) \cdot \chi(t), \quad \psi(2t) = \frac{n+3}{n+1} \cdot \psi(t); \quad (3.7.4)$$

$$\begin{aligned} \chi(1+\tau) + \chi(1-\tau) &= 2 + (n+1)\chi(\tau), \\ \chi(1+\tau) - \chi(1-\tau) &= 2\frac{n+1}{n}\psi(\tau) + \frac{(n-1)(n+2)}{n}\chi(\tau); \end{aligned} \quad (3.7.5)$$

$$\begin{aligned} \psi(1+\tau) + \psi(1-\tau) &= 2 - \frac{n-1}{n+1}\chi(\tau), \\ \psi(1+\tau) - \psi(1-\tau) &= \frac{2}{n}\psi(\tau) + \frac{(n-1)(n+2)}{n(n+1)}\chi(\tau). \end{aligned} \quad (3.7.6)$$

Эти соотношения позволяют развить арифметическую теорию базисных функций для любого¹ целочисленного значения n аналогично случаю $n = 2$.

В частности, функция $\chi(t)$ всегда принимает целые значения в целых точках.

Некоторые значения n представляют особый интерес.

При $n = 1$ мы получаем $\chi(t) = t^2$, $\varphi(t) = \psi(t) = t$, $\xi(t) = 2t - t^2$.

При $n = 0$ мы получаем $y = 0$, поэтому $\chi(t) = \varphi(t) = \psi(t) = \xi(t)$ и эта функция удовлетворяет равенствам

$$\chi(2t) = 3\chi(t), \quad \chi(2^m + k) + \chi(2^m - k) = 2 \cdot 3^m + \chi(k). \quad (3.7.7)$$

Здесь, как и в двумерном случае, полезно ввести новую функцию D по формуле

$$D(k) := \chi(k+1) - 2\chi(k) + \chi(k-1) \quad \text{для любого целого } k > 0. \quad (3.7.8)$$

ТЕОРЕМА 3.6. Функция $D(k)$ обладает свойствами:

$$D(2k) = D(k), \quad D(2^n + k) + D(2^n - k) = D(k) \quad \text{для } 0 < k < 2^n. \quad (3.7.9)$$

Детальное изучение функции $D(k)$ очень интересно и может быть предметом самостоятельного исследования.

При $n = -1$ мы получаем $\chi(t) = t$ и неясно, как определить остальные базисные функции.

Наконец, при $n = -2$ мы получаем $\chi(k) = \begin{cases} 1, & \text{если } k \not\equiv 0 \pmod{3}, \\ 0, & \text{если } k \equiv 0 \pmod{3}. \end{cases}$

Аналогичные формулы справедливы для остальных базисных функций.

Мы оставляем читателю исследовать остальные значения n и поискать интересные примеры.

Схолия Е. Числовые системы

Е.1. Стандартные системы. Большинство вещественных чисел иррациональны, то есть не могут быть записаны как отношение двух целых чисел. Более того, вещественные числа образуют несчетное множество и поэтому не могут быть занумерованы *словами*, содержащими конечное число букв из конечного или счетного алфавита.

¹ Я не знаю геометрической интерпретации этих функций для $n \leq 0$.

С другой стороны, есть много числовых систем, позволяющих записывать все вещественные числа, используя бесконечные (счетные) слова с буквами из конечного или счетного алфавита. Самые известные системы — это обычная десятичная система и ее двоичный аналог.

Для построения системы S такого типа нужны следующие начальные данные:

- вещественное или комплексное число b , $|b| > 1$, называемое базой;
- множество $D = \{d_1, d_2, \dots\}$ используемых цифр, которые могут быть любыми вещественными или комплексными числами, но обычно выбираются из натурального ряда, включая 0.

Каждой полубесконечной последовательности вида

$$a = a_n a_{n-1} \dots a_1 a_0, a_{-1} a_{-2} \dots a_{-n} \dots, \quad a_k \in \mathbb{Z}_+,$$

числовая система S ставит в соответствие число

$$\text{val}(a) = \sum_{-\infty}^{k=n} d_{a_k} \cdot b^k. \quad (3.7.1)$$

В стандартных числовых системах базой является натуральное число m , а цифрами — элементы множества $X_m = \{0, 1, \dots, m-1\}$. Хорошо известно, что любое неотрицательное вещественное число x записывается в форме

$$x = \text{val}(a) = \sum_{-\infty}^{j=n} a_j \cdot m^j. \quad (3.7.2)$$

Более точно, каждое неотрицательное целое число записывается в виде (3.7.2) однозначно и имеет дополнительное свойство: $a_k = 0$ для $k < 0$.

В то же время каждое вещественное число из отрезка $[0, 1]$ почти однозначно записывается в виде (3.7.2) и имеет дополнительное свойство: $a_k = 0$ для $k \geq 0$. Неединственность возникает из тождества

$$\sum_{k \geq 1} (m-1) \cdot m^{-k} = 1.$$

Общепринятый способ избежать этой неединственности — запретить использование бесконечной последовательности цифр $m-1$.

Руководствуясь этим примером, мы определим для любой числовой системы S такого типа множество $\mathbb{C}(S)$ целых чисел, которые обладают свойством $a_k = 0$ для $k < 0$, и множество $\mathbb{D}(S)$ дробных чисел, которые обладают свойством $a_k = 0$ для $k \geq 0$.

Для стандартных систем S мы имеем $\mathbb{C}(S) = \mathbb{Z}_+$, $\mathbb{D}(S) = [0, 1)$.

Е.2. Нестандартные системы. Нестандартные системы используются редко, но их свойства более интересны.

УПРАЖНЕНИЕ 21. Рассмотрим числовую систему S с базой $b = -2$ и цифрами $\{0, 1\}$.

Проверьте, что для этой системы $\mathcal{C}(S) = \mathbb{Z}$, а $\mathcal{D}(S) = \left[-\frac{2}{3}, \frac{1}{3}\right]$. Покажите, что любое вещественное число — положительное или отрицательное — записывается в виде (3.7.2) почти однозначно.

УПРАЖНЕНИЕ 22. Рассмотрим систему S с базой $b = 1 + i$ и цифрами $\{0, 1\}$.

Покажите, что $\mathcal{C}(S) = \mathbb{Z}[i]$ — множество целых гауссовых чисел вида $a + ib$, $a, b \in \mathbb{Z}$.

Что касается $\mathcal{D}(S)$, то это — компактное фрактальное множество размерности 2, определяемое свойством:

$$\mathcal{D} = \frac{1-i}{2} \cdot (\mathcal{D} \cup (1 + \mathcal{D})). \quad (3.7.3)$$

Здесь, как обычно, когда арифметические операции применяются к некоторому множеству чисел, это значит, что они применяются к каждому элементу множества. Изображение этого множества можно найти во многих книгах о фракталах (см., например, [Edg]) или построить самому, пользуясь равенством (3.7.3). Это множество показано на рис. 3.8.

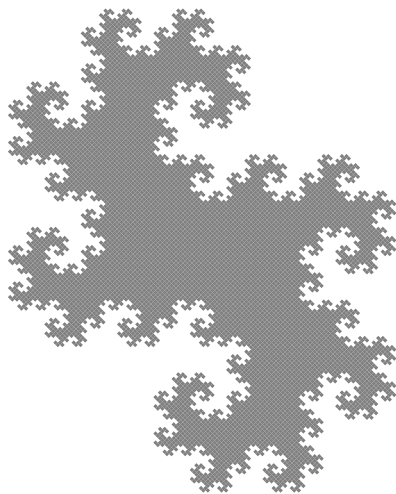


Рис. 3.8. Множество $\mathcal{D}(S)$

УПРАЖНЕНИЕ 23. $\varepsilon = e^{2\pi i/3}$ — кубический корень из 1. Существует ли нестандартная числовая система с базой и цифрами из $\mathbb{Z}[\varepsilon]$, для которой $\mathcal{C}(S) = \mathbb{Z}[\varepsilon]$? Каково множество $D(S)$ для такой системы?

Е.3. *Непрерывные дроби.* Существует еще более нестандартная числовая система, связанная с понятием *непрерывной дроби*. Пусть $k = \{k_1, k_2, \dots\}$ — конечная или счетная последовательность натуральных чисел. Мы поставим в соответствие этой последовательности число

$$\text{val}(k) = \frac{1}{k_1 + \frac{1}{k_2 + \frac{1}{k_3 + \dots + \frac{1}{k_n}}}}, \quad (3.7.4)$$

если последовательность k конечна, или предел выражения (3.7.4) при $n \rightarrow \infty$, если последовательность k бесконечна.

Хорошо известно, что предел в данном случае всегда существует. Более того, любое иррациональное число x из интервала $(0, 1)$ соответствует единственной бесконечной последовательности k . Говорят, что x является значением бесконечной непрерывной дроби, задаваемой последовательностью k .

Что касается рациональных чисел $r \in (0, 1)$, то каждое из них является значением ровно двух конечных непрерывных дробей:

$$k = \{k_1, \dots, k_{n-1}, 1\} \quad \text{и} \quad k' = \{k_1, \dots, k_{n-1} + 1\}.$$

Существует простой алгоритм, как восстановить последовательность k по значению $\text{val}(k)$. Именно, пусть $[x]$ означает так называемую *целую часть* числа x , то есть наибольшее целое число, не превосходящее x . Через $\{x\}$ обозначается *дробная часть* x , равная $x - [x]$. Для каждого числа $x \in (0, 1)$ мы определяем последовательно

$$x_1 = \frac{1}{x}, \quad k_1 = [x_1]; \quad x_2 = \frac{1}{\{x_1\}}, \quad k_2 = [x_2], \quad \dots,$$

$$x_n = \frac{1}{\{x_{n-1}\}}, \quad k_n = [x_n], \quad \dots$$

Для рационального x этот процесс останавливается, когда для некоторого n мы получаем $\{x_{n+1}\} = 0$. Тогда непрерывная дробь, соответствующая последовательности $k = \{k_1, \dots, k_n\}$, имеет значение x .

Для иррационального x процесс продолжается бесконечно и определяет бесконечную непрерывную дробь со значением x .

ПРИМЕР. Пусть $k_n = 2$ для всех n . Тогда $x = \text{val}(k)$, очевидно, удовлетворяет уравнению $\frac{1}{x} = 2 + x$. Поэтому $x^2 + 2x - 1 = 0$ и $x = -1 \pm \sqrt{2}$. Поскольку $x \in (0, 1)$, мы заключаем, что $x = \sqrt{2} - 1$.

Таким образом, квадратный корень из 2 задается выражением

$$\sqrt{2} = 1 + \frac{1}{2 + \frac{1}{2 + \frac{1}{2 + \frac{1}{2 + \dots}}}}$$

следовательно, не является рациональным числом.

Этот результат¹ был известен еще во времена Пифагора и держался в секрете, поскольку он мог подорвать веру в могущество (рациональных) чисел.

Есть несколько случаев, когда значение бесконечной непрерывной дроби выражается через известные функции одной переменной. Я знаю два таких случая.

1. Первый случай: исходная дробь k периодична или хотя бы смешанно-периодична.

В этом случае число $\text{val}(k)$ удовлетворяет квадратному уравнению с целыми коэффициентами и может быть вычислено явно.

Обратное также верно: любой вещественный корень квадратного уравнения с целыми коэффициентами² может быть записан в виде смешанно периодической непрерывной дроби.

2. Второй случай: последовательность $\{k_n\}$ является арифметической прогрессией или ее модификацией. Здесь я ограничусь тремя примерами:

$$\text{th } 1 = \frac{e^2 - 1}{e^2 + 1} = \frac{1}{1 + \frac{1}{3 + \frac{1}{5 + \frac{1}{7 + \frac{1}{9 + \dots}}}}};$$

¹ Более точно, его геометрическая интерпретация, показывающая, что диагональ квадрата несоизмерима с его стороной.

² Все такие числа имеют вид $\frac{a + \sqrt{b}}{c}$, $a, b, c \in \mathbb{Z}$, и называются *квадратичными иррациональностями*.

$$\operatorname{th} \frac{1}{2} = \frac{e-1}{e+1} = \frac{1}{2 + \frac{1}{6 + \frac{1}{10 + \frac{1}{14 + \frac{1}{18 + \dots}}}}};$$

$$e = 2 + \frac{1}{1 + \frac{1}{2 + \frac{1}{1 + \frac{1}{1 + \frac{1}{4 + \frac{1}{1 + \frac{1}{1 + \frac{1}{6 + \dots}}}}}}}}.$$

Е.4. Обобщенные числовые системы. Оказывается, все числовые системы, рассмотренные нами до сих пор, являются частными случаями следующей общей схемы. Выберем множество $D \subset \mathbb{Z}$ возможных цифр. Каждой цифре $d \in D$ соответствует вещественная или комплексная матрица A_d формата $n \times n$. Выберем также $(n \times n)$ -матрицу Ω .

Каждой полубесконечной последовательности цифр $a = \{a_0, a_1, \dots\}$ мы поставим в соответствие число

$$\operatorname{val}(a) = \operatorname{tr}(\Omega \cdot A_{a_0} \cdot A_{a_1} \cdot \dots) \quad (3.7.5)$$

в случае, когда бесконечное произведение имеет смысл.

Покажем, как эта конструкция связана с предыдущими числовыми системами.

Пусть $A_a = \begin{pmatrix} m & 0 \\ a & 1 \end{pmatrix}$, $0 \leq a \leq m-1$. Тогда

$$A_{a_n} \dots A_{a_1} A_{a_0} = \begin{pmatrix} m^{n+1} & 0 \\ \sum_{j=0}^n a_j m^j & 1 \end{pmatrix}.$$

Таким образом, если мы положим $\Omega = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, мы получим:

$$\begin{aligned} \operatorname{val}(a_0, a_1, \dots, a_n) &= \operatorname{tr}(\Omega \cdot A_{a_0} \dots A_{a_n}) = \operatorname{tr}(A_{a_0} \cdot A_{a_1} \dots A_{a_n} \cdot \Omega) = \\ &= a_0 + a_1 m + \dots + a_n m^n. \end{aligned} \quad (3.7.6)$$

Пусть теперь $A_k = \begin{pmatrix} k & 1 \\ 1 & 0 \end{pmatrix}$. Рассмотрим матрицы

$$A_k = \begin{pmatrix} k & 1 \\ 1 & 0 \end{pmatrix}, \quad A_k A_l = \begin{pmatrix} kl+1 & k \\ l & 1 \end{pmatrix}, \quad A_k A_l A_m = \begin{pmatrix} klm+m+k & kl+1 \\ lm+1 & l \end{pmatrix}$$

и сравним их с непрерывными дробями:

$$\frac{1}{k}, \quad \frac{1}{k + \frac{1}{l}} = \frac{l}{kl + 1}, \quad \frac{1}{k + \frac{1}{l + \frac{1}{m}}} = \frac{lm + 1}{klm + m + k}.$$

Это сравнение подсказывает общее тождество:

ЛЕММА Е.6. Значение непрерывной дроби $k = \{k_1, \dots, k_n\}$ можно вычислить по формуле:

$$\text{val}(k) = \frac{1}{k_1 + \frac{1}{k_2 + \frac{1}{k_3 + \dots + \frac{1}{k_n}}}} = \frac{\text{tr}(\Omega \cdot A_{k_1} \cdot A_{k_2} \dots A_{k_n})}{\text{tr}(\Omega_0 \cdot A_{k_1} \cdot A_{k_2} \dots A_{k_n})}, \quad (3.7.7)$$

где матрица Ω определена выше и $\Omega_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

3.8. ПРИМЕНЕНИЯ ОБОБЩЕННЫХ ЧИСЛОВЫХ СИСТЕМ

Применение к ковру Серпинского. Мы начнем с параметризации точек \mathcal{S} , несколько отличной от той, которую мы ввели в начале этой главы.

Рассмотрим алфавит из трех букв: $-1, 0, 1$. Каждому конечному слову $a = a_1 a_2 \dots a_n$ мы поставим в соответствие комплексное число

$$\text{val}(a) = \frac{\varepsilon^{a_1}}{2} + \frac{\varepsilon^{a_2}}{4} + \dots + \frac{\varepsilon^{a_n}}{2^n}, \quad \text{где } \varepsilon = e^{2\pi i/3}. \quad (3.8.1)$$

Пустой последовательности соответствует число 0.

Подумаем, какое множество заполняют значения $\text{val}(a)$, когда a пробегает множество всех бесконечных последовательностей. Пусть сначала все члены последовательности a одинаковы и равны m . Тогда значения $\text{val}(a) = \varepsilon^m$ — это три кубических корня из 1. Если последовательность a содержит только цифры 1 и -1 , то вещественная часть ряда (3.8.1) имеет вид $\sum_{k=1}^{\infty} -\frac{1}{2^{k+1}} = \frac{1}{2}$. Значит, величины $\text{val}(a)$ в этом случае лежат на прямой $x = \frac{1}{2}$. Более внимательное рассмотрение показывает, что значения $\text{val}(a)$ пробегают отрезок, соединяющий кубические корни ε и $\varepsilon^{-1} = \bar{\varepsilon}$. Соображения симметрии (относительно умножения на ε) показывают, что наше множество лежит внутри выпуклой оболочки точек 1, ε и $\bar{\varepsilon}$. Я оставляю читателю доказать, что

на самом деле искомое множество — это ковер Серпинского \mathcal{S} с вершинами в трех кубических корнях из 1.

УПРАЖНЕНИЕ 24. Покажите, что значения $\text{val}(a)$ для 3^n последовательностей длины n расположены в центрах треугольников ранга $n - 1$, дополнительных к \mathcal{S} .

УПРАЖНЕНИЕ 25. Покажите, что значения $\text{val}(a)$ и $\text{val}(a')$ для бесконечных последовательностей a и a' совпадают, только если a' получается из a заменой хвоста вида $xууу\dots$ на хвост $уххх\dots$.

УПРАЖНЕНИЕ 26. Какие бесконечные последовательности a соответствуют: а) граничным точкам; б) точкам ребер, соединяющих граничные точки; в) вершинам \mathcal{S}_n ; г) точкам ребер, соединяющих вершины \mathcal{S}_n ?

Теперь поставим следующую основную задачу:

Вычислить значение гармонической функции на \mathcal{S} с данными граничными значениями b_{-1}, b_0, b_1 в точке, соответствующей данной бесконечной последовательности a .

Ответ дается в следующей форме:

Существует обобщенная числовая система S , заданная тремя матрицами A_{-1}, A_0, A_1 и матрицей $\Omega = \Omega(b_{-1}, b_0, b_1)$ формата 3×3 , такая, что искомое значение равно $\text{val}(a)$ в системе S .

Применение к функции «вопросительный знак». Так называемая функция «вопросительный знак» была определена Минковским в 1904 году для нумерации всех квадратичных иррациональностей на интервале $(0, 1)$ с помощью рациональных чисел на том же интервале с сохранением непрерывности и порядка. Позже, в 1938 году, А. Данжуа дал определение этой функции на всей вещественной прямой.

Согласно одному из возможных определений (см. ниже), функция $?(\cdot)$ переводит число a , заданное бесконечной непрерывной дробью

$$a = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots + \frac{1}{a_k} + \dots}}},$$

в число, записываемое в двоичной системе как

$$?(a) := \sum_k \frac{(-1)^{k-1}}{2^{a_1 + \dots + a_k - 1}} = \overbrace{0,0\dots 0}^{a_1} \overbrace{1\dots 1}^{a_2} \overbrace{0\dots 0}^{a_3} \dots$$

Например, $\varphi\left(\frac{\sqrt{2}}{2}\right) = 0,11001100\dots = \frac{4}{5}$, $\varphi\left(\frac{e^2-1}{e^2+1}\right) = \sum_{k \geq 0} 2^{-k^2}$.

Мы сможем сказать больше об этой функции во второй части книги. Здесь мы только заметим, что это еще одна функция, которая наиболее просто выглядит при использовании обобщенных числовых систем.

Часть II



КОВЕР АПОЛЛОНИЯ

В этой части книги мы рассмотрим другой замечательный фрактал — так называемый ковер Аполлония \mathcal{A} . Он выглядит совсем по-другому, чем ковер Серпинского \mathcal{S} . Например, \mathcal{A} не является самоподобным фракталом, хотя для любого $k \geq 0$ он может быть представлен как объединение $3k + 2$ подмножеств, гомеоморфных \mathcal{S} .

Тем не менее, существует глубокая и красивая связь между обоими фракталами, и наша цель, только частично достигнутая здесь, состоит в том, чтобы понять и объяснить эту связь.

Многие факты, обсуждаемые ниже, относятся к элементарной геометрии. Однако в современном математическом образовании евклидова геометрия занимает неоправданно малое место, так что мы не можем полагаться на информацию, полученную в школе. Поэтому иногда нам приходится использовать более сложные средства для вывода нужных нам результатов.

Как и в первой части, мы изучаем фрактал с разных точек зрения: геометрической, теоретико-групповой и теоретико-числовой. Именно взаимодействие этих подходов делает нашу задачу очень интересной и многообещающей.

КРУГИ НА СФЕРАХ

4.1. ТЕОРЕМА ДЕКАРТА

Мы начнем с простого геометрического вопроса:

Описать все конфигурации четырех попарно касающихся окружностей на плоскости.

Примеры таких конфигураций показаны ниже на рис. 4.1. Мы включили сюда и те случаи, когда одна или две из наших окружностей вырождаются в прямую линию (которую можно рассматривать как окружность бесконечного радиуса), а также случаи, когда все окружности касаются одной изнутри. Впоследствии мы покажем, что в этом случае удобно рассматривать внешнюю окружность как окружность с отрицательным радиусом.

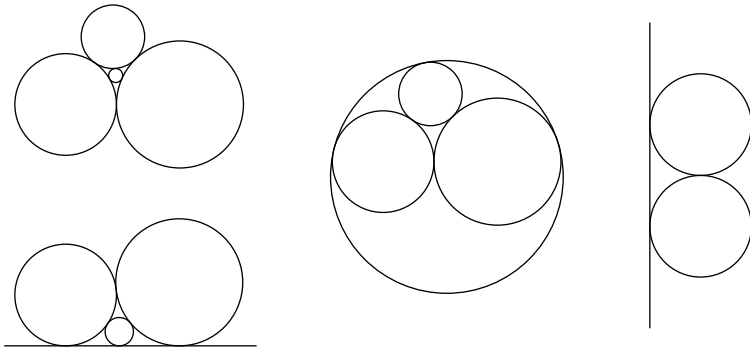


Рис. 4.1. Четверки касающихся окружностей

Наряду с этим существуют конфигурации, которые нам хотелось бы исключить, поскольку без них теория становится более простой и стройной. Несколько таких нежелательных конфигураций показано на рис. 4.2. Во всех этих конфигурациях все четыре окружности имеют общую точку, конечную или бесконечно удаленную. Причина, по которой нужно исключить эти конфигурации, станет яснее, когда мы точнее сформулируем задачу и перейдем от окружностей к кругам.

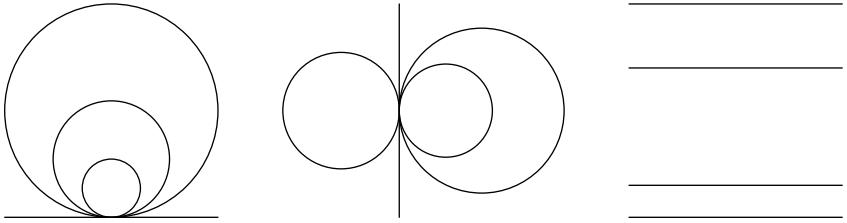


Рис. 4.2. «Неправильные» четверки

Оказывается, что полное и ясное решение нашей задачи использует сведения из разных областей математики. Более того, задача имеет естественное многомерное обобщение и требует более точной формулировки.

Мы начнем с того, что на вполне элементарном уровне покажем необходимость изменения и уточнения формулировки. Для этого сделаем шаг назад и рассмотрим вместо четверки тройку попарно касающихся окружностей. Имеются три типа таких троек — см. ниже рис. 4.3. Эти три типа различаются по виду треугольника, образуемого точками касания. А именно, этот треугольник остроугольный в случае а), прямоугольный в случае б) и тупоугольный в случае в).

В случае а) довольно ясно, что наши три окружности могут иметь произвольные положительные радиусы r_1, r_2, r_3 . В самом деле, пусть O_1, O_2, O_3 — центры искомых окружностей. Мы всегда можем построить треугольник $O_1O_2O_3$, поскольку длины его сторон известны: $|O_iO_j| = r_i + r_j$ и удовлетворяют неравенству треугольника:

$$|O_iO_j| + |O_jO_k| = (r_i + r_j) + (r_j + r_k) > r_i + r_k = |O_iO_k|. \quad (4.1.1)$$

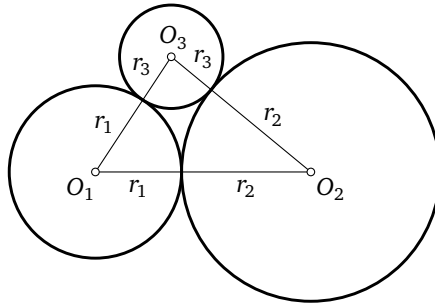
В случае в) мы имеем $|O_1O_2| = r_1 + r_2$, $|O_2O_3| = r_3 - r_2$, $|O_3O_1| = r_3 - r_1$ и $r_1 + r_2 < r_3$.

Оказывается, если заменить в наших формулах r_3 на $-r_3$, то можно объединить случаи а) и в) в одной общей формуле:

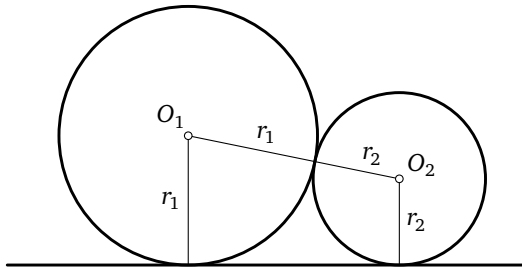
$$|O_iO_j| = |r_i + r_j|. \quad (4.1.2)$$

В случае б) центр O_3 расположен в бесконечно удаленной точке. Мы полагаем в этом случае $r_3 = \infty$, и равенство (4.1.2) при естественной его интерпретации также выполняется.

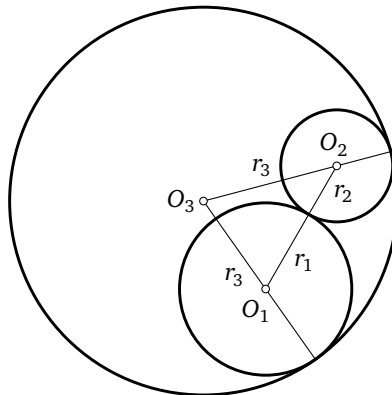
В случае, когда четыре окружности попарно касаются, их радиусы не произвольны, а удовлетворяют некоторому уравнению. Это



Тройка касающихся окружностей типа а)



Тройка касающихся окружностей типа б)



Тройка касающихся окружностей типа в)

Рис. 4.3.

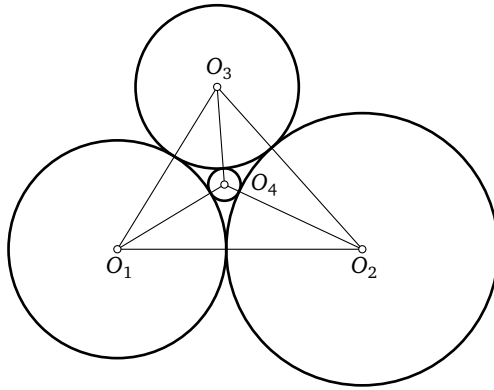


Рис. 4.4. К доказательству формулы Декарта

уравнение и/или некоторые его следствия были известны уже математикам Древней Греции более чем две тысячи лет тому назад.

В более близкое для нас время это уравнение было явно выписано Рене Декартом, знаменитым французским математиком и философом первой половины XVII столетия.

Если вместо радиусов r_i рассматривать обратные величины

$$c_i := r_i^{-1}, \quad 1 \leq i \leq 4,$$

то уравнение Декарта выглядит совсем просто:

$$(c_1 + c_2 + c_3 + c_4)^2 - 2(c_1^2 + c_2^2 + c_3^2 + c_4^2) = 0. \quad (4.1.3)$$

Геометрический смысл величины c_i — это кривизна окружности¹ с радиусом r_i .

Мы оставляем любителям геометрии дать доказательство формулы Декарта, используя сведения, еще недавно входившие в обычную школьную программу. Следующее упражнение, а также рис. 4.4 будут для этого полезными.

УПРАЖНЕНИЕ 27. Найдите общую формулу, пригодную для вычисления площади треугольника $O_1O_2O_3$, введенного выше, в случаях а) и в).

Подсказка. Используйте формулу Герона

$$S = \sqrt{p(p-a)(p-b)(p-c)}.$$

¹ Причина, по которой кривизны ведут себя проще, чем радиусы, будет объяснена позже, когда мы введем теоретико-групповой подход к нашей задаче.

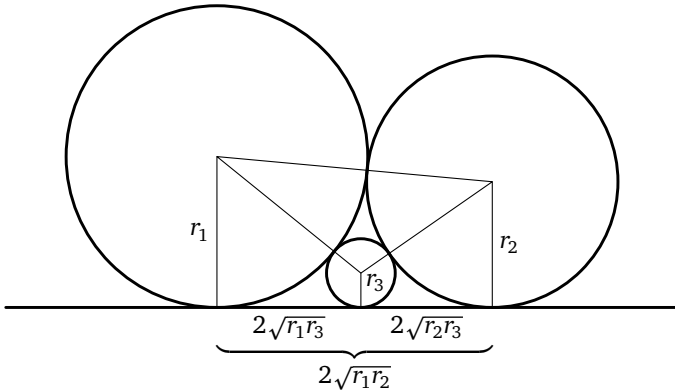


Рис. 4.5. Вырожденное уравнение Декарта

Ответ. $S = \sqrt{r_1 r_2 r_3 (r_1 + r_2 + r_3)}$. Заметим, что выражение под знаком корня всегда положительно.

Существует частный случай теоремы Декарта, который гораздо легче доказать, чем общий случай. А именно, предположим, что одна из четырех окружностей вырождается в прямую линию. Пусть, например, $c_4 = 0$, так что зависимость между оставшимися кривизнами имеет вид

$$(c_1 + c_2 + c_3)^2 - 2(c_1^2 + c_2^2 + c_3^2) = 0. \quad (4.1.4)$$

По счастью, левая часть (4.1.4) разлагается на простые множители, которые легко выражаются через кривизны. Чтобы увидеть это, рассмотрим левую часть как квадратичный многочлен от c_1 :

$$-c_1^2 + 2c_1(c_2 + c_3) - c_2^2 + 2c_2c_3 - c_3^2.$$

Этот многочлен имеет корни

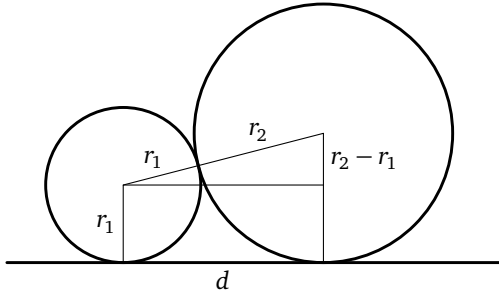
$$c_1 = c_2 + c_3 \pm 2\sqrt{c_2c_3} = (\sqrt{c_2} \pm \sqrt{c_3})^2.$$

Поэтому он может быть записан как

$$-(c_1 - (\sqrt{c_2} + \sqrt{c_3})^2)(c_1 - (\sqrt{c_2} - \sqrt{c_3})^2) = (\sqrt{c_1} + \sqrt{c_2} + \sqrt{c_3}) \times \\ \times (-\sqrt{c_1} + \sqrt{c_2} + \sqrt{c_3})(\sqrt{c_1} - \sqrt{c_2} + \sqrt{c_3})(\sqrt{c_1} + \sqrt{c_2} - \sqrt{c_3}).$$

Отсюда следует, что (4.1.4) справедливо тогда и только тогда, когда при подходящем выборе знаков выполнены следующие равенства:

$$\sqrt{c_1} \pm \sqrt{c_2} \pm \sqrt{c_3} = 0, \quad \text{или} \quad \sqrt{r_2 r_3} \pm \sqrt{r_1 r_2} \pm \sqrt{r_1 r_3} = 0. \quad (4.1.5)$$

Рис. 4.6. Вырожденная тройка с $d = 2\sqrt{r_1 r_2}$

Правильный выбор знаков зависит от относительных размеров радиусов. Если, например, мы занумеруем радиусы так, что $r_1 \geq r_2 \geq r_3 > 0$, то правильное равенство имеет вид $\sqrt{r_1 r_2} = \sqrt{r_2 r_3} + \sqrt{r_3 r_1}$. Вы можете сами доказать это равенство, используя рис. 4.5 и 4.6.

В следующем разделе мы докажем более общее утверждение, используя матричную алгебру и геометрию пространства Минковского. Но перед этим мы должны исправить неаккуратности в нашем первоначальном изложении и дать точную постановку задачи.

До сих пор мы не учитывали знак кривизны, а это может привести к ошибочному истолкованию формулы Декарта. Рассмотрим, например, четыре окружности, показанные ниже на рис. 4.7.

Если мы здесь положим $c_1 = 1$, $c_2 = c_3 = 2$, $c_4 = 3$, то придем к неверному равенству

$$64 = (1 + 2 + 2 + 3)^2 = 2(1 + 4 + 4 + 9) = 36.$$

Но если мы будем считать кривизну внешней окружности равной -1 , то получим верное равенство

$$36 = (-1 + 2 + 2 + 3)^2 = 2(1 + 4 + 4 + 9) = 36.$$

Еще раз взглянув на рисунок 4.7, мы видим, что внешняя окружность находится в особом положении: остальные окружности касаются ее изнутри. А мы уже говорили, что в этом случае удобно считать кривизну отрицательной величиной.

Чтобы сделать изложение строгим, есть две возможности: либо ввести ориентацию окружностей, либо рассматривать вместо окружностей двумерные круги, ограниченные этими окружностями. Оба подхода на самом деле эквивалентны. В самом деле, каждый круг наследует ориентацию у двумерной плоскости (или сферы), на которой

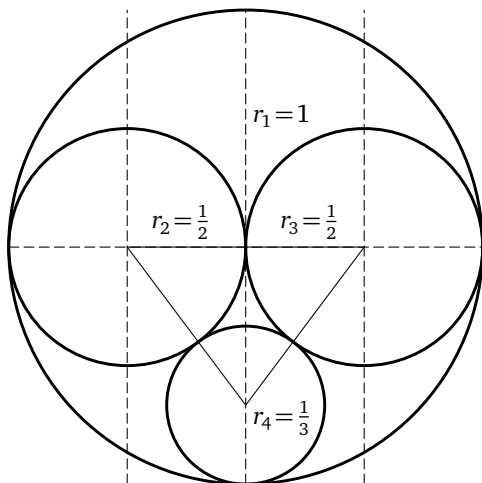


Рис. 4.7. «Нарушение» уравнения Декарта

он находится, а граница ориентированного круга сама имеет выделенную ориентацию. В нашем случае она определяется известным «правилом левой руки»: когда мы обходим границу в положительном направлении, область должна оставаться слева.

В частности, внешняя окружность на рис. 4.7 является границей области, дополнительной к единичному кругу. Поэтому ее ориентация противоположна ориентациям других окружностей, которые ограничивают обычные круги.

Возникает вопрос, нужно ли вообще рассматривать неограниченные области и их границы и нельзя ли обойтись обычными кругами. Ответ здесь — твердое «нет». Дело в том, что после перехода от плоскости \mathbb{R}^2 к расширенной плоскости $\overline{\mathbb{R}^2}$ разница между ограниченными и неограниченными областями стирается (потому что бесконечно удаленная точка теряет свое привилегированное положение и уравнивается в правах со всеми остальными точками)¹.

Таким образом, мы окончательно установили, что является основным предметом наших исследований: это — множество \mathcal{D} кругов на двумерной сфере S^2 , или на расширенной плоскости $\overline{\mathbb{R}^2}$, или на рас-

¹ Разумеется, можно держаться старых убеждений и настаивать на «особом положении» бесконечно удаленной точки. При этом тоже получается непротиворечивая теория. Но формулы этой теории сильно проигрывают в простоте и ясности — как сложная теория циклов Птолемея уступает простой и понятной теории Коперника.

ширенной комплексной плоскости $\overline{\mathbb{C}}$. Каждому такому кругу соответствует ориентированная окружность на сфере (окружность или прямая линия на $\overline{\mathbb{R}^2}$ или $\overline{\mathbb{C}}$).

Будем говорить, что два круга *касаются*, если они имеют ровно одну общую точку. В терминах ориентированных окружностей или прямых это означает «трицательное» касание: в общей точке ориентации двух границ противоположны. Теперь становится понятно, почему мы исключили конфигурации, показанные на рис. 4.2: они дают примеры «положительного» касания и не соответствуют конфигурациям четырех попарно касающихся кругов.

Теперь разберемся с понятием кривизны. Пусть C — ориентированная окружность с (обычным) радиусом r на $\overline{\mathbb{C}}$. Мы приписываем окружности C положительную кривизну $c = r^{-1}$, если C ограничивает, по правилу левой руки, обычный диск. Если же C — граница дополнения к обычному диску, то мы считаем, что кривизна окружности C отрицательна и равна $c = -r^{-1}$. Кривизна прямой линии считается равной нулю.

Оказывается, при таком понимании касания и кривизны формула Декарта верна без всяких исключений.

Замечание. Посмотрим более внимательно на знаки чисел c_i , $1 \leq i \leq 4$, составляющих решение уравнения (4.1.3). Заметим, что если четверка (c_1, c_2, c_3, c_4) является решением, то противоположная ей четверка $(-c_1, -c_2, -c_3, -c_4)$ — тоже решение. Однако мы увидим вскоре, что из двух противоположных четверок только одна реализуется как набор кривизн для четверки границ попарно касающихся кругов.

Уравнение (4.1.3) может быть переписано в форме

$$2(c_1 + c_2)(c_3 + c_4) = (c_1 - c_2)^2 + (c_3 - c_4)^2. \quad (4.1.6)$$

Отсюда мы видим, что либо $c_1 + c_2 \geq 0$ и $c_3 + c_4 \geq 0$, либо $c_1 + c_2 \leq 0$ и $c_3 + c_4 \leq 0$. Предположим, что нумерация чисел $\{c_i\}$ выбрана так, что $c_1 \geq c_2 \geq c_3 \geq c_4$.

Тогда в первом случае мы имеем $|c_4| \leq c_3 \leq c_2 \leq c_1$, а во втором $c_4 \leq c_3 \leq c_2 \leq -|c_1|$.

Читатель убедится сам, что только первый случай соответствует четверке кривизн границ попарно касающихся кругов. Поэтому только такие решения мы будем рассматривать в дальнейшем. При этом мы не теряем информации об остальных решениях уравнения Декарта: они состоят из противоположных четверок.

Таким образом, начиная с этого места мы предположим, что имеет место одна из следующих возможностей:

- а) все числа c_i положительны, или
- б) три числа положительны, а четвертое отрицательно и по абсолютной величине меньше каждого из остальных, или
- в) три числа положительны, четвертое равно нулю, или, наконец,
- г) два числа положительны и равны друг другу, а остальные два равны нулю.

Этот результат отражает очевидный геометрический факт: из четырех попарно касающихся обобщенных кругов по крайней мере два являются обыкновенными ограниченными кругами. ♡

Схолия F. КОНФОРМНАЯ ГРУППА И СТЕРЕОГРАФИЧЕСКАЯ ПРОЕКЦИЯ

F.1. Стереографическая проекция. В этом разделе мы рассматриваем общий n -мерный случай. Однако все рассуждения и вычисления практически одинаковы во всех размерностях. Поэтому читатель, незнакомый с предметом, может для начала полагать, что $n = 1$ или 2 .

Пусть $\overline{\mathbb{R}^n}$ — множество, получаемое из \mathbb{R}^n добавлением бесконечно удаленной точки ∞ . Топология в $\overline{\mathbb{R}^n}$ определяется так, что

$$x_k \rightarrow \infty \text{ в топологии } \overline{\mathbb{R}^n} \iff x_k \rightarrow \infty \text{ в смысле обычного анализа.}$$

Существует замечательное взаимно однозначное и взаимно непрерывное соответствие между n -мерной сферой S^n и $\overline{\mathbb{R}^n}$. Оно называется *стереографической проекцией*. Здесь мы дадим определение этого соответствия и укажем его основные свойства.

Пусть \mathbb{R}^{n+1} — евклидово пространство с координатами $(\alpha_0, \alpha_1, \dots, \alpha_n)$. Единичная сфера $S^n \subset \mathbb{R}^{n+1}$ задается уравнением $\alpha_0^2 + \alpha_1^2 + \dots + \alpha_n^2 = 1$. Точку $P = (1, 0, \dots, 0) \in S^n$ мы будем называть северным полюсом.

Пусть \mathbb{R}^n — другое евклидово пространство с координатами (x_1, x_2, \dots, x_n) . Удобно считать, что \mathbb{R}^n — подпространство в \mathbb{R}^{n+1} , состоящее из точек с координатами $(0, x_1, \dots, x_n)$.

Определим отображение s из $S^n \setminus \{P\}$ на \mathbb{R}^n по формуле:

$$s(\alpha) = \left(0, \frac{\alpha_1}{1 - \alpha_0}, \frac{\alpha_2}{1 - \alpha_0}, \dots, \frac{\alpha_n}{1 - \alpha_0} \right). \tag{4.1.1}$$

Обратное отображение имеет вид:

$$s^{-1}(x) = \left(\frac{|x|^2 - 1}{|x|^2 + 1}, \frac{2x_1}{1 + |x|^2}, \frac{2x_2}{1 + |x|^2}, \dots, \frac{2x_n}{1 + |x|^2} \right), \tag{4.1.2}$$

где $|x|^2 = x_1^2 + x_2^2 + \dots + x_n^2$.

УПРАЖНЕНИЕ 28. Проверьте, что точки P , $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_n)$ и $s(\alpha) = (0, x_1(\alpha), x_2(\alpha), \dots, x_n(\alpha))$ лежат на одной прямой в \mathbb{R}^{n+1} .

Таким образом, геометрически отображение s является проекцией $S^n \setminus \{P\}$ из северного полюса на координатную плоскость $\mathbb{R}^n \in \mathbb{R}^{n+1}$, определяемую уравнением $\alpha_0 = 0$.

Как аналитическое, так и геометрическое определения неприменимы к самому северному полюсу. Мы постановим отдельно, что $s(P) = \infty \in \overline{\mathbb{R}^n}$. Так определенное отображение s будет, очевидно, биекцией S^n на $\overline{\mathbb{R}^n}$. Эта биекция непрерывна по определению топологии на $\overline{\mathbb{R}^n}$.

Одно из замечательных свойств стереографической проекции — ее конформность. Это значит, что углы между пересекающимися кривыми сохраняются.

Второе замечательное свойство состоит в том, что все окружности на S^n переходят при стереографической проекции в окружности или прямые на $\overline{\mathbb{R}^n}$. Более того, все k -мерные сферы на S^n переходят в k -мерные сферы или k -мерные плоскости на $\overline{\mathbb{R}^n}$. При этом k -мерные шары на S^n переходят в обобщенные k -мерные шары на $\overline{\mathbb{R}^n}$.

Мы дадим здесь доказательство последнего свойства для n -мерных шаров на S^n . Любой такой шар $D \subset S^n$ можно рассматривать как пересечение S^n с полупространством в \mathbb{R}^{n+1} , заданным в координатах $\alpha_0, \alpha_1, \dots, \alpha_n$ линейным неравенством

$$p_0\alpha_0 + p_1\alpha_1 + \dots + p_n\alpha_n + p_{n+1} \leq 0. \quad (4.1.3)$$

Заметим, что гиперплоскость $p_0\alpha_0 + p_1\alpha_1 + \dots + p_n\alpha_n + p_{n+1} = 0$ пересекает единичную сферу S^n по нетривиальной $(n-1)$ -мерной сфере тогда и только тогда, когда

$$p_{n+1}^2 - p_0^2 - p_1^2 - \dots - p_n^2 < 0. \quad (4.1.4)$$

Мы видим, что естественно рассматривать вектор p как элемент пространства $\mathbb{R}^{1,n+1}$ (см. ниже с. 87) и писать левую часть неравенства (4.1.4) как $|p|^2$. Поскольку умножение p на положительный множитель не меняет смысла неравенства (4.1.3), мы можем нормализовать p условием¹ $|p|^2 = -1$.

Выражая $\{\alpha_i\}$ через координаты $\{x_j\}$ точки $s(\alpha)$, мы запишем неравенство, задающее $s(D)$, в виде

$$p_0(|x|^2 - 1) + 2p_1x_1 + \dots + 2p_nx_n + p_{n+1}(|x|^2 + 1) \leq 0.$$

¹ Не спутайте вектор $p \in \mathbb{R}^{1,n+1}$ с вектором $\vec{p} \in \mathbb{R}^n$, который будет определен ниже.

В терминах скалярного произведения в \mathbb{R}^n это значит

$$a + 2(\vec{p}, \vec{x}) + c|\vec{x}|^2 \leq 0, \tag{4.1.5}$$

где $a = p_{n+1} - p_0$, $c = p_{n+1} + p_0$, $\vec{p} = (p_1, \dots, p_n)$ и $\vec{x} = (x_1, \dots, x_n)$.

Теперь мы используем равенство $|p|^2 = ac - |\vec{p}|^2$ и нормализацию $|p|^2 = -1$, чтобы придать нашему неравенству окончательную форму

$$c \cdot \left| x + \frac{\vec{p}}{c} \right|^2 \leq c^{-1}. \tag{4.1.6}$$

Это неравенство при $c > 0$ задает замкнутый шар в \mathbb{R}^n с центром в точке $-\frac{\vec{p}}{c}$ и с кривизной c . В этом случае исходный диск D не содержит северного полюса.

Если же $c < 0$, то (4.1.6) задает дополнение к открытому шару с центром в точке $-\frac{\vec{p}}{c}$ и радиусом $-\frac{1}{c}$. Мы условились считать кривизной границы этого обобщенного шара отрицательную величину c . В этом случае исходный шар D содержит северный полюс внутри.

Наконец, если $c = 0$, то (4.1.6) не имеет смысла, а (4.1.5) задает замкнутое полупространство в R^n . В этом случае северный полюс лежит на границе исходного шара D .

F.2. Конформная группа. Существует большая группа $Conf_n$ (или, короче, C_n) так называемых *конформных преобразований*, действующая на S^n и на $\overline{\mathbb{R}^n}$ так, что стереографическая проекция s является C_n -ковариантным отображением. Это значит, что для любого $g \in C_n$ следующая диаграмма коммутативна:

$$\begin{array}{ccc} S^n & \xrightarrow{s} & \overline{\mathbb{R}^n} \\ g \downarrow & & \downarrow g \\ S^n & \xrightarrow{s} & \overline{\mathbb{R}^n}, \end{array} \tag{4.1.7}$$

где g означает действие $g \in C_n$ на S^n и на $\overline{\mathbb{R}^n}$.

Напомним, что диаграмма, состоящая из множеств и отображений, называется *коммутативной*, если для любого пути, составленного из стрелок диаграммы, произведение соответствующих отображений зависит только от начальной и конечной точек пути. В нашем случае есть два пути, ведущих из левого верхнего угла в правый нижний угол диаграммы, и коммутативность значит, что $s \circ g = g \circ s$.

Группа C_n может быть определена многими способами. Мы дадим здесь три эквивалентных определения.

ГЕОМЕТРИЧЕСКОЕ ОПРЕДЕЛЕНИЕ (только для $n > 1$). Напомним, что гладкое отображение $g: \mathbb{R}^n \rightarrow \mathbb{R}^n$ имеет в каждой точке $x \in \mathbb{R}^n$ производную $g'(x)$, которая является линейным оператором $g'(x): \mathbb{R}^n \rightarrow \mathbb{R}^n$. Мы говорим, что g *конформно*, если $g'(x)$ в каждой точке x является композицией вращения и подобия.

Таким образом, в бесконечно малом конформные отображения сохраняют форму фигур. Это объясняет происхождение термина «конформный».

В случае $n = 1$ группа конформных в этом смысле отображений слишком велика (бесконечномерна). А именно, она совпадает с группой всех гладких отображений S^1 .

В случае $n = 2$ группа всех конформных преобразований расширенной плоскости \mathbb{C} уже конечномерна: каждое взаимно однозначное конформное преобразование \mathbb{C} является дробно-линейным (см. ниже).

Теперь мы дадим другое, теоретико-групповое определение, которое для $n > 1$ эквивалентно геометрическому и задает конечномерную группу в любой размерности n .

Пусть $E_n(\mathbb{R})$ — группа всех движений \mathbb{R}^n , то есть вращений и параллельных переносов. Действие этой группы на \mathbb{R}^n естественно продолжается на $\overline{\mathbb{R}^n}$ так, что бесконечно удаленная точка остается неподвижной.

Назовем *инверсией*, или *отражением в единичной сфере*, преобразование $Inv: \overline{\mathbb{R}^n} \rightarrow \overline{\mathbb{R}^n}$, задаваемое формулой:

$$Inv(x) = \begin{cases} \frac{x}{|x|^2}, & \text{если } x \neq 0, \infty, \\ \infty, & \text{если } x = 0, \\ 0, & \text{если } x = \infty. \end{cases} \quad (4.1.8)$$

ТЕОРЕТИКО-ГРУППОВОЕ ОПРЕДЕЛЕНИЕ. Сначала мы определим так называемую *расширенную конформную группу* $Conf_n$ как группу, порожденную $E_n(\mathbb{R})$ и Inv .

Группа $Conf_n$ состоит из двух связных компонент. Преобразования из одной компоненты сохраняют ориентацию $\overline{\mathbb{R}^n}$. Они являются композицией движений и четного числа инверсий: $g_1 \circ Inv \circ g_2 \circ Inv \circ \dots \circ g_{2n} \circ Inv$. Эта компонента сама является группой и, по определению, называется конформной группой $Conf_n$.

Преобразования из другой компоненты обращают ориентацию $\overline{\mathbb{R}^n}$. Они иногда называются *конформными преобразованиями второго рода* (в случае $n = 2$ это антианалитические отображения).

Общий элемент второй компоненты является произведением движений и нечетного числа инверсий: $g_1 \circ Inv \circ g_2 \circ Inv \circ \dots \circ g_{2n+1} \circ Inv$.

Наконец, наиболее рабочее определение — это следующее, которое мы назовем *матричным определением*.

Пусть $\mathbb{R}^{1,n+1}$ означает вещественное векторное пространство с координатами $(x_0, x_1, \dots, x_{n+1})$ и со скалярным произведением, заданным симметричной билинейной формой

$$B(x, y) = x_0y_0 - x_1y_1 - \dots - x_{n+1}y_{n+1}.$$

Группа линейных преобразований, сохраняющих эту форму, называется *псевдоортогональной группой* и обозначается $O(1, n + 1; \mathbb{R})$. Относительно выбранного базиса элементы этой группы задаются блочными матрицами вида $g = \begin{pmatrix} a & \vec{b}^t \\ \vec{c} & D \end{pmatrix}$, где a — вещественное число, \vec{b}, \vec{c} — вектор-столбцы размера $n + 1$, индекс t означает транспонирование, а D — матрица порядка $n + 1$. Тот факт, что $g \in O(1, n + 1; \mathbb{R})$, выражается уравнениями

$$a^2 = 1 + |\vec{c}|^2, \quad D^t D = 1_{n+1} + \vec{b}\vec{b}^t, \quad D^t \vec{c} = \vec{b}a, \quad (4.1.9)$$

где 1_{n+1} означает единичную матрицу порядка $n + 1$.

Из (4.1.9) следует, что a и D обратимы (проверьте самостоятельно, что $D^{-1} = 1_{n+1} - \frac{\vec{b}\vec{b}^t}{1 + |\vec{b}|^2}$). Поэтому группа $O(1, n + 1; \mathbb{R})$ разбивается на четыре части согласно знакам a и $\det D$. На самом деле эти части являются связными компонентами. Более точно, наша группа диффеоморфна произведению $O(n, \mathbb{R}) \times S^1 \times \mathbb{R} \times \mathbb{Z}_2$: каждой четверке $(A, \vec{v}, \tau, \pm 1)$ соответствует элемент

$$g = \pm \begin{pmatrix} \text{ch } \tau & \vec{v}^t \cdot \text{sh } \tau \\ \text{sh } \tau \cdot A\vec{v} & \text{ch } \tau \cdot A \end{pmatrix} \in O(1, n + 1; \mathbb{R}) \quad (4.1.10)$$

и, наоборот, каждая матрица из $O(1, n + 1; \mathbb{R})$ записывается в такой форме.

Псевдоортогональная группа действует на пространстве $\mathbb{R}^{1,n+1}$, сохраняя конус

$$C: x_0^2 = x_1^2 + \dots + x_{n+1}^2.$$

Она действует также на проективном пространстве, связанном с $\mathbb{R}^{1,n+1}$. Поскольку скалярные матрицы действуют тривиально, мы имеем фактически действие соответствующей проективной группы

$PO(1, n + 1; \mathbb{R}) = O(1, n + 1; \mathbb{R}) / \{\pm 1\}$, которая является факторгруппой $O(1, n + 1; \mathbb{R})$ по ее центру $\{\pm 1_{n+2}\}$. Эта группа имеет только две связные компоненты $PO_{\pm}(1, n + 1; \mathbb{R})$, различаемые знаком величины $\det(a^{-1}D)$.

Проективизация конуса C (с выброшенной вершиной) может быть отождествлена со сферой S^n с помощью координат $\alpha_i = \frac{x_i}{x_0}$, $1 \leq i \leq n + 1$, и с $\overline{\mathbb{R}^n}$ с помощью координат $w_j = \frac{x_j}{x_0 - x_{n+1}}$, $1 \leq j \leq n$.

Обратно, координаты точки конуса могут быть восстановлены, с точностью до пропорциональности, по формулам:

$$x_0 = \frac{\sum_j w_j^2 + 1}{2}, \quad x_j = w_j \text{ для } 1 \leq j \leq n, \quad x_{n+1} = \frac{\sum_j w_j^2 - 1}{2}. \quad (4.1.11)$$

Следующий факт хорошо известен, и мы используем его для матричного определения конформной группы.

ТЕОРЕМА F.1. Группа $PO(1, n + 1; \mathbb{R})$, действующая на S^n (или на $\overline{\mathbb{R}^n}$), совпадает с \overline{Conf}_n , а ее связная подгруппа $PO_+(1, n + 1; \mathbb{R})$ совпадает с $Conf_n$.

УПРАЖНЕНИЕ 29. Инверсия $Inv \in \overline{Conf}_n$ в последней реализации соответствует некоторому элементу $PO(1, n + 1; \mathbb{R})$, то есть паре матриц $\pm g \in O(1, n + 1; \mathbb{R})$. Найдите эту пару.

Подсказка. Используйте (4.1.11).

ОТВЕТ. $g = \text{diag}(1, 1, \dots, 1, -1)$.

F.3. Малые размерности. В основном тексте мы рассмотрим подробнее случай $n = 2$, а также $n = 3$, $n = 4$. Во всех этих случаях конформная группа имеет дополнительные свойства, которые мы обсудим здесь.

Случай $n = 2$. Группа $Conf_2$ изоморфна $PO_+(1, 3; \mathbb{R})$, а также группе Мёбиуса $PSL(2, \mathbb{C})$. Расширенная конформная группа \overline{Conf}_2 изоморфна $PO(1, 3; \mathbb{R})$, а также расширенной группе Мёбиуса.

Напомним, что группа Мёбиуса действует на расширенной комплексной плоскости $\overline{\mathbb{C}}$ так называемыми *дробно-линейными*, или *мёбусовыми*, преобразованиями:

$$w \rightarrow \frac{\alpha w + \beta}{\gamma w + \delta}, \quad \text{где } \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in SL(2, \mathbb{C}). \quad (4.1.12)$$

Расширенная группа Мёбиуса содержит также комплексное сопряжение и, следовательно, все преобразования вида

$$w \rightarrow \frac{\alpha \bar{w} + \beta}{\gamma \bar{w} + \delta}, \quad \text{где } \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in SL(2, \mathbb{C}). \quad (4.1.13)$$

Среди этих преобразований выделяются так называемые *отражения* r , обладающие свойством $r^2 = 1$ и имеющие в качестве множества неподвижных точек окружность или прямую линию. Обозначим это множество M_r и назовем его *зеркалом*, а преобразование r — отражением в этом зеркале. Обратное, для каждой окружности или прямой линии M существует единственное отражение r , для которого $M_r = M$; мы обозначим его r_M .

Если зеркало M является прямой линией l , преобразование r_M является обычным евклидовым отражением в прямой l .

Если M — единичная окружность S с центром в начале координат, то r_M совпадает с инверсией Inv , определенной формулой (4.1.8).

В общем случае отражение в окружности M можно задать формулой $r_M = g \circ Inv \circ g^{-1}$, где $g \in Conf_2$ — любое конформное преобразование, переводящее S в M .

УПРАЖНЕНИЕ 30. Покажите, что все отражения образуют один класс сопряженности в группе $Conf_2$.

Подсказка. Проверьте, что $Conf_2$ действует транзитивно на \mathcal{D} .

УПРАЖНЕНИЕ 31. Покажите, что группа $Conf_2$ порождается отражениями.

Подсказка. Используйте хорошо известный факт, что $SL(2, \mathbb{C})$ порождается элементами

$$g(t) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \quad t \in \mathbb{C}, \quad \text{и} \quad s = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (4.1.14)$$

УПРАЖНЕНИЕ 32. Покажите, что классы сопряженных элементов в $Conf_2$ — это в точности множества уровня $I(g) = \text{const}$ для функции

$$I(g) := \frac{(\text{tr } g)^2}{\det g} - 4 \quad (4.1.15)$$

с единственным исключением: множество $I(g) = 0$ является объединением двух классов: $\{e\}$ и класса, содержащего жорданов блок J_2 .

УПРАЖНЕНИЕ 33. Покажите, что все инволютивные отображения в $Conf_2$ образуют два класса сопряженных элементов: единичный класс и класс, состоящий из вращений сферы S^2 на 180° вокруг какой-нибудь оси.

УПРАЖНЕНИЕ 34. Покажите, что все инволюции в $PO_-(1, 3; \mathbb{R})$, которые не являются отражением в окружности или прямой, образуют один класс сопряженных элементов, содержащий антиподальное отображение S^2 .

Мы приведем здесь два основных свойства группы Мёбиуса G .

Предложение F.1. Для любых троек различных точек (z_1, z_2, z_3) и (w_1, w_2, w_3) на $\overline{\mathbb{C}}$ существует единственное преобразование $g \in G$ такое, что $g(z_i) = w_i$, $i = 1, 2, 3$.

Доказательство. Сначала проверим это утверждение для случая $w_1 = 0$, $w_2 = 1$, $w_3 = \infty$. Соответствующее преобразование g_{z_1, z_2, z_3} может быть выписано явно:

$$g_{z_1, z_2, z_3}(z) = \frac{z - z_1}{z - z_3} : \frac{z_2 - z_1}{z_2 - z_3}. \quad (4.1.16)$$

В общем случае искомое преобразование g равно $g = g_{w_1, w_2, w_3}^{-1} \circ g_{z_1, z_2, z_3}$. \square

Предложение F.2. Любая окружность или прямая линия переходит после преобразования $g \in G$ в окружность или прямую линию. (Или: любой круг переходит в круг.)

Для доказательства нам понадобится

Лемма F.1. Пусть a, c — вещественные числа, b — комплексное число такие, что $ac - |b|^2 < 0$. Тогда неравенство

$$a + \bar{b}w + b\bar{w} + cw\bar{w} \leq 0 \quad (4.1.17)$$

задает круг $D \in \mathcal{D}$. Более точно, это:

- а) замкнутый круг с центром $-\frac{b}{c}$ и радиусом $r = c^{-1}$, если $c > 0$;
- б) дополнение к открытому кругу с центром $-\frac{b}{c}$ и радиусом $r = -c^{-1}$, если $c < 0$;
- в) замкнутая полуплоскость, если $c = 0$.

Более того, каждый обобщенный круг $D \in \mathcal{D}$ задается неравенством вида (4.1.17).

Доказательство. Это — частный случай (4.1.5). \square

Предложение F.1 следует из леммы F.1, поскольку неравенство (4.1.17) сохраняет свою форму при преобразованиях вида (4.1.14), следовательно, при всех дробно-линейных преобразованиях.

Замечание. Множество $\overline{\text{Conf}}_2 \setminus \text{Conf}_2$ конформных преобразований второго рода не образует группы. Оно является двусторонним смежным классом в $\overline{\text{Conf}}_2$ относительно Conf_2 .

Стоит отметить, что это множество преобразований обладает обоими свойствами, упомянутыми в предложениях F.1 и F.2: оно действует просто транзитивно на тройках различных точек в $\overline{\mathbb{C}}$ и сохраняет обобщенные окружности и диски. \heartsuit

Случай $n = 3$. Группа $\text{Conf}_3 = \text{PSO}_+(1, 4; \mathbb{R})$ изоморфна группе $\text{PU}(1, 1; \mathbb{H})$, которая является факторгруппой кватернионной псевдоунитарной группы $U(1, 1; \mathbb{H})$ по ее центру $\{\pm 1_2\}$.

Группа $U(1, 1; \mathbb{H})$ состоит из кватернионных матриц второго порядка $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, удовлетворяющих соотношениям:

$$|a|^2 = |d|^2 = 1 + |b|^2 = 1 + |c|^2, \quad \bar{a}b = \bar{c}d.$$

Положим $a = u \operatorname{ch} t$, $d = v \operatorname{ch} t$, где $t \in \mathbb{R}$, а u, v — кватернионы единичной нормы. Тогда существует такой кватернион единичной нормы w , что $b = w \operatorname{sh} t$ и $c = \bar{w}u\bar{v} \operatorname{sh} t$.

Если матрица g не диагональна, параметры u, v, w и t определены однозначно. Для диагональных матриц мы имеем $t = 0$ и значение w не влияет на g . Таким образом, наша группа является объединением $S^3 \times S^3 \times S^3 \times (\mathbb{R} \setminus \{0\})$ и $S^3 \times S^3$.

Группа $U(1, 1; \mathbb{H})$ действует на двумерном правом кватернионном пространстве \mathbb{H}^2 . Элементы \mathbb{H}^2 — вектор-столбцы $\begin{pmatrix} q_1 \\ q_2 \end{pmatrix}$ с двумя кватернионными координатами; умножение на числа (кватернионы) производится справа, а действие группы — умножением слева на матрицу из нашей группы.

Это действие сохраняет конус $|q_1|^2 - |q_2|^2 = 0$ и переносится на его кватернионную проективизацию, которая изоморфна сфере S^3 . Явный вид последнего действия дается формулой: $u \mapsto (au + b)(cu + d)^{-1}$.

УПРАЖНЕНИЕ 35. Проверьте, что если $|u| = 1$ и матрица $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in U(1, 1; \mathbb{H})$, то $|(au + b)(cu + d)^{-1}| = 1$.

Случай $n = 4$. Группа $Conf_4 = PO_+(1, 5; \mathbb{R})$ изоморфна другой проективной кватернионной группе $PGL(2, \mathbb{H}) = GL(2, \mathbb{H})/\mathbb{R}^\times \cdot 1_2$.

Явная формула действия снова имеет вид дробно-линейного преобразования $q \mapsto (aq + b)(cq + d)^{-1}$. Но теперь наша группа действует на кватернионной проективной прямой $\mathbb{P}^1(\mathbb{H}) \simeq \overline{\mathbb{H}} \simeq \mathbb{R}^4 \simeq S^4$.

Глава 5

СТРОГОЕ ОПРЕДЕЛЕНИЕ КОВРА АПОЛЛОНИЯ

5.1. ОСНОВНЫЕ ФАКТЫ

Рассмотрим четыре попарно касающихся круга на двумерной сфере S^2 . Если мы удалим из сферы внутренние точки всех четырех кругов, то останется четыре криволинейных треугольника. Впишем в каждый из них максимальный возможный круг и удалим его внутренние точки. Вместо каждого криволинейного треугольника останется 3 меньших треугольника, а всего — 12 треугольников. Опять впишем в каждый из них максимальный возможный круг и удалим его внутренние точки. Останется 36 треугольников...

Продолжая этот процесс, мы удалим из сферы счетное множество открытых кругов. Оставшееся замкнутое подмножество \mathcal{A} имеет фрактальную природу и называется *ковром Аполлония* в честь древнегреческого математика Аполлония Пергского, жившего в III–II столетиях до нашей эры. Разумеется, можно было бы заменить сферу на \mathbb{R}^2 или на \mathbb{C} ; получаемое множество также именуется ковром Аполлония.

Согласно сложившейся практике, термином «ковер Аполлония» пользуются также для обозначения множества кругов (открытых и замкнутых) и множества окружностей, участвующих в построении \mathcal{A} .

Посмотрим, насколько различные формы может иметь ковер Аполлония \mathcal{A} . На первый взгляд, различные выборы начальных четырех кругов приводят к разным и непохожим друг на друга коврам. Тем не менее, все получаемые картины в определенном смысле эквивалентны.

Чтобы понять это, рассмотрим группу G конформных преобразований расширенной комплексной плоскости \mathbb{C} (см. формулу (4.1.12)).

Упражнение 36. Покажите, что любые две неупорядоченные четверки попарно касающихся кругов на сфере могут быть преобразованы одна в другую с помощью конформного отображения.

Подсказка. Покажите, что тройка попарно касающихся кругов на сфере однозначно определяется тройкой точек касания (воспользуйтесь предложением F.1). Затем покажите, что тройка попарно касающихся окружностей может быть ровно одним способом дополне-

на до четверки попарно касающихся окружностей и что эти две четверки переводятся одна в другую конформным преобразованием.

Таким образом, с точностью до конформного преобразования существует только один ковер Аполлония.

ТЕОРЕМА 5.1. *Ковер Аполлония \mathcal{A} однозначно определяется любой тройкой попарно касающихся кругов в нем. (Другими словами, если два ковра Аполлония имеют общую тройку попарно касающихся кругов, то они совпадают.)*

Утверждение теоремы выглядит достаточно очевидным, и я рекомендую читателю попробовать придумать собственное доказательство. То, которое придумал я и которое приводится ниже, довольно длинно и основано на специальной нумерации кругов в \mathcal{A} .

Нумерация, о которой идет речь, подсказывается самой процедурой построения ковра. Именно, назовем начальные четыре круга D^1, D^2, D^3, D^4 кругами уровня 0. После удаления из S^2 внутренних точек четырех кругов останется объединение четырех замкнутых криволинейных треугольников; назовем их *треугольниками уровня 1* и обозначим их T_1, T_2, T_3, T_4 так, чтобы T_i не имел общих точек с D^i . Впишем в каждый из этих треугольников максимальный возможный круг; назовем эти круги *кругами уровня 1* и обозначим их D_i .

После удаления из T_i внутренности круга D_i останется объединение трех треугольников. Мы будем называть их *треугольниками уровня 2* и обозначать $T_{ij}, j \neq i$. В каждый из треугольников уровня 2 впишем максимальный возможный круг; эти круги будем обозначать D_{ij} и называть *кругами уровня 2*.

Продолжим эту процедуру до бесконечности. На n -м шаге мы рассматриваем треугольник $T_{i_1 i_2 \dots i_{n-1}}$, вписываем в него максимальный возможный круг $D_{i_1 i_2 \dots i_{n-1}}$ и удаляем его внутренность. Остается объединение трех треугольников уровня n , обозначаемых $T_{i_1 i_2 \dots i_{n-1} i_n}, i_n \neq i_{n-1}$.

Заметим, что два различных круга одного и того же уровня $n \geq 1$ никогда не касаются.

Итак, мы пронумеровали все треугольники (или круги) уровня $n \geq 1$ последовательностями вида $i_1 i_2 \dots i_n$, где i_k принимают значения 1, 2, 3, 4 и $i_k \neq i_{k+1}$ (рис. 5.1). Наряду с кругами уровня ≥ 1 имеются только четыре исходных круга $D^i, 1 \leq i \leq 4$.

ЛЕММА 5.1. *Пусть D, D', D'', D''' — четыре попарно касающихся круга на S^2 . Если три из них принадлежат некоторому ковра \mathcal{A} , то и четвертый принадлежит тому же ковра.*

Доказательство. Предположим: что круги D, D', D'' принадлежат ковра \mathcal{A} и имеют в нем уровни m, m', m'' соответственно; можно

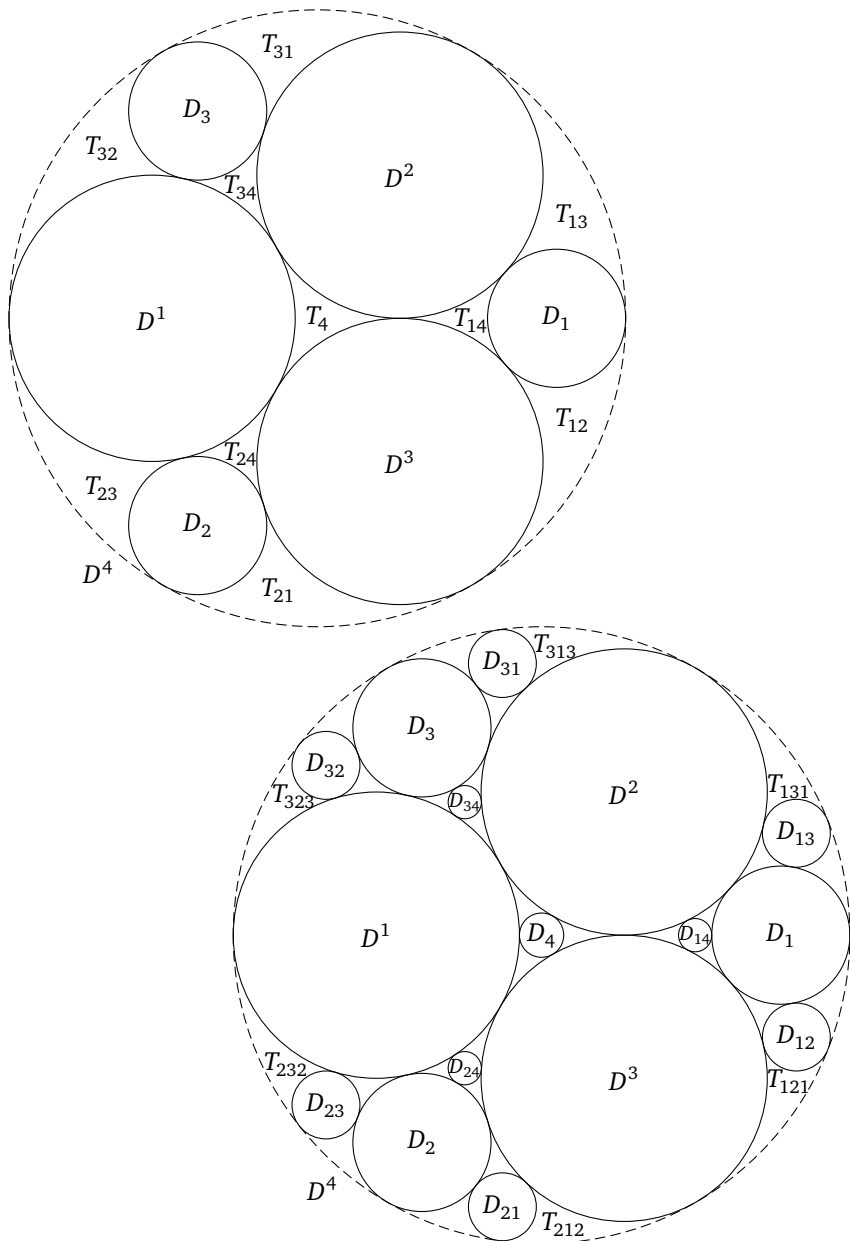


Рис. 5.1. Нумерация кругов

считать, что $t \leq m' \leq m''$. Поскольку диски одного уровня касаться не могут, возможны следующие четыре случая: 1) $0 < m < m' < m''$; 2) $0 = m < m' < m''$; 3) $0 = m = m' < m''$; 4) $0 = m = m' = m''$.

В первом случае можно считать, что $D = D_{i_1 i_2 \dots i_m}$, $D' = D_{j_1 j_2 \dots j_{m'}}$, и $D'' = D_{k_1 k_2 \dots k_{m''}}$. По построению круг D'' вписан в в треугольник $T_{k_1 k_2 \dots k_{m''}}$, ограниченный дугами трех кругов. Один из них — круг $D_{k_1 k_2 \dots k_{m''-1}}$ уровня $m'' - 1$; уровни двух других обозначим через p и q , причем можно считать, что $p \leq q < m'' - 1$ (равенство возможно только при $p = 0$).

Из конструкции ковра ясно, что уровни всех кругов, касающихся D'' , кроме трех упомянутых выше, превосходят m'' . Однако же нам известно, что D и D' касаются D'' . Стало быть, $m = p$, $m' = q$ и три круга, касающихся D'' , суть D , D' и $D_{k_1 k_2 \dots k_{m''-1}}$. Отсюда вытекает, что круг $D_{k_1 k_2 \dots k_{m''-1}}$ касается кругов D , D' и D'' . Кроме того, этих кругов касается круг $D_{k_1 k_2 \dots k_{m''} k_{m''-1}}$, имеющий уровень $m'' + 1$. Тем самым оба круга, касающиеся D , D' и D'' , принадлежат ковра.

В остальных случаях доказательство аналогично, но проще. Например, круг, касающийся D^1 , D^2 и D_{34} , совпадает либо с D_3 , либо с D_{343} и тем самым принадлежит ковра. \square

Доказательство теоремы. Пусть и ковер \mathcal{A} , и ковер $\tilde{\mathcal{A}}$ содержит три попарно касающихся круга D , D' и D'' . Обозначим их уровни в ковре \mathcal{A} через $l \leq m \leq n$. Докажем включение $\mathcal{A} \subset \tilde{\mathcal{A}}$ индукцией по n .

Если $n = 0$, то три наши круга суть исходные круги D^1 , D^2 и D^3 для \mathcal{A} . Поскольку эти три круга принадлежат $\tilde{\mathcal{A}}$, лемма 5.1 показывает, что D^4 and D_4 также принадлежат $\tilde{\mathcal{A}}$.

Чтобы показать, что всякий круг из ковра \mathcal{A} принадлежит $\tilde{\mathcal{A}}$, проведем индукцию по уровню. Именно, если это верно для всех кругов уровня $\leq n - 1$, то всякий диск уровня n также принадлежит $\tilde{\mathcal{A}}$, поскольку он касается трех дисков уровня $n - 1$, принадлежащих $\tilde{\mathcal{A}}$.

Вернемся к исходной индукции. Предположим, что доказано следующее утверждение: если три диска из условия имеют уровень $< n$ в \mathcal{A} , то они принадлежат $\tilde{\mathcal{A}}$. Пусть теперь D , D' и D'' — попарно касающиеся диски из \mathcal{A} , имеющие уровни $k \leq l < n$. Из доказательства леммы 5.1 видно, что среди дисков, касающихся D , D' и D'' , найдется диск D''' уровня $n - 1$. По предположению индукции D , D' и D''' принадлежат $\tilde{\mathcal{A}}$; следовательно, D'' принадлежит $\tilde{\mathcal{A}}$ ввиду леммы.

Итак, $\mathcal{A} \subset \tilde{\mathcal{A}}$. Но условия теоремы симметричны относительно \mathcal{A} и $\tilde{\mathcal{A}}$. Поэтому $\tilde{\mathcal{A}} \subset \mathcal{A}$ и, следовательно, эти ковры совпадают. \square

ЛЕММА 5.2. Треугольник $T_{i_1 i_2 \dots i_n}$ содержится в $T_{j_1 j_2 \dots j_m}$, если и только если $t \leq n$ и $i_k = j_k$ для $1 \leq k \leq t$.

Доказательство. Заметим, что два разных треугольника одного уровня могут иметь не более одной общей точки, так что первый из наших треугольников может содержаться только в одном треугольнике уровня m . Поскольку он содержится в $T_{i_1 i_2 \dots i_m}$ и $T_{j_1 j_2 \dots j_m}$, лемма отсюда следует. \square

Есть три наиболее симметричных выбора для начальной четверки попарно касающихся кругов на плоскости. Соответствующие ковры Аполлония показаны на рис. 5.2, 5.3, 5.4.

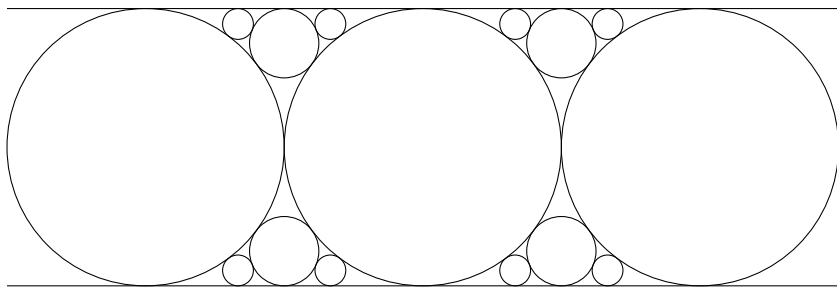


Рис. 5.2. Ковер-лента

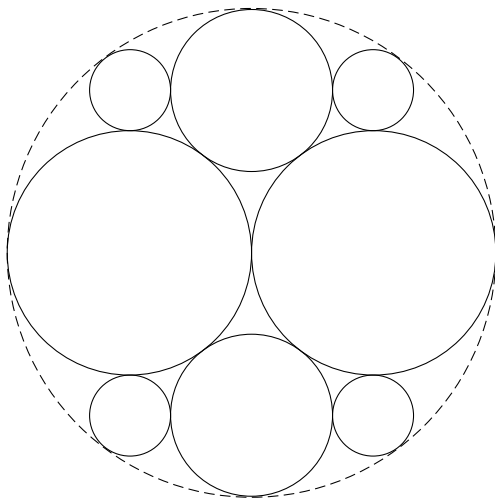


Рис. 5.3. Прямоугольный ковер

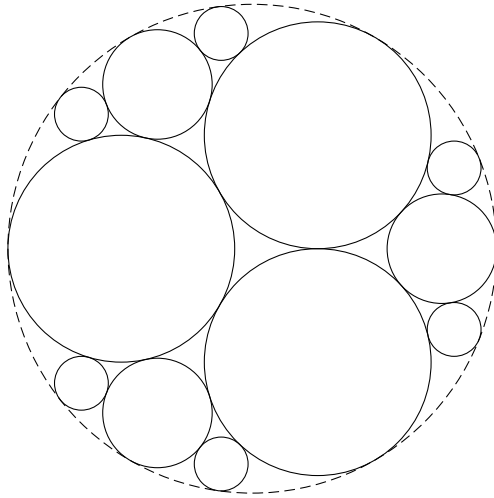


Рис. 5.4. Треугольный ковер

Все эти ковры являются стереографическими проекциями наиболее симметричного ковра на сфере S^2 , порожденного тремя попарно касающимися кругами одинакового размера. См. рис. 5.5.

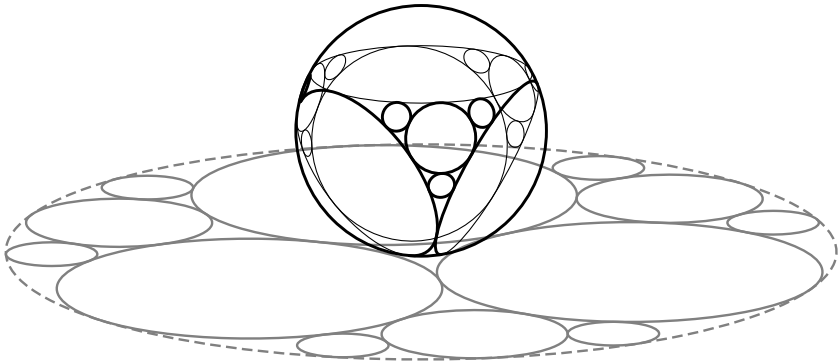


Рис. 5.5. Сферический ковер

Есть также несколько других интересных реализаций ковра Аполлония, из которых мы отметим две. При их исследовании используются некоторые свойства так называемых чисел Фибоначчи.

Схолия Г. Числа Фибоначчи

Знаменитый итальянский математик Леонардо Пизанский, известный также под именем Фибоначчи, жил в XIII столетии. Среди прочих вещей он исследовал последовательность целых чисел $\{\Phi_k\}$, удовлетворяющую рекуррентному уравнению

$$\Phi_{k+1} = \Phi_k + \Phi_{k-1} \quad (5.1.1)$$

и начальному условию $\Phi_1 = \Phi_2 = 1$. Она выглядит следующим образом:

$$\begin{array}{cccccccccccccccc} n: & -7 & -6 & -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \Phi_n: & 13 & -8 & 5 & -3 & 2 & -1 & 1 & 0 & 1 & 1 & 2 & 3 & 5 & 8 & 13 \end{array}$$

Позднее эти числа появлялись во многих алгебраических и комбинаторных задачах и получили имя *числа Фибоначчи*. Мы коротко опишем основные факты из теории таких последовательностей.

Рассмотрим совокупность V всех двусторонних последовательностей $\{v_n\}_{n \in \mathbb{Z}}$, удовлетворяющих рекуррентному уравнению вида (5.1.1), то есть $v_{n+1} = v_n + v_{n-1}$. Это — вещественное векторное пространство, в котором операции сложения и умножения на число определены покомпонентно:

$$(v + w)_n = v_n + w_n, \quad (c \cdot v)_n = c \cdot v_n.$$

Размерность этого пространства равна 2, поскольку каждая последовательность из V однозначно определяется своими начальными значениями и эти значения можно выбирать произвольно. Мы можем, следовательно, рассматривать числа v_1, v_2 как координаты последовательности $\{v_n\}$ в V . Например, последовательность Фибоначчи как точка из V имеет координаты $(1, 1)$. Другая известная последовательность чисел Люка имеет координаты $(1, 3)$.

Пусть T означает преобразование пространства V , переводящее последовательность $\{v_n\}$ в последовательность $\{v_{n+1}\}$ (также принадлежащую V , — проверьте!). Наглядный смысл T — сдвиг всей последовательности на один номер влево.

Это — линейный оператор в пространстве V . Найдем спектр этого оператора. Если число λ принадлежит спектру, то для некоторой ненулевой последовательности $\{v_n\}$ мы имеем $v_{n+1} = \lambda \cdot v_n$ для всех $n \in \mathbb{Z}$. Отсюда $v_n = c \cdot \lambda^n$, $c \neq 0$, и рекуррентное условие дает квадратное уравнение на λ : $\lambda^2 = \lambda + 1$. Мы видим, что спектр состоит из двух

чисел:

$$\varphi = \frac{1+\sqrt{5}}{2} \approx 1,618\dots \quad \text{и} \quad -\varphi^{-1} = \frac{1-\sqrt{5}}{2} = 1 - \varphi \approx -0,618\dots$$

Первое из них получило специальное название *золотого сечения*, поскольку прямоугольник с таким отношением сторон считается наиболее приятным для человеческого глаза.

Для использования в дальнейшем мы введем также величины

$$c = \varphi^2 = \frac{3+\sqrt{5}}{2} = \varphi + 1 \quad \text{и} \quad \theta = \sqrt{\varphi} = \sqrt{\frac{1+\sqrt{5}}{2}}.$$

Собственные векторы оператора T являются геометрическими прогрессиями $v'_n = \varphi^n$ и $v''_n = (-\varphi)^{-n}$. Они линейно независимы и образуют базис в пространстве V . Поэтому каждый вектор $v \in V$ является линейной комбинацией v' и v'' . В частности, n -е число Фибоначчи может быть записано как

$$\Phi_n = \alpha \cdot \varphi^n + \beta \cdot (-\varphi^{-1})^n \quad \text{для подходящих } \alpha \text{ и } \beta.$$

Из начальных условий $\Phi_1 = \Phi_2 = 1$ мы находим $\alpha = -\beta = \frac{1}{\varphi + \varphi^{-1}} = \frac{1}{\sqrt{5}}$. Поэтому

$$\Phi_{2k} = \frac{\varphi^{2k} - \varphi^{-2k}}{\sqrt{5}} = \frac{c^k - c^{-k}}{\sqrt{5}}; \quad \Phi_{2k+1} = \frac{\varphi^{2k+1} + \varphi^{-2k-1}}{\sqrt{5}} = \frac{c^{k+1/2} + c^{-k-1/2}}{\sqrt{5}}. \quad (5.1.2)$$

Обратно,

$$\varphi^n = (-1)^n \frac{\Phi_{n+1} + \Phi_{n-1} - \Phi_n \sqrt{5}}{2}; \quad c^n = \frac{\Phi_{2n+1} + \Phi_{2n-1} + \Phi_{2n} \sqrt{5}}{2}. \quad (5.1.3)$$

Отметим также, что $\Phi_{-2n} = -\Phi_{2n}$; $\Phi_{-2n-1} = \Phi_{2n+1}$.

Отсюда следует, что при больших n

$$\Phi_n \approx \frac{\varphi^n}{\sqrt{5}} \quad \text{и} \quad \lim_{n \rightarrow \infty} \frac{\Phi_{n+1}}{\Phi_n} = \varphi. \quad (5.1.4)$$

Числа Люка даются еще более простой формулой: $L_n = \varphi^n + (-\varphi^{-1})^{-n}$. Они выглядят так:

$$\begin{array}{cccccccccccccccc} n: & -7 & -6 & -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ L_n: & -29 & 18 & -11 & 7 & -4 & 3 & -1 & 2 & 1 & 3 & 4 & 7 & 11 & 18 & 29 \quad \diamond \end{array}$$

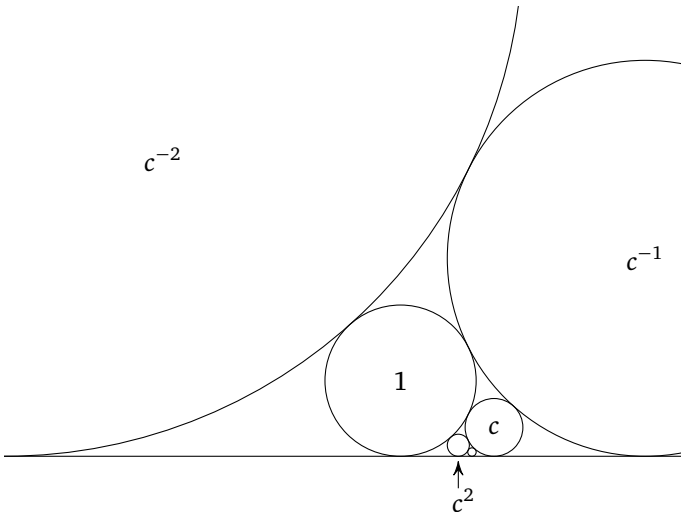
5.2. КОВРЫ С НЕОГРАНИЧЕННЫМИ РАЗМЕРАМИ КРУГОВ

Рассмотрим четверку q_1 попарно касающихся кругов на расширенной плоскости, один из которых является нижней полуплоскостью, а кривизны граничных окружностей остальных трех образуют геометрическую прогрессию. Запишем четыре кривизны в форме $0 < x^{-1} < 1 < x$. Из уравнения Декарта следует, что число x удовлетворяет уравнению

$$(x + 1 + x^{-1})^2 = 2(x^2 + 1 + x^{-2}), \quad \text{или} \quad x^2 - 2(x + x^{-1}) + x^{-2} = 1. \quad (5.2.1)$$

Полагая $y := x + x^{-1}$, мы получаем квадратное уравнение для y : $y^2 - 2y - 3 = 0$. Оно имеет два корня: 1 и 3. Только второе значение y приводит к вещественному значению x . А именно, $x = \frac{3 + \sqrt{5}}{2} = \frac{2}{3 - \sqrt{5}}$, что совпадает с величиной c , введенной в схолии G.

Ковер \mathcal{A}_1 , порожденный четверкой q_1 , имеет следующее свойство: после растяжения в c раз он переходит в свое зеркальное отражение в вертикальной прямой. А если его растянуть в c^2 раз, он переходит в ковер, получаемый из начального параллельным переносом в горизонтальном направлении, см. рис. 5.6 и 5.8 (здесь и далее на рисунках для части кругов указана кривизна их границ).

Рис. 5.6. Ковер \mathcal{A}_1

Это легко вывести из того, что ковер \mathcal{A}_1 инвариантен относительно преобразования $w \mapsto -c\bar{w}$. Последнее свойство вытекает из того, что \mathcal{A}_1 и $-c \cdot \bar{\mathcal{A}}_1$ имеют три общих попарно касающихся круга.

В частности, ковер \mathcal{A}_1 содержит последовательность кругов D_k с кривизнами границ c^k , $k \in \mathbb{Z}$. Соответствующие круги задаются равенствами

$$\left| c^k w + (-1)^k \frac{2}{\sqrt{5}} + i \right| \leq 1, \tag{5.2.2}$$

а соответствующие эрмитовы матрицы имеют вид

$$\begin{aligned} M_k &= \begin{pmatrix} \frac{4}{5}c^{-k} & (-1)^k \frac{2}{\sqrt{5}} - i \\ (-1)^k \frac{2}{\sqrt{5}} + i & c^k \end{pmatrix} = \\ &= \begin{pmatrix} (-\varphi)^{-k} & 0 \\ 0 & \varphi^k \end{pmatrix} \cdot \begin{pmatrix} \frac{4}{5} & \frac{2}{\sqrt{5}} - i(-1)^k \\ \frac{2}{\sqrt{5}} + i(-1)^k & 1 \end{pmatrix} \cdot \begin{pmatrix} (-\varphi)^k & 0 \\ 0 & \varphi^k \end{pmatrix}, \end{aligned} \tag{5.2.3}$$

где $\varphi := \sqrt{c} \approx 1,618034\dots$ — знаменитое золотое сечение (см. схемку G).

Каждое из равенств (5.2.2) и (5.2.3) имеет следствием тот факт, что преобразование $w \rightarrow -c \cdot \bar{w}$ переводит D_n в D_{n-1} и, следовательно, сохраняет ковер \mathcal{A}_1 .

УПРАЖНЕНИЕ 37. Найдите матрицу $g \in SL(2, \mathbb{C})$, которая переводит ковер \mathcal{A}_1 в ковер-ленту.

Подсказка. Найдите матрицу g , которая сохраняет вещественную ось и переводит круг D_0 в горизонтальную прямую. Покажите, что образы $g \cdot D_k$ расположены как на рис. 5.7.

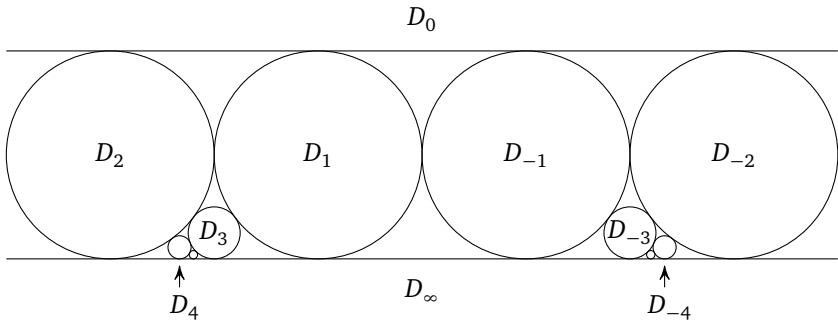


Рис. 5.7. Образ ковra \mathcal{A}_1 в ковre-ленте

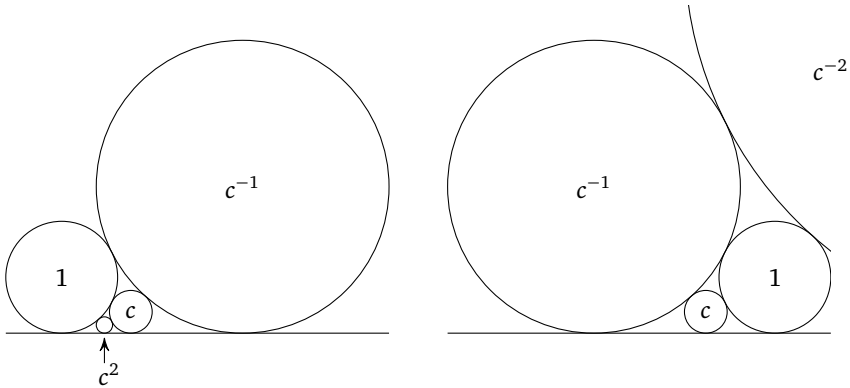


Рис. 5.8. Ковры \mathcal{A}_1 и $c \cdot \mathcal{A}_1$

Другой интересный ковер \mathcal{A}_2 с неограниченными кривизнами можно определить так. Попробуем найти четверку попарно касающихся кругов q_2 , для которой все четыре кривизны образуют геометрическую прогрессию $(1, \rho, \rho^2, \rho^3)$, где для определенности $\rho > 1$. Из уравнения Декарта получаем:

$$(1 + \rho + \rho^2 + \rho^3)^2 = 2(1 + \rho^2 + \rho^4 + \rho^6). \quad (5.2.4)$$

Упрощая это уравнение, перепишем его в форме

$$0 = 1 - 2\rho - \rho^2 - 4\rho^3 - \rho^4 - 2\rho^5 + \rho^6,$$

или

$$4 + (\rho + \rho^{-1}) + 2(\rho^2 + \rho^{-2}) = (\rho^3 + \rho^{-3}).$$

Введем новую неизвестную величину $u = \rho + \rho^{-1}$ и получим

$$4 + u + 2(u^2 - 2) = (u^3 - 3u), \quad \text{или} \quad u^3 - 2u^2 - 4u = 0.$$

Это уравнение имеет три корня: $u = 0, 1 - \sqrt{5}, 1 + \sqrt{5}$. Только последний корень дает вещественное значение для ρ , которое равно

$$\begin{aligned} \rho &= \varphi + \sqrt{\varphi} = \theta^2 + \theta \approx 2,890054\dots; \\ \rho^{-1} &= \varphi - \sqrt{\varphi} = \theta^2 - \theta \approx 0,346014\dots \end{aligned}$$

Так же, как и в предыдущем примере, начальные круги включаются в бесконечную двустороннюю последовательность D_k , образующую спираль. Когда $k \rightarrow -\infty$, спираль стремится к некоторой конечной точке a . Если мы возьмем a в качестве начала координат, наша

спираль будет инвариантна относительно умножения на некоторое комплексное число λ с $|\lambda| = \rho$. Обозначим аргумент λ через 2α . Тогда соответствующие матрицы M_k должны иметь вид

$$M_k = \begin{pmatrix} a\rho^k & be^{2ika} \\ \bar{b}e^{-2ika} & c\rho^{-k} \end{pmatrix}, \quad ac - |b|^2 = -1. \quad (5.2.5)$$

Условие, что круги D_k и D_{k+m} касаются, имеет вид $\det(M_k + M_{k+m}) = 0$. Это условие в действительности не зависит от k и равносильно уравнению

$$\frac{|b|^2}{ac} = \frac{\rho^m + \rho^{-m} + 2}{e^{im\alpha} + e^{-im\alpha} + 2}.$$

Положим $s = \frac{1}{2} \ln \rho$. Тогда правая часть уравнения будет

$$\frac{1 + \operatorname{ch} 2ms}{1 + \cos 2m\alpha} = \left(\frac{\operatorname{ch} ms}{\cos m\alpha} \right)^2.$$

Мы знаем, что D_0 касается D_m для $m = 1, 2, 3$. Таким образом, справедливы равенства

$$\frac{|b|}{\sqrt{ac}} = \frac{\operatorname{ch} s}{|\cos \alpha|} = \frac{\operatorname{ch} 2s}{|\cos 2\alpha|} = \frac{\operatorname{ch} 3s}{|\cos 3\alpha|}. \quad (5.2.6)$$

Поскольку $\operatorname{ch} 3s = \operatorname{ch} s(2\operatorname{ch} 2s - 1)$ и $\cos 3\alpha = \cos \alpha(2\cos 2\alpha - 1)$, мы заключаем, сравнивая второй и последний члены в (5.2.6), что

$$2\operatorname{ch} 2s - 1 = |2\cos 2\alpha - 1|.$$

Это возможно лишь при $2\cos 2\alpha - 1 < 0$. Поэтому мы получаем

$$2\operatorname{ch} 2s - 1 = 1 - 2\cos 2\alpha,$$

или $\operatorname{ch} 2s = 1 - \cos 2\alpha$, что возможно лишь при $\cos 2\alpha \leq 0$.

Используя соотношение $\operatorname{ch} 2s = 1 - \cos 2\alpha$, мы получаем, сравнивая второй и третий члены в (5.2.6),

$$\operatorname{ch} s = \pm \frac{\cos \alpha \cdot (1 - \cos 2\alpha)}{\cos 2\alpha}.$$

Теперь тождество $2\operatorname{ch}^2 s = \operatorname{ch} 2s + 1$ приводит к уравнению:

$$2\left(\frac{\cos \alpha \cdot (1 - \cos 2\alpha)}{\cos 2\alpha}\right)^2 = 2 - \cos 2\alpha.$$

Обозначим $\cos 2\alpha$ через x и перепишем уравнение в алгебраической форме:

$$\frac{(x+1)(1-x)^2}{x^2} = 2-x, \quad \text{или} \quad (x+1)(1-x)^2 = 2x^2 - x^3,$$

или

$$2x^3 - 3x^2 - x + 1 = 0.$$

Оно имеет корень $x = 1/2$, и это позволяет нам переписать его в простой форме $(2x-1)(x^2-x-1) = 0$. Таким образом, другие два корня — это φ и $-\varphi^{-1} = 1 - \varphi$. Только один из трех корней отрицателен: $x = -\varphi^{-1}$.

Окончательно получаем $\cos 2\alpha = -\varphi^{-1}$, $\operatorname{ch} 2s = \varphi$. Следовательно, $\rho + \rho^{-1} = 2\varphi$ и $\rho = \varphi + \sqrt{\varphi^2 - 1} = \theta^2 + \theta$. Заодно мы получаем $\frac{|b|}{\sqrt{ac}} = \varphi^2$, поэтому

$$|b|^2 = \frac{\varphi^2}{\sqrt{5}}, \quad ac = \frac{\varphi^{-2}}{\sqrt{5}}. \quad (5.2.7)$$

Таким образом, мы знаем матрицы M_k с точностью до комплексного сопряжения, а также сопряжения диагональной матрицей. Геометрически это значит, что мы знаем ковер \mathcal{A}_2 с точностью до поворотов, растяжений и отражений в прямой линии. В частности, можно положить

$$M_k = \frac{1}{\sqrt[4]{5}} \begin{pmatrix} \varphi^{-1} \cdot \rho^k & \varphi \cdot e^{2ika} \\ \varphi \cdot e^{2ika} & \varphi^{-1} \cdot \rho^{-k} \end{pmatrix}, \quad (5.2.8)$$

так что

$$D_0 = \left\{ w: \left| w + 1 + \frac{1}{\sqrt{5}} \right| \leq \sqrt{\frac{1+2\sqrt{5}}{5}} \right\}. \quad (5.2.9)$$

Далее, найдем число λ , определенное с точностью до комплексного сопряжения.

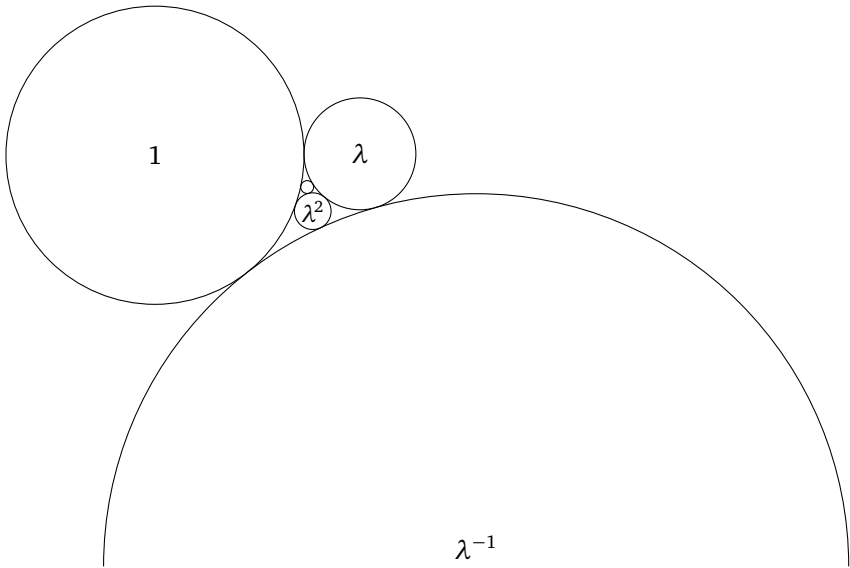
Мы имеем

$$2 \sin^2 \alpha = 1 - \cos 2\alpha = \varphi \quad \text{и} \quad 2 \cos^2 \alpha = 1 + \cos 2\alpha = 1 - \varphi^{-1} = \varphi^{-2}.$$

Поэтому $\sin^2 \alpha = \varphi^{-1}$ и $\sin 2\alpha = \pm \theta^{-1}$. Таким образом, мы имеем $e^{2i\alpha} = \cos 2\alpha + i \sin 2\alpha = -\varphi^{-1} \pm i\theta^{-1}$. Наконец,

$$\lambda = \rho e^{2i\alpha} = -(1 + \theta^{-1})(1 \mp i\theta). \quad (5.2.10)$$

Соответствующая картинка показана на рис. 5.9.

Рис. 5.9. Ковёр \mathcal{A}_2

5.3. ТРИ ИНТЕРПРЕТАЦИИ МНОЖЕСТВА \mathcal{D}

Символом $\mathbb{R}^{1,3}$ обычно обозначают 4-мерное вещественное векторное пространство с координатами t, x, y, z и с индефинитным¹ скалярным произведением

$$(p_1, p_2) = t_1 t_2 - x_1 x_2 - y_1 y_2 - z_1 z_2. \quad (5.3.1)$$

В физической литературе это пространство называют *пространством Минковского* в честь замечательного немецкого математика и физика Германа Минковского.

Вектор $p \in \mathbb{R}^{1,3}$ называется *временеподобным*, *световым* или *пространственноподобным*, если его квадрат положителен, равен нулю или отрицателен. Световые векторы, кроме того, делятся на векторы будущего, для которых $t > 0$, и векторы прошлого, для которых $t < 0$.

Физический смысл вектора p — это *событие*, которое происходит в момент времени t в точке $(x, y, z) \in \mathbb{R}^3$.

¹ Скалярное произведение называют индефинитным, если скалярный квадрат вектора может быть и положительным, и отрицательным.

Физики называют *полной группой Лоренца* L группу всех линейных обратимых преобразований пространства $\mathbb{R}^{1,3}$, сохраняющих скалярное произведение (5.3.1). Эта группа состоит из четырех связанных компонент; та из них, которая содержит единицу, сама является группой и называется *собственной группой Лоренца* L_0 . В математических текстах те же группы обозначаются $O(1, 3)$ и $SO_+(1, 3)$ соответственно.

Принцип относительности утверждает, что все физические законы инвариантны относительно собственной группы Лоренца.

Алгебраически элементы $g \in O(1, 3)$ задаются вещественными матрицами $\|g_{ij}\|$ формата 4×4 , у которых строки (столбцы) — попарно ортогональные векторы из $\mathbb{R}^{1,3}$, причем первая строка (первый столбец) имеет скалярный квадрат 1, а остальные строки (столбцы) имеют скалярный квадрат¹ -1 .

Мы кратко напомним факт, объясняемый в схолии F: матрица $g \in O(1, 3)$ принадлежит собственной группе Лоренца, если выполнены два дополнительных условия: $\det g = 1$ и $g_{0,0} > 0$.

Теперь мы покажем, как использовать пространство Минковского для параметризации множества \mathcal{D} всех кругов на двумерной сфере.

Каждый круг на S^2 можно определить как пересечение S^2 с подходящим полупространством $H_{u,\tau}$, задаваемым линейным (неоднородным) неравенством

$$H_{u,\tau} = \{v \in \mathbb{R}^3 : (u, v) + \tau \leq 0\}, \quad \text{где } u \in S^2 \text{ и } \tau \in (-1, 1). \quad (5.3.2)$$

Вместо пары $(u, \tau) \in S^2 \times (-1, 1)$ мы можем использовать один пространственноподобный вектор $p = (t, x, y, z) \in \mathbb{R}^{1,3}$, задаваемый формулой

$$p = \frac{1}{\sqrt{1+\tau^2}} \cdot (\tau, u).$$

В терминах этого вектора полупространство $H_{u,\tau}$ принимает вид

$$H_p = \{v \in \mathbb{R}^3 : xv^1 + yv^2 + zv^3 + t \leq 0\}. \quad (5.3.3)$$

Ясно, что $H_{p_1} = H_{p_2}$ тогда и только тогда, когда $p_1 = c \cdot p_2$ для некоторого $c > 0$. Поэтому мы можем нормализовать p условием $|p|^2 = -1$.

Итак, множество \mathcal{D} кругов на сфере отождествляется с множеством P_{-1} всех пространственноподобных векторов $p \in \mathbb{R}^{1,3}$ с $|p|^2 = -1$.

¹ Сравните со свойством ортогональных матриц: их строки (столбцы) попарно ортогональны и имеют длину 1.

Хорошо известно, что P_{-1} — однополостный трехмерный гиперboloид в \mathbb{R}^4 и что группа $L_0 \simeq SO_+(1, 3; \mathbb{R})$ действует транзитивно на нем. Стабилизатором точки $(0, 0, 0, 1)$ является группа $SO_+(1, 2; \mathbb{R})$, естественно вложенная в $SO_+(1, 3; \mathbb{R})$.

Это и есть первая интерпретация множества \mathcal{D} .

УПРАЖНЕНИЕ 38. Покажите, что гиперboloид $P_{-1} \subset \mathbb{R}^{1,3}$, определенный уравнением $|p|^2 = -1$, диффеоморфен $S^2 \times \mathbb{R}$.

Подсказка. Вспомните о параметрах u, τ , введенных выше.

Вторая интерпретация множества \mathcal{D} использует комплексные матрицы второго порядка. Мы начнем с неравенства (4.1.17) и построим из его коэффициентов матрицу $M = \begin{pmatrix} a & b \\ \bar{b} & c \end{pmatrix}$.

Напомним, что мы наложили на коэффициенты a, b, c условие $ac - |b|^2 < 0$. Поэтому M — эрмитова матрица с отрицательным определителем. Снова мы можем и будем нормализовать эту матрицу условием $\det M = -1$.

Таким образом, множество \mathcal{D} отождествляется с совокупностью H_{-1} всех эрмитовых матриц M формата 2×2 с $\det M = -1$.

УПРАЖНЕНИЕ 39. Покажите, что две построенные интерпретации связаны следующим образом: вектору $p = (t, x, y, z) \in \mathbb{R}^{1,3}$ соответствует матрица $M = \begin{pmatrix} a & b \\ \bar{b} & c \end{pmatrix}$, в которой

$$a = t - z, \quad b = x + iy, \quad c = t + z. \quad (5.3.4)$$

Подсказка. Сравните (4.1.5) и (4.1.17).

Третья интерпретация описывает \mathcal{D} как однородное пространство.

Мы уже знаем, что группа $G = PSL(2, \mathbb{C})$ действует на $\overline{\mathbb{C}}$ дробно-линейными преобразованиями. Более того, согласно предложению F.2, G действует на \mathcal{D} .

С другой стороны, группа $SL(2, \mathbb{C})$ действует на множестве H эрмитовых матриц второго порядка по правилу:

$$g: M \mapsto gMg^*. \quad (5.3.5)$$

Это действие сохраняет определитель матрицы и, следовательно, переводит в себя множество H_{-1} эрмитовых матриц с определителем -1 . (Фактически это действие группы G , поскольку центр \mathbb{C} группы $SL(2, \mathbb{C})$ действует тривиально.)

ТЕОРЕМА 5.2. Существует такой гомоморфизм $\pi: SL(2, \mathbb{C}) \rightarrow L_0 \simeq SO_+(1, 3; \mathbb{R})$, что следующая диаграмма коммутативна:

$$\begin{array}{ccccc}
 G & \times & \mathcal{D} & \longrightarrow & \mathcal{D} \\
 p \uparrow & & \uparrow \parallel & & \uparrow \parallel \\
 SL(2, \mathbb{C}) & \times & H_{-1} & \longrightarrow & H_{-1} \\
 \pi \downarrow & & \downarrow \parallel & & \downarrow \parallel \\
 L_0 & \times & P_{-1} & \longrightarrow & P_{-1}.
 \end{array}$$

Здесь p означает естественную проекцию $SL(2, \mathbb{C})$ на $PSL(2, \mathbb{C}) \simeq G$, а горизонтальные стрелки означают действие группы на множестве.

Мы оставляем читателю проверку деталей, но приведем явную формулу для гомоморфизма π .

УПРАЖНЕНИЕ 40. Покажите, что гомоморфизм π дается формулой:

$$\pi \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \frac{|a|^2 + |b|^2 + |c|^2 + |d|^2}{2} & \operatorname{Re}(a\bar{b} + c\bar{d}) & \operatorname{Im}(a\bar{b} + c\bar{d}) & \frac{|b|^2 - |a|^2 - |c|^2 + |d|^2}{2} \\ \operatorname{Re}(a\bar{c} + b\bar{d}) & \operatorname{Re}(a\bar{d} + b\bar{c}) & \operatorname{Im}(a\bar{d} - b\bar{c}) & \operatorname{Re}(b\bar{d} - a\bar{c}) \\ \operatorname{Im}(a\bar{c} + b\bar{d}) & \operatorname{Im}(a\bar{d} + b\bar{c}) & \operatorname{Re}(a\bar{d} - b\bar{c}) & \operatorname{Im}(b\bar{d} - a\bar{c}) \\ \frac{|c|^2 - |a|^2 - |b|^2 + |d|^2}{2} & \operatorname{Re}(c\bar{d} - a\bar{b}) & \operatorname{Im}(c\bar{d} - a\bar{b}) & \frac{|a|^2 - |b|^2 - |c|^2 + |d|^2}{2} \end{pmatrix}.$$

ЗАМЕЧАНИЕ. Обратное отображение $SO_+(1, 3; \mathbb{R}) \rightarrow PSL(2, \mathbb{C})$ также определено, но поднять его до отображения в $SL(2, \mathbb{C})$ можно только с точностью до знака. Это — так называемое *спиновое представление* группы $SO_+(1, 3; \mathbb{R})$.

В частности, все попарные произведения матричных элементов вида $2a\bar{a}$, $2a\bar{b}$, ... и т. д. однозначно определены и приводятся в таблице 5.1. ♡

УПРАЖНЕНИЕ 41. Опишите образ при гомоморфизме π следующих подгрупп группы G : а) $PGL(2, \mathbb{R})$; б) $PSU(2, \mathbb{C})$; в) $PSU(1, 1; \mathbb{C})$.

Подсказка. Используйте тот факт, что все эти подгруппы являются стабилизаторами некоторых векторов.

ОТВЕТЫ: а) $\pi(PGL(2, \mathbb{R})) = \operatorname{Stab}(0, 0, 1, 0) \simeq SO_+(1, 2; \mathbb{R})$;

б) $\pi(PSU(2, \mathbb{C})) = \operatorname{Stab}(1, 0, 0, 0) \simeq SO(3, \mathbb{R})$;

в) $\pi(PSU(1, 1; \mathbb{C})) = \operatorname{Stab}(0, 0, 0, 1) \simeq SO_+(1, 2; \mathbb{R})$.

Интересной проблемой является сравнение π -образа подгруппы $SL(2, \mathbb{Z} + i\mathbb{Z}) \subset SL(2, \mathbb{C})$ с подгруппой $SO_+(1, 3; \mathbb{Z}) \subset SO_+(1, 3)$.

Таблица 5.1

Попарные произведения матричных элементов

	\bar{a}	\bar{b}
2a	$g_{00} - g_{03} - g_{30} + g_{33}$	$g_{01} - g_{31} + i(g_{32} - g_{02})$
2b	$g_{01} - g_{31} + i(g_{02} - g_{32})$	$g_{00} + g_{03} - g_{30} - g_{33}$
2c	$g_{10} - g_{13} + i(g_{23} - g_{20})$	$g_{11} - g_{22} - i(g_{12} + g_{21})$
2d	$g_{11} + g_{22} + i(g_{12} - g_{21})$	$g_{10} + g_{13} - i(g_{20} + g_{23})$
	\bar{c}	\bar{d}
2a	$g_{10} - g_{13} + i(g_{20} - g_{23})$	$g_{11} + g_{22} + i(g_{21} - g_{12})$
2b	$g_{11} - g_{22} + i(g_{12} + g_{21})$	$g_{10} + g_{13} + i(g_{20} + g_{23})$
2c	$g_{00} - g_{03} + g_{30} - g_{33}$	$g_{01} + g_{31} - i(g_{02} + g_{32})$
2d	$g_{01} + g_{31} + i(g_{02} + g_{32})$	$g_{00} + g_{03} + g_{30} + g_{33}$

5.4. ОБОБЩЕННАЯ ТЕОРЕМА ДЕКАРТА

Пусть D_i , $1 \leq i \leq 4$, — четыре попарно касающихся круга на сфере. Обозначим через p_i соответствующие пространственноподобные векторы с $|p_i|^2 = -1$ и через M_i соответствующие эрмитовы матрицы с $\det M_i = -1$.

Лемма 5.3. *Круги D_1 и D_2 касаются тогда и только тогда, когда выполняются следующие эквивалентные условия:*

- а) $p_1 + p_2$ — световой вектор будущего;
- б) $(p_1, p_2) = 1$ и $p_1 + p_2$ имеет положительную t -координату;
- в) $\det(M_1 + M_2) = 0$ и $\text{tr}(M_1 + M_2) > 0$.

Доказательство. Сначала мы покажем, что ориентированные окружности $C_i = \partial D_i$, $i = 1, 2$, отрицательно (соответственно положительно) касаются тогда и только тогда, когда $|p_1 \pm p_2|^2 = 0$, или, что эквивалентно, когда $\det(M_1 \pm M_2) = 0$.

Используя подходящее мебиусово преобразование, мы можем предполагать, что C_1 — это вещественная ось со стандартной ориентацией. Соответствующие вектор и матрица имеют вид $p_1 = (0, 0, -1, 0)$ и $M_1 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$.

Пусть C_2 — ориентированная окружность, касающаяся C_1 . Обозначим точку касания через a . Тогда преобразование $w \mapsto \frac{c}{a-w}$ с вещественным параметром c сохраняет C_1 и при подходящем c переводит C_2 в горизонтальную прямую $2i + \mathbb{R}$ с некоторой ориента-

цией. Соответствующие вектор и матрица будут $p_2 = \pm(1, 0, -1, 1)$ и $M_2 = \pm \begin{pmatrix} 2 & -i \\ i & 0 \end{pmatrix}$, где знак плюс соответствует стандартной ориентации, а минус — противоположной. Мы видим, что условия леммы выполняются.

Обратно, если условия леммы выполняются, то мы можем сделать такое преобразование Мёбиуса, что векторы p_1 и p_2 примут указанную выше форму. Тогда данные окружности касаются нужным образом.

Доказательство леммы следует той же схеме. Заметим только, что знак координаты t для световых векторов сохраняется под действием группы G и то же самое справедливо для следа $\text{tr } M$, если $\det M = 0$. \square

Вернемся к теореме Декарта. Рассмотрим матрицу Грама скалярных произведений векторов p_i . Согласно лемме 5.3 она дается формулой:

$$G_{ij} := (p_i, p_j) = 1 - 2\delta_{ij}. \quad (5.4.1)$$

Хорошо известно, что определитель матрицы Грама равен квадрату определителя матрицы, составленной из координат этих векторов. То же самое верно с точностью до знака и в псевдоевклидовом пространстве, например, в $\mathbb{R}^{1,3}$.

В нашем случае квадрат матрицы G равен $4 \cdot 1_4$ и $\det G = 16$. Отсюда следует, что векторы p_i линейно независимы и, следовательно, образуют базис в $\mathbb{R}^{1,3}$.

Для любого вектора $v \in \mathbb{R}^{1,3}$ мы определим два сорта координат:

- ковариантные координаты v_i ,
- контравариантные координаты v^j

по формулам

$$v_i = (v, p_i); \quad v = \sum_{j=1}^4 v^j \cdot p_j. \quad (5.4.2)$$

Найдем соотношения между этими координатами. Подставляя второе равенство (5.4.2) в первое и учитывая (5.4.1), мы выводим:

$$v_i = \left(\sum_{j=1}^4 v^j \cdot p_j, p_i \right) = \sum_{j=1}^4 G_{ij} v^j = \sum_{j=1}^4 v^j - 2v^i. \quad (5.4.3)$$

Суммируя последнее равенство по i , мы получаем:

$$\sum_{i=1}^4 v_i = 4 \sum_{j=1}^4 v^j - 2 \sum_{i=1}^4 v^i = 2 \sum_{j=1}^4 v^j$$

и, окончательно,

$$v^j = \frac{1}{2} \sum_{i=1}^4 v_i - \frac{1}{2} v_j. \quad (5.4.4)$$

Из (5.4.3) мы также выводим выражение для $|v|^2$ в терминах координат:

$$|v|^2 = \left(\sum_j v^j \right)^2 - 2 \sum_j (v^j)^2 = \frac{1}{4} \left(\sum_i v_i \right)^2 - \frac{1}{2} \sum_i v_i^2. \quad (5.4.5)$$

Отсюда следует, что для любого светового вектора v мы имеем равенство

$$\left(\sum_i v_i \right)^2 - 2 \sum_i v_i^2 = 0. \quad (5.4.6)$$

Положим, в частности, $v = (1, 0, 0, -1)$. Тогда $v_i = (v, p_i) = t_i + z_i = c_i$, и (5.4.6) дает в точности утверждение теоремы Декарта. Если же положить $v = (1, 0, 0, 1)$, то мы получим

$$\left(\sum_{i=1}^4 a_i \right)^2 = 2 \sum_{i=1}^4 a_i^2.$$

Тот же подход позволяет доказать больше.

ТЕОРЕМА 5.3 (обобщенная теорема Декарта). *Матрицы M_i , отвечающие четырем попарно касающимся кругам, удовлетворяют соотношению*

$$\left(\sum_i M_i \right)^2 - 2 \sum_i M_i^2 = -8 \cdot 1_4. \quad (5.4.7)$$

Доказательство. Введем скалярное произведение в пространстве эрмитовых матриц формата 2×2 , соответствующее квадратичной форме $Q(M) := \det M$. Явный вид этого произведения такой:

$$(M_1, M_2) = \frac{\det(M_1 + M_2) - \det M_1 - \det M_2}{2}. \quad (5.4.8)$$

В частности, мы имеем: $(M, 1) = \frac{\det(M+1) - \det M - 1}{2} = \frac{1}{2} \operatorname{tr} M$.

Напомним также тождество Кэли, справедливое для любых матриц и в случае матриц порядка 2 имеющее вид

$$M^2 - M \cdot \operatorname{tr} M + \det M \cdot 1 = 0. \quad (5.4.9)$$

Пусть теперь M_1, M_2, M_3, M_4 — четыре эрмитовых матрицы, соответствующие четырем попарно касающимся кругам и нормализованные условием $\det M_i = -1$. Тогда (5.4.9) принимает вид

$$M_i^2 = M_i \cdot \operatorname{tr} M_i + 1. \quad (5.4.10)$$

Введем обозначения:

$$\Sigma_1 := \sum_{i=1}^4 M_i, \quad \Sigma_2 := \sum_{i=1}^4 M_i^2.$$

Мы видели выше, что $(M_i, M_j) = (p_i, p_j) = 1 - 2\delta_{ij}$. В частности, отсюда следует, что $(\Sigma_1, M_i) = 2$ и $(\Sigma_1, \Sigma_1) = 8$. Далее, взяв скалярное произведение обеих частей равенства (5.4.10) с M_j и суммируя по i , мы получаем

$$(\Sigma_2, M_j) = \operatorname{tr} \Sigma_1. \quad (5.4.11)$$

С другой стороны, мы имеем $\Sigma_1^2 = \Sigma_1 \cdot \operatorname{tr} \Sigma_1 - 8 \cdot 1$. Взяв опять скалярное произведение с M_j , мы получим

$$(\Sigma_1^2, M_j) = 2 \operatorname{tr} \Sigma_1 - 4 \operatorname{tr} M_j. \quad (5.4.12)$$

Вычитая из (5.4.12) удвоенное (5.4.11), окончательно получаем

$$(\Sigma_1^2 - 2\Sigma_2, M_j) = -8(1, M_j), \quad \text{или} \quad (\Sigma_1^2 - 2\Sigma_2 + 8 \cdot 1, M_j) = 0.$$

Поскольку матрицы M_i образуют базис в пространстве всех эрмитовых матриц второго порядка, мы получаем искомое равенство (5.4.7). \square

Соотношение (5.4.7) можно рассматривать как матричную форму теоремы Декарта. Оно дает нам информацию не только о радиусах касающихся кругов, но и об их взаимном расположении. Отметим одно следствие, полезное в вычислениях.

ТЕОРЕМА 5.4. Пусть D_+ и D_- — два касающихся круга на плоскости. Предположим, что некоторая последовательность кругов $\{D_k\}$, $k \in \mathbb{Z}$, обладает следующим свойством: каждый круг D_k касается данных кругов D_{\pm} , а также соседних кругов $D_{k \pm 1}$.

Тогда последовательность соответствующих эрмитовых матриц $\{M_k\}$, $k \in \mathbb{Z}$, зависит квадратично от параметра k :

$$M_k = A \cdot k^2 + B \cdot k + C, \quad (5.4.13)$$

где $A = M_+ + M_-$, $B = \frac{M_1 - M_{-1}}{2}$, $C = M_0$.

Иллюстрации к этой теореме вы можете видеть на рис. 5.10 и 6.5.

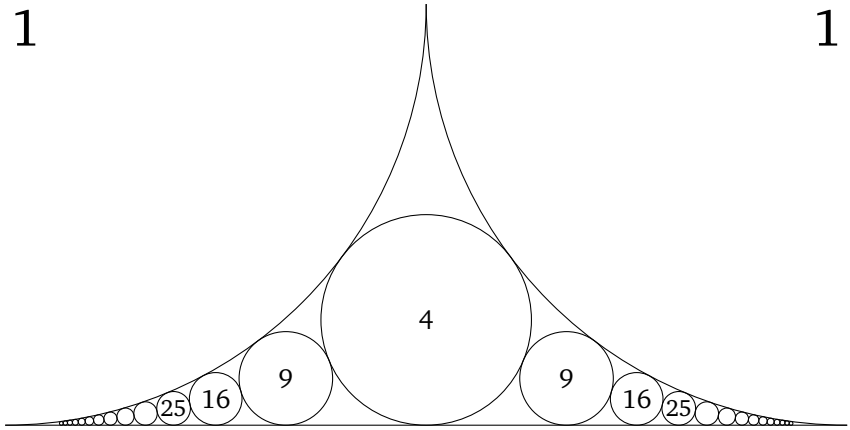


Рис. 5.10. Квадратичные последовательности кривизн

Глава 6

АРИФМЕТИЧЕСКИЕ СВОЙСТВА КОВРОВ АПОЛЛОНИЯ

Мы рассмотрим здесь некоторые теоретико-числовые вопросы, возникающие при изучении кривизн граничных окружностей для кругов, составляющих данный ковер Аполлония.

6.1. ЦЕЛОЧИСЛЕННЫЕ РЕШЕНИЯ УРАВНЕНИЯ ДЕКАРТА

Сначала рассмотрим арифметические свойства множества решений уравнения Декарта (4.1.3). Сделаем замену переменных:

$$\begin{aligned} t &= \frac{c_0 + c_1 + c_2 + c_3}{2}, & x &= \frac{c_0 + c_1 - c_2 - c_3}{2}, \\ y &= \frac{c_0 - c_1 + c_2 - c_3}{2}, & z &= \frac{c_0 - c_1 - c_2 + c_3}{2}. \end{aligned} \quad (6.1.1)$$

Тогда мы получим:

$$t^2 - x^2 - y^2 - z^2 = \frac{(c_0 + c_1 + c_2 + c_3)^2}{2} - (c_0^2 + c_1^2 + c_2^2 + c_3^2),$$

и уравнение (4.1.3) примет вид

$$t^2 - x^2 - y^2 - z^2 = 0. \quad (6.1.2)$$

Другими словами, решения уравнения Декарта в новых координатах являются световыми векторами в пространстве Минковского.

Лемма 6.1. Целочисленные решения уравнения Декарта соответствуют целочисленным световым векторам в $\mathbb{R}^{1,3}$, то есть световым векторам с целыми координатами.

Доказательство. Из (4.1.3) ясно, что суммы $c_0 \pm c_1 \pm c_2 \pm c_3$ принимают всегда четные значения. Поэтому каждому целочисленному решению (4.1.3) соответствует световой вектор с целыми координатами. Обратное, из (6.1.2) следует, что суммы $t \pm x \pm y \pm z$ всегда четны. Поэтому из равенств

$$\begin{aligned} c_0 &= \frac{t + x + y + z}{2}, & c_1 &= \frac{t + x - y - z}{2}, \\ c_2 &= \frac{t - x + y - z}{2}, & c_3 &= \frac{t - x - y + z}{2} \end{aligned}$$

мы выводим, что каждый целочисленный световой вектор соответствует целочисленному решению (4.1.3). \square

Таким образом, мы приходим к задаче:

Задача 8. Описать множество целых точек на световом конусе в $\mathbb{R}^{1,3}$.

Решение аналогичной задачи для рациональных точек хорошо известно и довольно просто. Каждой рациональной точке (t, x, y, z) светового конуса соответствует рациональная точка $\left(\frac{x}{t}, \frac{y}{t}, \frac{z}{t}\right)$ единичной сферы S^2 . Стереографическая проекция на $\overline{\mathbb{C}}$ переводит эту точку в $\frac{x+iy}{t-z} \in P^1(\mathbb{Q}[i])$.

Обратно, любая рациональная точка $(r+is) \in P^1(\mathbb{Q}[i])$ происходит из рациональной точки

$$\left(\frac{2r}{r^2+s^2+1}, \frac{2s}{r^2+s^2+1}, \frac{r^2+s^2-1}{r^2+s^2+1}\right) \in S^2.$$

Полагая $r = \frac{k}{n}, s = \frac{m}{n}$, мы видим, что каждый целочисленный световой вектор пропорционален (но не обязательно равен!) вектору вида

$$t = k^2 + m^2 + n^2, \quad x = 2kn, \quad y = 2mn, \quad z = k^2 + m^2 - n^2 \quad (6.1.3)$$

с целыми k, m, n .

Заметим, что для каждого целочисленного светового вектора p все его кратные $np, n \in \mathbb{Z}$, — также целочисленные световые векторы. Поэтому мы можем ограничиться изучением *примитивных* векторов, для которых наибольший общий делитель его координат равен 1.

Лемма 6.2. *Каждый примитивный целочисленный световой вектор имеет нечетную координату t и ровно одну нечетную координату среди x, y, z .*

Доказательство. Если t четно, то $x^2 + y^2 + z^2$ делится на 4. Поскольку каждый квадрат имеет вычет 0 или 1 mod 4, мы видим, что все координаты x, y, z должны быть четны. Но тогда p не примитивен.

Если t нечетно, то $x^2 + y^2 + z^2 \equiv 1 \pmod{4}$. Отсюда следует, что в точности одно из чисел x, y, z нечетно. \square

Задача 9. Найти удобную параметризацию всех примитивных целочисленных световых векторов.

Например, допустим, что t, z нечетны, а x, y — четны. Верно ли, что в этом случае имеет место равенство (6.1.3)?

Теперь рассмотрим подгруппу Γ группы Лоренца, которая сохраняет множество целочисленных световых векторов.

УПРАЖНЕНИЕ 42. Покажите, что подгруппа Γ совпадает с подгруппой $SO_+(1, 3; \mathbb{Z})$, состоящей из целочисленных матриц в $SO_+(1, 3; \mathbb{R})$.

Подсказка. Пусть $g \in \Gamma$. Покажите, что сумма и разность любых двух столбцов g является целочисленным вектором. Покажите, что целочисленный световой вектор имеет хотя бы одну четную координату.

Таким образом, группа Γ действует на множестве целочисленных световых векторов и сохраняет множество P примитивных векторов.

УПРАЖНЕНИЕ 43. а) Найдите индекс подгруппы $PSL(2, \mathbb{Z}[i])$ в $PGL(2, \mathbb{Z}[i])$.

б)* Каковы образы этих подгрупп в $O_+(1, 3; \mathbb{R})$?

УПРАЖНЕНИЕ 44. Покажите, что введенный выше гомоморфизм $\pi: PGL(2, \mathbb{C}) \rightarrow SO_+(1, 3; \mathbb{R})$ может быть продолжен до гомоморфизма $\tilde{\pi}: \bar{G} \rightarrow O_+(1, 3; \mathbb{R})$.

Подсказка. Покажите, что диагональная матрица $\text{diag}(1, 1, -1, 1)$ может быть выбрана в качестве образа комплексного сопряжения s при гомоморфизме $\tilde{\pi}$.

Задача 10. Описать Γ -орбиты в P .

Схолия Н. СТРУКТУРА НЕКОТОРЫХ ГРУПП, ПОРОЖДЕННЫХ ОТРАЖЕНИЯМИ

Теория групп, порожденных отражениями, — большая и очень интересная область современной математики. Мы приведем здесь только несколько фактов из этой теории, которые нужны нам в связи с ковром Аполлония.

Сначала мы опишем структуру так называемой *свободной* группы F_n с n образующими x_1, x_2, \dots, x_n . Эта группа однозначно определяется следующим свойством универсальности.

Для любой группы G с n образующими y_1, y_2, \dots, y_n существует и единствен гомоморфизм $\alpha: F_n \rightarrow G$, для которого $\alpha(x_i) = y_i$, $1 \leq i \leq n$.

Это определение имеет много достоинств (как мы увидим ниже), но не является эффективным. Единственность такой группы легко выводится из определения. В самом деле, предположим, что имеются две такие группы: F_n с образующими x_1, x_2, \dots, x_n и F'_n с образующими x'_1, x'_2, \dots, x'_n . Тогда из свойства универсальности мы выводим, что существуют гомоморфизмы $\alpha: F_n \rightarrow F'_n$ и $\alpha': F'_n \rightarrow F_n$ такие, что $\alpha(x_i) = x'_i$ и $\alpha'(x'_i) = x_i$. Рассмотрим композицию $\alpha' \circ \alpha$. Это — гомоморфизм F_n на себя, сохраняющий образующие. Согласно свойству универсальности, такой гомоморфизм должен быть тождественным. То же самое верно для композиции $\alpha \circ \alpha'$. Следовательно, F_n и F'_n изоморфны.

Покажем теперь, что группа с искомым свойством действительно существует. Для этого мы рассмотрим совокупность W_n всех слов в алфавите $x_1, x_1^{-1}, \dots, x_n, x_n^{-1}$, удовлетворяющих условию

(*) буквы x_i и x_i^{-1} не могут стоять рядом.

Обозначим через $l(w)$ длину слова w , то есть число букв в этом слове. Пусть $W_n^{(k)}$ означает множество слов длины k в W_n . Ясно, что $W_n^{(0)}$ содержит только пустое слово, а $W_n^{(1)}$ состоит из $2n$ однобуквенных слов — образующих и обратных к ним.

УПРАЖНЕНИЕ 45. Покажите, что $\#(W_n^{(k)}) = 2n(2n - 1)^{k-1}$ для $k \geq 1$.

Мы хотим определить на W_n структуру группы. Для этого мы определим произведение двух слов w_1, w_2 индуктивно по длине первого сомножителя. А именно, если $l(w_1) = 0$, мы полагаем $w_1 w_2 := w_2$.

Теперь предположим, что произведение уже определено для всех пар w_1, w_2 с $l(w_1) < k$, и рассмотрим случай $l(w_1) = k \geq 1$. Пусть последней буквой слова w_1 является $x_i^{\varepsilon_1}$, $1 \leq i \leq n$, $\varepsilon_1 = \pm 1$, а первой буквой слова w_2 является $x_j^{\varepsilon_2}$, $1 \leq j \leq n$, $\varepsilon_2 = \pm 1$.

Если $i \neq j$ или $i = j$, $\varepsilon_1 + \varepsilon_2 \neq 0$, мы определяем произведение $w_1 w_2$ просто приписывая слово w_2 справа к слову w_1 . Это новое слово имеет длину $l(w_1) + l(w_2)$ и удовлетворяет условию (*).

Если же $i = j$ и $\varepsilon_1 + \varepsilon_2 = 0$, мы обозначим через \tilde{w}_1 (соответственно через \tilde{w}_2) слово, получаемое из w_1 отбрасыванием последней буквы (соответственно слово, получаемое из w_2 отбрасыванием первой буквы). После этого мы полагаем $w_1 w_2 := \tilde{w}_1 \tilde{w}_2$.

Например, если $w_1 = x_1$, $w_2 = x_1^{-1} x_2$, то $\tilde{w}_1 = \emptyset$, $\tilde{w}_2 = x_2$ и $w_1 w_2 = x_2$.

Из этого определения легко выводится, что всегда $l(w_1 w_2) \leq l(w_1) + l(w_2)$ и $l(w_1 w_2) \equiv l(w_1) + l(w_2) \pmod{2}$.

Теперь мы должны проверить, что W_n действительно является группой относительно введенной операции умножения. Это значит, что нужно проверить ассоциативность умножения, наличие единичного и обратного элемента. Первое делается опять индукцией по длине среднего сомножителя; роль единичного элемента играет пустое слово, а обратный элемент получается, если написать данное слово в обратном порядке и заменить все показатели на противоположные.

Традиционно этот труд поручается читателю.

Проверим теперь, что построенная таким образом группа обладает свойством универсальности. Пусть G — любая группа, порожденная x_1, x_2, \dots, x_n . Тогда существует единственный гомоморфизм

$\alpha: W_n \rightarrow G$, для которого $\alpha(\{x_i\}) = x_i$. (Здесь $\{x_i\}$ означает однобуквенное слово.) А именно, слово $w = x_{i_1}^{\varepsilon_1} x_{i_2}^{\varepsilon_2} \dots x_{i_k}^{\varepsilon_k}$ обязано переходить в $\alpha(w) = x_{i_1}^{\varepsilon_1} \cdot x_{i_2}^{\varepsilon_2} \cdot \dots \cdot x_{i_k}^{\varepsilon_k}$, где знак \cdot означает умножение в группе G . С другой стороны, легко проверить, что определенное таким образом отображение α является гомоморфизмом W_n на G .

Мы доказали существование свободной группы F_n и заодно доказали

Предложение Н.1. *Любой элемент F_n единственным образом записывается в виде произведения*

$$g = x_{i_1}^{\varepsilon_1} x_{i_2}^{\varepsilon_2} \dots x_{i_k}^{\varepsilon_k}, \quad (6.1.1)$$

удовлетворяющего условию (*).

Нам понадобится еще другое семейство групп Γ_n , $n \geq 1$, которые свободно порождены n инволюциями s_1, \dots, s_n . По определению, группа Γ_n имеет следующее свойство универсальности:

Для любой группы G , порожденной n инволюциями t_1, \dots, t_n , существует единственный гомоморфизм α группы Γ_n на G , для которого $\alpha(s_i) = t_i$, $1 \leq i \leq n$.

Существование и единственность (с точностью до изоморфизма) группы Γ_n доказывается аналогично тому, как мы это сделали для F_n . Единственное отличие состоит в том, что теперь множество W_n состоит из всех слов в алфавите s_1, \dots, s_n без повторения букв подряд.

Предложение Н.2. *Любой элемент Γ_n однозначно записывается в виде произведения*

$$g = s_{i_1} s_{i_2} \dots s_{i_k}, \quad k \geq 0, \quad \text{где } i_a \neq i_{a+1} \text{ для } 1 \leq a \leq k-1. \quad (6.1.2)$$

Упражнение 46. а) Покажите, что в этом случае

$$\#(W_n^{(k)}) = \begin{cases} 1 & \text{при } k = 0, \\ n(n-1)^{k-1} & \text{при } k \geq 1. \end{cases} \quad (6.1.3)$$

б) Покажите, что Γ_n изоморфна F_n/J , где F_n — свободная группа с образующими s_1, \dots, s_n , а J — наименьшая нормальная подгруппа, содержащая s_1^2, \dots, s_n^2 .

Теорема Н.1. *Всякая нетривиальная (то есть отличная от e) инволюция в Γ_n сопряжена в точности одной образующей s_1, \dots, s_n .*

Доказательство. Пусть $g \in \Gamma_n$ — инволюция. Согласно предложению Н.2, она может быть записана в виде $g = s_{i_1} s_{i_2} \dots s_{i_n}$. Тогда $g^{-1} = s_{i_n} s_{i_{n-1}} \dots s_{i_1}$. Но $g^{-1} = g$, поэтому $s_{i_{n-k}} = s_{i_{k+1}}$ для $k = 0, 1, \dots, n-1$.

Если $n = 2k$ чётно, отсюда следует, что $k = 0$ и $g = e$.

Если $n = 2k - 1$ нечётно, мы получаем $g = ws_{i_k}w^{-1}$, где $w = s_{i_1} \dots s_{i_{k-1}}$. Следовательно, g сопряжена с s_{i_k} .

Наконец, покажем, что образующая s_i не сопряжена с s_j при $i \neq j$. Предположим противное. Тогда существует слово w такое, что $ws_i = s_jw$. Пусть w_0 — наиболее короткое из таких слов. Из равенства $w_0s_i = s_jw_0$ мы заключаем, что первой буквой w_0 является s_j , а последней — s_i , так что $w_0 = s_jw's_i$ для некоторого слова w' . Но тогда $s_jw' = w's_i$, что противоречит выбору w_0 , поскольку $l(w') = l(w_0) - 2 < l(w_0)$. \square

Для малых значений n группа Γ_n допускает более простое описание. Например, Γ_1 совпадает с группой $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$ порядка 2.

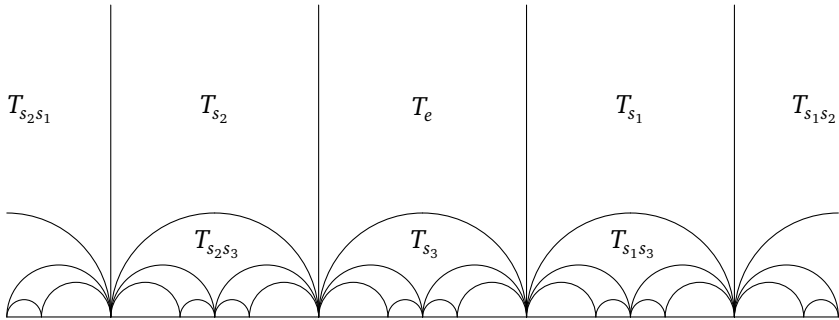
Группа Γ_2 изоморфна группе $\text{Aff}(1, \mathbb{Z})$ аффинных преобразований целочисленной решетки. Она имеет матричную реализацию вида $\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$, где $a = \pm 1$, $b \in \mathbb{Z}$. Мы оставляем читателю проверку того, что матрицы $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ и $\begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix}$ могут быть выбраны в качестве образующих инволюций s_1, s_2 .

Для $n = 3$ группа Γ_3 может быть реализована как дискретная группа преобразований, действующая на плоскости Лобачевского (или гиперболической плоскости) L_2 . Рассмотрим, например, модель Пуанкаре, изображающую L_2 в виде верхней полуплоскости $y > 0$ (см. схематично J ниже).

Три образующие инволюции для Γ_3 — это отражения в трех попарно касающихся зеркалах. Например, мы можем выбрать в качестве этих зеркал единичную (полу)окружность M_0 и две вертикальные прямые (два луча) $M_{\pm 1}$: $x = \pm 1$, касающиеся M_0 . Эти три зеркала ограничивают на плоскости Лобачевского треугольник T конечной площади с тремя бесконечно удаленными вершинами. Для любого слова w , составленного из s_1, s_2, s_3 без повторений букв, обозначим через T_w образ T под действием элемента $\gamma \in \Gamma_3$, соответствующего слову w .

Можно доказать, используя индукцию по $l(w)$, что все треугольники T_w различны, не имеют общих внутренних точек и покрывают всю плоскость Лобачевского. (См. рис. 6.1.)

Случай $n = 4$ — более сложный. Однако именно этот случай встречается в нашем исследовании. Более того, группа Γ_4 возникает при описании двух различных множеств: кругов, составляющих ковер Аполлония \mathcal{A} , и четверок попарно касающихся кругов из \mathcal{A} . Мы уже рассматривали эту ситуацию в разделе 5.1.

Рис. 6.1. Действие Γ_3 на L_2 6.2. СТРУКТУРА МНОЖЕСТВА $\overline{\mathbb{Q}}$

Здесь мы рассмотрим в деталях строение рациональной проективной прямой $P^1(\mathbb{Q}) = \overline{\mathbb{Q}}$. Как множество, эта проективная прямая получается добавлением к множеству \mathbb{Q} рациональных чисел бесконечно удаленной точки ∞ . Иногда эту проективную прямую называют *рациональной окружностью*.

Сначала подумаем, как удобнее всего параметризовать $\overline{\mathbb{Q}}$. Каждое число $r \in \overline{\mathbb{Q}}$ можно записать в виде $r = \frac{p}{q}$, где $p, q \in \mathbb{Z}$ не обращаются в нуль одновременно. Но такая запись неоднозначна. Мы можем наложить дополнительное условие: $\text{НОД}(p, q) = 1$, то есть потребовать, чтобы p и q были взаимно просты, или, что то же, чтобы дробь $\frac{p}{q}$ была несократима.

Отображение $\alpha: (p, q) \rightarrow \frac{p}{q}$ является двукратным накрытием. А именно, прообраз точки $r = \frac{p}{q}$ состоит из двух пар взаимно простых чисел: (p, q) и $(-p, -q)$. К сожалению, нет никакого естественного способа выбрать единственного представителя из каждого прообраза. Хотя для всех конечных точек можно условиться считать $q > 0$, но для $\infty = \frac{\pm 1}{0}$ этот способ не годится.

Замечание 1. Для аналитически думающего читателя мы можем сказать, что ситуация здесь похожа на риманову поверхность двузначной функции $f(w) = \sqrt{w}$. Отображение $z \mapsto w = z^2$ имеет два прообраза для каждого $w \in \mathbb{C}^\times$, но эта двузначная функция не имеет аналитической (или хотя бы непрерывной) однозначной ветви. \heartsuit

Замечание 2. Замечательный способ занумеровать все положительные рациональные числа обнаружили недавно Нил Калкин и Херберт Вилф¹.

Пусть $b(n)$ — число разбиений $n \geq 0$ в сумму степеней двойки, в которых никакая степень не используется больше двух раз. Тогда отношение $r_n = \frac{b(n)}{b(n+1)}$ принимает каждое положительное рациональное значение в точности один раз! Начальный кусок этой нумерации выглядит так:

n	0	1	2	3	4	5	6	7	8	9	10	11	12
$b(n)$	1	1	2	1	3	2	3	1	4	3	5	2	5
r_n	1	$\frac{1}{2}$	2	$\frac{1}{3}$	$\frac{3}{2}$	$\frac{2}{3}$	3	$\frac{1}{4}$	$\frac{4}{3}$	$\frac{3}{5}$	$\frac{5}{2}$	$\frac{2}{5}$	$\frac{5}{3}$
n	13	14	15	16	17	18	19	20	21	22	23	24	25
$b(n)$	3	4	1	5	7	3	8	5	7	2	7	5	8
r_n	$\frac{3}{4}$	4	$\frac{1}{5}$	$\frac{5}{4}$	$\frac{7}{4}$	$\frac{3}{7}$	$\frac{8}{3}$	$\frac{5}{8}$	$\frac{7}{5}$	$\frac{2}{7}$	$\frac{7}{2}$	$\frac{5}{7}$	$\frac{8}{5}$
n	26	27	28	29	30	31	32	33					
$b(n)$	3	7	4	5	1	6	5	4					
r_n	$\frac{3}{8}$	$\frac{7}{3}$	$\frac{4}{7}$	$\frac{5}{4}$	$\frac{1}{5}$	6	$\frac{5}{6}$	$\frac{9}{5}$					

Интересно сравнить эту нумерацию с той, которая получается из рассмотрения рядов Фарея (см. ниже). ♡

Наш следующий шаг в изучении $\overline{\mathbb{Q}}$ — введение естественного расстояния между точками. Далее мы всегда предполагаем, что рациональные числа (включая ∞) записаны в виде несократимой дроби.

Назовем два числа $r_i = \frac{p_i}{q_i}$, $i = 1, 2$, из $\overline{\mathbb{Q}}$ близкими, если выполнены следующие эквивалентные условия:

$$\text{а) } |p_1q_2 - p_2q_1| = 1, \quad \text{б) } |r_1 - r_2| = \frac{1}{|q_1q_2|}. \quad (6.2.1)$$

Стоит отметить, что отношение близости не является² отношением эквивалентности: каждое целое число близко к бесконечности, но только соседние целые числа близки друг к другу.

¹ N. Kalkin, H. Wilf. Recounting the rationals // The American Mathematical Monthly. 2000. Vol. 107. P. 360–363.

² Как и в обычной жизни.

Теперь мы отметим, что группа $PGL(2, \mathbb{Z})$ действует на $\overline{\mathbb{Q}}$ дробно-линейными преобразованиями и это действие сохраняет отношения близости. Впоследствии мы будем часто использовать этот факт.

ЛЕММА 6.3. *Группа $PSL(2, \mathbb{Z})$ действует просто транзитивно на множестве X всех упорядоченных пар близких чисел. Группа $PGL(2, \mathbb{Z})$ действует на X транзитивно с нетривиальным стабилизатором точки, изоморфным \mathbb{Z}_2 .*

ДОКАЗАТЕЛЬСТВО. Предположим, что (r_1, r_2) — пара близких чисел. Пусть для определенности $r_i = \frac{p_i}{q_i}$, $i = 1, 2$. Мы должны проверить, что существует единственный элемент $\gamma \in PSL(2, \mathbb{Z})$, который переводит данную пару (r_1, r_2) в стандартную пару $(\infty, 0)$. Пусть $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ — представитель элемента γ в $SL(2, \mathbb{Z})$. Тогда мы имеем: $\gamma(0) = \frac{b}{d}$, $\gamma(\infty) = \frac{a}{c}$. Из условий $\gamma(\infty) = r_1$, $\gamma(0) = r_2$ следует, что $(a, c) = k_1 \cdot (p_1, q_1)$, $(b, d) = k_2 \cdot (p_2, q_2)$. Поэтому $1 = \det g = ad - bc = k_1 k_2 \cdot (p_1 q_2 - p_2 q_1)^{-1} = k_1 k_2$ и $k_1 = k_2 = \pm 1$. Значит, матрица $g = \pm \begin{pmatrix} p_1 & p_2 \\ q_1 & q_2 \end{pmatrix}$ определена с точностью до знака и определяет единственный элемент $PSL(2, \mathbb{Z})$.

Стабилизатор пары $(\infty, 0)$ в $PGL(2, \mathbb{Z})$ состоит из классов матриц $\begin{pmatrix} 1 & 0 \\ 0 & \pm 1 \end{pmatrix}$. □

УПРАЖНЕНИЕ 47. Найдите все рациональные числа, которые близки: а) к 0; б) к ∞ ; в) к 1.

Мы определим *расстояние* между точками $\overline{\mathbb{Q}}$ следующим образом. Для данных двух чисел $r' \neq r''$ назовем расстоянием между ними наименьшее $n \in \mathbb{Z}_+$, для которого существует цепочка

$$r' = r_0, r_1, \dots, r_{n-1}, r_n = r''$$

такая, что r_k близко к r_{k+1} для всех k . Мы обозначим это расстояние $d(r', r'')$. Если $r' = r''$, мы полагаем $d(r', r'') = 0$.

УПРАЖНЕНИЕ 48. а) Покажите, что $(\overline{\mathbb{Q}}, d)$ является дискретным метрическим пространством, на котором группа $PGL(2, \mathbb{Z})$ действует, сохраняя расстояние.

б) Найдите стабилизатор точки ∞ .

ОТВЕТ. б) Группа $\text{Aff}(1, \mathbb{Z})$ аффинных преобразований $r \mapsto ar + b$, $a = \pm 1$, $b \in \mathbb{Z}$.

УПРАЖНЕНИЕ 49. Вычислите расстояния:

а) $d(\infty, n)$; б) $d(0, n)$; в) $d\left(0, \frac{5}{8}\right)$.

ОТВЕТ. а) 1; б) 0 для $n = 0$; 1 для $n = \pm 1$; 2 для $|n| > 1$; в) 3.

УПРАЖНЕНИЕ 50. а) Покажите, что для всех $r', r'' \in \overline{\mathbb{Q}}$ расстояние $d(r', r'')$ конечно.

б) Ограничено ли пространство $\overline{\mathbb{Q}}$?

Ответ. а) См. теорему 6.2 ниже; в) нет.

Довольно интересные и нетривиальные задачи возникают при изучении геометрии шаров и сфер в $\overline{\mathbb{Q}}$. Как обычно, мы определяем шар с центром a и радиусом r как множество

$$B_r(a) = \{b \in \overline{\mathbb{Q}} : d(a, b) \leq r\}.$$

Аналогично, сфера с тем же центром и радиусом — это множество

$$S_r(a) = \{b \in \overline{\mathbb{Q}} : d(a, b) = r\}.$$

ТЕОРЕМА 6.2. Шар $B_n(\infty)$ состоит из ∞ и всех рациональных чисел, которые могут быть записаны в виде непрерывной дроби длины $\leq n$, то есть как

$$r = k_1 + \frac{1}{k_2 + \frac{1}{k_3 + \dots + \frac{1}{k_n}}}, \quad (6.2.2)$$

где k_i — любые целые числа (положительные или отрицательные).

Доказательство. Прежде всего, покажем, что для каждого числа r вида (6.2.2) расстояние $d(\infty, r)$ не превосходит n .

Для $n = 1$ это следует из упражнения 49. Предположим теперь, что утверждение доказано для всех непрерывных дробей длины $\leq n - 1$, и рассмотрим дробь длины n , заданную формулой (6.2.2). Обозначим через r' число $\frac{1}{r - k_1}$. Ясно, что r' записывается непрерывной дробью длины $n - 1$, поэтому $d(\infty, r') \leq n - 1$. Теперь, ввиду инвариантности расстояния относительно сдвигов $r \mapsto r + k$, $k \in \mathbb{Z}$, и относительно инверсии $r \mapsto r^{-1}$, мы заключаем

$$d(\infty, r) = d(\infty, r - k_1) = d(0, r') \leq d(0, \infty) + d(\infty, r') \leq 1 + (n - 1) = n.$$

Первое неравенство — это просто неравенство треугольника, а второе следует из упражнения 49 а) и из предположения индукции.

Теперь проверим обратное утверждение: все точки шара $B_n(\infty)$ записываются в виде (6.2.2). Случай $n = 1$ опять сводится к упражнению 49, а общий случай также рассматривается с помощью индукции. Ключевое соображение: для каждой точки $r \in B_n(\infty)$ существует точка $r' \in B_{n-1}(\infty)$, близкая к r . \square

Строение сфер — это более сложный вопрос. «Сложность» сферы возрастает с ростом ее радиуса.

Например, $S_1(\infty) = \mathbb{Z}$. Это — однородное пространство относительно группы $\text{Aff}(1, \mathbb{Z})$, которая играет роль «группы вращений» вокруг бесконечной точки — см. упражнение 49 а).

Сфера $S_2(\infty)$ состоит из точек $k_1 + \frac{1}{k_2}$, где $k_1, k_2 \in \mathbb{Z}$ и $k_2 \neq 0, \pm 1$. Под действием группы $\text{Aff}(1, \mathbb{Z})$ она распадается на бесконечное число орбит Ω_m , нумеруемых числом $m = |k_2| \geq 2$. Стабилизатор точки $k + \frac{1}{m} \in \Omega_m$ тривиален при $m > 2$ и содержит один нетривиальный элемент $r \mapsto 2k + 1 - r$ для $m = 2$.

Задача 11. Описать орбиты $\text{Aff}(1, \mathbb{Z})$ на сфере $S_k(\infty)$ для $k > 2$.

6.3. РАЦИОНАЛЬНАЯ ПАРАМЕТРИЗАЦИЯ ОКРУЖНОСТИ

Хорошо известно, что окружность как вещественное алгебраическое многообразие рационально эквивалентна вещественной проективной прямой. Это значит, что можно установить взаимно однозначное соответствие между окружностью и проективной прямой с помощью рациональных функций.

Например, окружность $x^2 + y^2 = 1$ отождествляется с проективной прямой с однородными координатами $(t_0 : t_1)$ по формуле

$$x = \frac{t_1^2 - t_0^2}{t_1^2 + t_0^2}, \quad y = \frac{2t_0t_1}{t_1^2 + t_0^2}, \quad t = \frac{t_1}{t_0} = \frac{y}{1-x} = \frac{1+x}{y}. \quad (6.3.1)$$

В частности, когда проективный параметр t пробегает $\overline{\mathbb{Q}}$, соответствующая точка (x, y) пробегает все рациональные точки окружности¹.

Отсюда можно вывести хорошо известное описание всех примитивных целочисленных решений уравнения $x^2 + y^2 = z^2$. Именно, в каждом примитивном решении ровно одно из чисел x, y четно; если это y , то

$$x = a^2 - b^2, \quad y = 2ab, \quad \pm z = a^2 + b^2, \quad (6.3.2)$$

где a, b — взаимно простые числа.

Аналогично, проективизация будущего светового конуса — это просто двумерная сфера, рационально эквивалентная расширенной плоскости $\overline{\mathbb{R}^2}$. Поэтому все будущие световые векторы с целыми коэффициентами с точностью до пропорциональности даются формулой

$$t = k^2 + l^2 + m^2, \quad x = 2km, \quad y = 2lm, \quad z = \pm(k^2 + l^2 - m^2). \quad (6.3.3)$$

¹ То есть точки с рациональными координатами (x, y) .

Я не знаю, верно ли, что любой целочисленный световой вектор может быть записан в виде (6.3.3) для некоторых целых взаимно простых k, l, m .

Теперь мы учтем, что на вещественной проективной прямой есть естественная ориентация. Для наших целей удобно будет рассматривать ориентацию как циклический порядок для каждой тройки различных точек $x_1, x_2, x_3 \in \overline{\mathbb{R}}$. Геометрически, это понятие ориентации связано с обычным (то есть выбором положительного направления обхода) следующим образом: двигаясь от x_1 в положительном направлении, мы встречаем x_2 раньше, чем x_3 . Мы также будем использовать выражение « x_2 лежит между x_1 и x_3 ».

Предупреждение. Наше понятие «между» отличается от стандартного, которое существует на прямой, но не существует на окружности. Кроме того, наше понятие не симметрично: если x_2 лежит между x_1 и x_3 , то x_2 *не лежит* между x_3 и x_1 ! Зато наше понятие «между» инвариантно относительно циклических перестановок трех точек.

Упражнение 51. а) Покажите, что в случае, когда все три точки x_1, x_2, x_3 конечны, утверждение « x_2 лежит между x_1 и x_3 » эквивалентно неравенству

$$(x_1 - x_2)(x_2 - x_3)(x_3 - x_1) > 0.$$

б) Какие из следующих утверждений верны:

- (i) 1 лежит между 0 и ∞ ;
- (ii) ∞ лежит между 0 и 1;
- (iii) -1 лежит между 0 и ∞ ?

Теперь мы определим новую операцию¹ «вставки» на $\overline{\mathbb{Q}}$. Эта операция сопоставляет паре рациональных чисел (r_1, r_2) третье рациональное число, обозначаемое $r_1 \downarrow r_2$, так, что выполняются условия: существуют такие целые p_1, p_2, q_1, q_2 , удовлетворяющие условиям

$$\text{НОД}(p_1, q_1) = \text{НОД}(p_2, q_2) = 1,$$

что

$$r_1 = \frac{p_1}{q_1}, \quad r_2 = \frac{p_2}{q_2}, \quad r_1 \downarrow r_2 = \frac{p_1 + p_2}{q_1 + q_2}; \quad (6.3.4)$$

$$\text{точка } r_1 \downarrow r_2 \text{ лежит между } r_1 \text{ и } r_2. \quad (6.3.5)$$

¹ Как я узнал от Р. Борчердса, в Англии эту операцию называют «сложением дробей для филологов». Она также служит предметом одного из анекдотов, часто цитировавшихся на семинаре Гельфанда.

УПРАЖНЕНИЕ 52. Покажите, что условие (6.3.4) определяет две разных точки в зависимости от четырех разных выборов знаков p_i и q_i . Проверьте, что ровно одна из этих точек удовлетворяет (6.3.5).

УПРАЖНЕНИЕ 53. Вычислите следующие выражения:

а) $0 \downarrow \infty$; б) $\infty \downarrow 0$; в) $\infty \downarrow -2$; г) $1 \downarrow 2$; д) $2 \downarrow 1$; е) $\frac{1}{2} \downarrow -\frac{1}{3}$.

Ответ. а) 1; б) -1 ; в) -3 ; г) $\frac{3}{2}$; д) ∞ ; е) -2 .

Операция \downarrow имеет особенно интересные свойства, когда r_1 и r_2 близки. В этом случае $r_1 \downarrow r_2$ близко к обоим числам r_1 и r_2 .

УПРАЖНЕНИЕ 54. Покажите, что для близких чисел r_1, r_2 число $r_1 \downarrow r_2$ — это единственное рациональное число между (в нашем смысле) r_1 и r_2 , которое близко к обоим числам.

Эти рассуждения приводят к понятию так называемых *рядов Фарея*. Стандартный ряд Фарея F^n ранга n по определению состоит из всех рациональных чисел $0 < \frac{p}{q} < 1$ с $1 \leq q \leq n$, написанных в возрастающем порядке. Число членов в F^n равно $\sum_{k=2}^n \varphi(k)$, где $\varphi(k)$ — функция Эйлера, дающая количество натуральных чисел, меньших k и взаимно простых с k . Она дается формулой

$$\varphi(n) = n \cdot \prod_{p|n} (1 - p^{-1}),$$

где p пробегает все простые делители n . Например, ряд Фарея F^5 содержит $\varphi(2) + \varphi(3) + \varphi(4) + \varphi(5) = 1 + 2 + 2 + 4 = 9$ членов и выглядит так:

$$\frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}.$$

Мы отсылаем читателя к [Nev] по поводу многих известных свойств рядов Фарея и перечислим здесь только те, которые нам нужны.

УПРАЖНЕНИЕ 55. Покажите, что соседние члены ряда Фарея — близкие числа.

Для наших целей удобно несколько изменить понятие ряда Фарея. А именно, назовем *модифицированным рядом Фарея* подмножество $F^{(n)} \subset \mathbb{R}$, определяемое индуктивно следующим образом.

Ряд $F^{(0)}$ состоит из трех чисел: 0, 1 и ∞ с данным циклическим порядком. Ряд $F^{(n+1)}$, $n \geq 0$, получается из $F^{(n)}$ с помощью вставки между каждой парой последовательных чисел a, b числа $a \downarrow b$. Таким образом, общее число членов удваивается и модифицированный ряд Фарея $F^{(n)}$ содержит $3 \cdot 2^n$ циклически упорядоченных членов. Мы обозначим через $f_k^{(n)}$, $1 - 2^n \leq k \leq 2^{n+1}$, k -й член ряда $F^{(n)}$. Таким образом, для любого n мы имеем равенства $f_0^{(n)} = 0$, $f_2^{(n)} = 1$, $f_{2^{n+1}}^{(n)} = \infty$.

Модифицированные ряды Фарея ранга ≤ 3 показаны ниже:

$k :$																	0																	1																	2
$f_k^{(0)} :$																	$\frac{0}{1}$																	$\frac{1}{1}$																	$\frac{1}{0}$
$k :$					-1					0					1					2					3					4																					
$f_k^{(1)} :$					$-\frac{1}{1}$					$\frac{0}{1}$					$\frac{1}{2}$					$\frac{1}{1}$					$\frac{2}{1}$					$\frac{1}{0}$																					
$k :$	-3	-2			-1			0	1	2	3	4	5	6	7	8																																			
$f_k^{(2)} :$	$-\frac{2}{1}$	$-\frac{1}{1}$			$-\frac{1}{2}$			$\frac{0}{1}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{1}$	$\frac{3}{2}$	$\frac{2}{1}$	$\frac{3}{1}$	$\frac{4}{1}$																																			
$k :$	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16																											
$f_k^{(3)} :$	$-\frac{3}{1}$	$-\frac{2}{1}$	$-\frac{2}{2}$	$-\frac{1}{1}$	$-\frac{2}{3}$	$-\frac{1}{2}$	$-\frac{1}{3}$	$\frac{0}{1}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{2}{5}$	$\frac{1}{2}$	$\frac{3}{3}$	$\frac{2}{4}$	$\frac{3}{1}$	$\frac{4}{3}$	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{2}{1}$	$\frac{5}{2}$	$\frac{3}{1}$	$\frac{4}{1}$	$\frac{1}{1}$	$\frac{1}{0}$																											

Найти явную формулу для чисел $f_k^{(n)}$ — нетривиальная задача. Мы обсудим ее ниже.

УПРАЖНЕНИЕ 56. Покажите, что $f_k^{(n)} = f_{2k}^{(n+1)}$, так что $f_k^{(n)}$ фактически зависит не от k и n в отдельности, а только от двоично-рационального числа $r = \frac{k}{2^n}$. Поэтому мы часто будем писать f_r вместо $f_k^{(n)}$.

Чтобы упростить изложение, мы временно рассмотрим только часть $F^{(n)}$, лежащую между 0 и 1, то есть члены f_r с $0 \leq r \leq 1$.

Заметим, что если бы мы изменили процедуру построения $F^{(n)}$ и вставляли между двумя числами a, b не $a \downarrow b$, а их среднее арифметическое $\frac{a+b}{2}$, мы получили бы на n -м шагу арифметическую прогрессию с $2^n + 1$ членами, которая начинается с 0 и кончается 1. Вместо $f_k^{(n)}$ мы получили бы $a_k^{(n)} = \frac{k}{2^n}$, или, в обозначениях, введенных выше, $a_r = r$.

Теперь мы вполне готовы к определению замечательной функции, впервые введенной Германом Минковским, который назвал ее функцией ? (функцией «вопросительный знак») — см. схолию E в первой части.

ТЕОРЕМА (Минковского). Существует единственная непрерывная и монотонно возрастающая функция $?: [0, 1] \rightarrow [0, 1]$ такая, что

$$?(a \downarrow b) = \frac{?(a) + ?(b)}{2} \quad \text{для всех близких чисел } a, b \in [0, 1]. \quad (6.3.6)$$

СХЕМА ДОКАЗАТЕЛЬСТВА. Формула (6.3.6) показывает, что если искомая функция существует, она должна иметь свойство $?(f_k^{(n)}) = a_k^{(n)}$. Другими словами, $?(f_r) = r$ для всех $r \in \mathbb{Z} \left[\frac{1}{2} \right] \cap [0, 1]$.

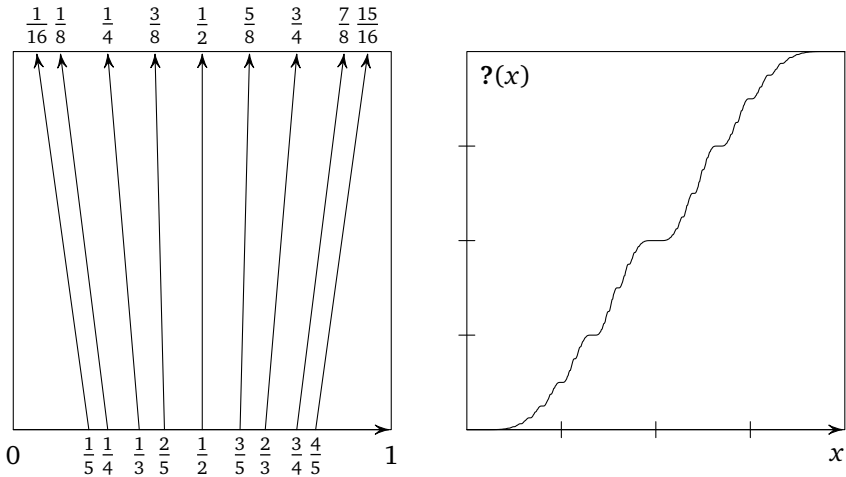


Рис. 6.2. График функции ?

С другой стороны, мы можем определить функцию ? на $\mathbb{Z}\left[\frac{1}{2}\right] \cap [0, 1]$ по формуле $?(f_r) = r$. Поскольку каждое из множеств $\{f_k^{(n)}\}$ и $\{a_k^{(n)}\}$ плотно в $[0, 1]$, наша функция продолжается единственным образом на $[0, 1]$ с сохранением монотонности. А именно, мы полагаем

$$?(x) = \lim_{n \rightarrow \infty} ?(x_n), \tag{6.3.7}$$

где $\{x_n\}$ — любая монотонная последовательность рациональных чисел, сходящаяся к x . \square

Функция p , обратная к $?$, решает поставленную выше задачу о вычислении $f_k^{(n)}$. А именно, $f_r = p(r)$.

На множестве $\mathbb{Z}\left[\frac{1}{2}\right] \cap [0, 1]$ двоично-рациональных чисел функция $p(x)$ может быть вычислена шаг за шагом, используя свойство

$$p\left(\frac{2k+1}{2^{n+1}}\right) = p\left(\frac{k}{2^n}\right) \downarrow p\left(\frac{k+1}{2^n}\right), \tag{6.3.8}$$

которое следует из (6.3.6). Этот способ вычисления по существу повторяет конструкцию модифицированного ряда Фарея.

ТЕОРЕМА 6.3. Функция $p := ?^{-1}$ имеет следующие свойства:

1. а) $p(1-x) = 1-p(x)$; б) $p\left(\frac{x}{2}\right) = \frac{p(x)}{1+p(x)}$; в) $p\left(\frac{1+x}{2}\right) = \frac{1}{2-p(x)}$.
2. $(p')\left(\frac{k}{2^n}\right) = \infty$ для любых n и k , $0 \leq k \leq 2^n$.

3. Для любого рационального, но не двоично-рационального числа $r \in [0, 1]$ значение $p(r)$ является квадратичной иррациональностью, то есть имеет вид $r_1 + \sqrt{r_2}$ для некоторых рациональных r_1, r_2 .

4. Имеет место замечательная формула:

$$p(\underbrace{0,0\dots 0}_{k_1} \underbrace{1\dots 1}_{l_1} \dots \underbrace{0\dots 0}_{k_n} \underbrace{1\dots 1}_{l_n} \dots) = \frac{1}{k_1 + \frac{1}{l_1 + \frac{1}{\ddots + \frac{1}{k_n + \frac{1}{l_n + \ddots}}}}}, \quad (6.3.9)$$

где в левой части используются бесконечные двоичные дроби, а в правой части — бесконечные непрерывные дроби. Формула (6.3.9) остается верной и для конечных двоичных дробей (угадайте правильную форму правую часть в этом случае).

СХЕМА ДОКАЗАТЕЛЬСТВА. Соотношения 1 а)–в) могут быть выведены с помощью следующего полезного факта.

ЛЕММА 6.4. Пусть $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL(2, \mathbb{Z})$. Тогда преобразование множества $\overline{\mathbb{Q}}$ по формуле

$$r \mapsto g \cdot r := \frac{ar+b}{cr+d}$$

перестановочно с операцией вставки, то есть

$$(g \cdot r_1) \downarrow (g \cdot r_2) = g \cdot (r_1 \downarrow r_2). \quad (6.3.10)$$

Мы оставляем проверку этого факта читателю и сделаем только два полезных замечания, каждое из которых можно положить в основу доказательства.

1. Преобразования, о которых идет речь, переводят близкие числа в близкие.

2. Группа $GL(2, \mathbb{Z})$ порождается двумя элементами:

$$g_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad g_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Теперь мы докажем 1 а). Рассмотрим следующую диаграмму

$$\begin{array}{ccc} [0, 1] & \xrightarrow{x \mapsto 1-x} & [0, 1] \\ p \downarrow & & \downarrow p \\ [0, 1] & \xrightarrow{x \mapsto 1-x} & [0, 1]. \end{array} \quad (6.3.11)$$

Соотношение 1 а) равносильно коммутативности этой диаграммы. Чтобы это проверить, выберем в качестве $x \in [0, 1]$ двоично-рациональное число $r = \frac{k}{2^n} = a_r$.

Левая вертикальная стрелка переводит это число в $p(a_r) = f_r$, а нижняя горизонтальная стрелка посылает f_r в $1 - f_r$.

С другой стороны, верхняя горизонтальная стрелка переводит r в $1 - r = a_{1-r}$, а правая вертикальная стрелка посылает a_{1-r} в f_{1-r} . Таким образом, соотношение 1 а) верно для любого двоично-рационального числа. По непрерывности оно верно всюду.

Рассмотрим теперь соотношение 1 б). Оно эквивалентно коммутативности диаграммы

$$\begin{array}{ccc} [0, 1] & \xrightarrow{x \mapsto x/2} & \left[0, \frac{1}{2}\right] \\ p \downarrow & & \downarrow p \\ [0, 1] & \xrightarrow{x \mapsto \frac{x}{1+x}} & \left[0, \frac{1}{2}\right]. \end{array} \quad (6.3.12)$$

Начнем с точки $r = a_r \in [0, 1]$. Верхняя горизонтальная стрелка переводит эту точку в $a_{r/2}$, а затем правая вертикальная стрелка посылает ее в $f_{r/2}$.

С другой стороны, левая вертикальная стрелка переводит a_r в f_r , а затем нижняя горизонтальная стрелка посылает результат в $\frac{f_r}{1+f_r}$.

Таким образом, мы должны проверить равенство $\frac{f_r}{1+f_r} = f_{r/2}$. Для этого мы заметим, что преобразование $x \mapsto \frac{x}{1+x}$ переводит отрезок $[0, 1]$ в отрезок $\left[0, \frac{1}{2}\right]$. Поскольку это преобразование принадлежит группе $PGL(2, \mathbb{Z})$, оно переводит ряд Фарея в себя, причем f_0 и f_1 переходят в f_0 и $f_{1/2}$ соответственно. Индукцией по n мы выводим, что $f_{k/2^n}$ переходит в $f_{k/2^{n+1}}$.

Соотношение 1 в) доказывается так же, используя диаграмму

$$\begin{array}{ccc} [0, 1] & \xrightarrow{x \mapsto \frac{1+x}{2}} & \left[\frac{1}{2}, 1\right] \\ p \downarrow & & \downarrow p \\ [0, 1] & \xrightarrow{x \mapsto \frac{1}{2-x}} & \left[\frac{1}{2}, 1\right]. \end{array} \quad (6.3.13)$$

Общая часть всех этих доказательств состоит в использовании двух фактов: аффинные преобразования сохраняют полусуммы, а

преобразования из $PGL(2, \mathbb{Z})$ сохраняют вставки. Я рекомендую читателю сформулировать и доказать другие свойства функций $?$ и p с помощью подходящих диаграмм.

Полезно также расширить область определения $?$ и p на все множество $\overline{\mathbb{R}}$, используя тождества:

$$p\left(\frac{1}{x}\right) = \frac{1}{p(x)}; \quad p(-x) = -p(x). \quad (6.3.14)$$

Свойство 2 достаточно проверить в точке $x = 0$. Общий случай, когда $x = \frac{k}{2^n}$, рассматривается аналогично, а также формально сводится к случаю $x = 0$ с помощью 1 а)–в).

Итак, пусть $x = 0$. Мы имеем $p(0) = 0$, $p\left(\frac{1}{2^n}\right) = \frac{1}{n+1}$. Таким образом, если $\frac{1}{2^n} \leq \Delta x \leq \frac{1}{2^{n-1}}$, то $\frac{1}{n+1} \leq \Delta p \leq \frac{1}{n}$. Поэтому $\frac{2^{n-1}}{n+1} \leq \frac{\Delta p}{\Delta x} \leq \frac{2^n}{n}$ для $\frac{1}{2^n} \leq \Delta x \leq \frac{1}{2^{n-1}}$ и $p'(0) = +\infty$.

Утверждение 3 следует из формулы (6.3.9). Что касается самой этой формулы, она проверяется по индукции для конечных дробей, используя свойства рядов Фарея. Общий случай следует по непрерывности (или монотонности). Отметим еще, что в конце части I мы использовали формулу (6.3.9) для определения функции $?$. \square

ЗАМЕЧАНИЕ. Рассмотрим теперь функцию $p := ?^{-1}$ как функцию распределения вероятностной меры μ на $[0, 1]$: мера отрезка $[a, b]$ равна $p(b) - p(a)$. Эта мера является слабым пределом последовательности дискретных мер μ_n , $n \geq 1$, сосредоточенных на подмножестве $F^{(n)} \cap [0, 1]$ так, что точка $f_k^{(n)}$ имеет массу $\frac{1}{2^n}$ для $1 \leq k \leq 2^n$. Ясно, что носителем меры μ является весь отрезок $[0, 1]$ (то есть мера любого интервала $(a, b) \subset [0, 1]$ положительна). В то время как для обычного ряда Фарея мера, определенная аналогичным образом, равномерна, в нашем случае это не так. Подробное изучение этой меры и соответствующей случайной величины было бы очень интересно (см., например, [dR1]). \heartsuit

УПРАЖНЕНИЕ 57. Найдите значения $?(x)$ и $?'(x)$ в точке $x = \frac{1}{3}$.

Подсказка. Используя равенство

$$\frac{1}{2} - \frac{1}{4} + \frac{1}{8} - \frac{1}{16} + \frac{1}{32} - \frac{1}{64} + \dots = \frac{1}{3},$$

покажите, что

$$?\left(\frac{1}{3} - \frac{1}{3 \cdot 4^n}\right) = \frac{\Phi_{2n-1}}{\Phi_{2n+1}}, \quad ?\left(\frac{1}{3} + \frac{2}{3 \cdot 4^n}\right) = \frac{\Phi_{2n}}{\Phi_{2n+2}},$$

где Φ_n — n -е число Фибоначчи, задаваемое формулой

$$\Phi_n = \frac{\varphi^n - (-\varphi)^{-n}}{\varphi + \varphi^{-1}},$$

где $\varphi = \frac{\sqrt{5}+1}{2} \approx 1,618\dots$ — «золотое сечение» (см. схолию G).

ОТВЕТ. $?'\left(\frac{1}{3}\right) = \frac{3-\sqrt{5}}{2}$; $?''\left(\frac{1}{3}\right) = 0$.

Задача 12. Верно ли, что $?'(x) = 0$ для всех рациональных чисел, кроме $a_k^{(n)}$?

Мы можем резюмировать полученный результат следующим образом: существует монотонная параметризация всех рациональных чисел отрезка $[0, 1]$ с помощью более простого множества двоично-рациональных чисел на том же отрезке. Если отбросить ограничение $r \in [0, 1]$, мы получаем параметризацию рациональной проективной прямой \mathbb{Q} с помощью двоично-рациональной проективной прямой $\mathbb{Z}\left[\frac{1}{2}\right]$, которая сохраняет циклический порядок точек.

Замечание 3. Существует интересная геометрическая интерпретация рядов Фарея и функции Минковского. Она была обнаружена Жоржем де Рамом [dR1].

Рассмотрим квадрат $[-1, 1] \times [-1, 1] \subset \mathbb{R}^2$. Разделим каждую сторону квадрата на три равные части. Затем соединим соседние точки деления. Мы получим восьмиугольник с равными углами, но неравными сторонами. Продолжим эту процедуру; разделим каждую сторону восьмиугольника на три равные части и соединим соседние точки деления. В результате получится выпуклый 16-угольник, который содержится в 8-угольнике. Продолжая таким образом, мы построим семейство вложенных выпуклых многоугольников Π_n , $n \geq 1$, с 2^{n+1} сторонами. Пересечение всех этих многоугольников будет выпуклой областью D , ограниченной некоторой кривой C (см. рис. 6.3). Де Рам обнаружил также, что граничная кривая принадлежит классу C^2 , то есть выглядит локально как график функции, имеющей непрерывную вторую производную. В частности, имеет смысл говорить о касательной и кривизне в каждой точке кривой. Основным результатом де Рама можно сформулировать так.

а) Середины сторон каждого Π_n принадлежат предельной кривой C . Занумеруем те из этих точек, которые принадлежат верхней половине кривой, числами $r_k = \frac{k}{2^n}$, $-2^n \leq k \leq 2^n$.

б) Пусть верхняя половина C является графиком функции $y = f(x)$, $|x| \leq 1$, и пусть x_k означает x -координату точки с номером r_k . Тогда

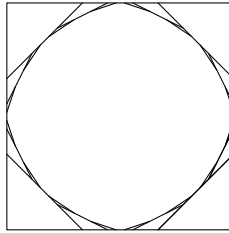


Рис. 6.3. Кривая де Рама

$f'(x_k) = -\text{sign}(r_k) f_{r_k}$, то есть значения производной функции f' на отрезке возрастания являются членами ряда Фарея. ♡

6.4. СОВЕРШЕННЫЕ ПАРАМЕТРИЗАЦИИ КРУГОВ, КАСАЮЩИХСЯ ДАННОГО КРУГА

Пусть \mathcal{A} — ковер Аполлония. Выберем какой-нибудь круг $D \in \mathcal{A}$ и пусть M — соответствующая эрмитова матрица. Рассмотрим все круги в \mathcal{A} , которые касаются D .

Точки касания этих кругов с D образуют счетное множество $T \subset \partial D$. Мы построим такую параметризацию множества T элементами \mathbb{Q} , что циклический порядок на T , индуцированный ориентацией ∂D , совпадает с циклическим порядком на \mathbb{Q} , определенным выше.

Пусть D_r означает круг из \mathcal{A} , касающийся D в точке $t_r \in T$, и пусть M_r — соответствующая эрмитова матрица.

Мы скажем, что параметризация $r \rightarrow t_r$ множества T совершенна, если она имеет следующие свойства:

1. Если $r = \frac{p}{q}$ — несократимая дробь, то

$$M_r = Ap^2 + 2Bpq + Cq^2 - M,$$

где A, B, C — фиксированные эрмитовы матрицы.

2. Круг D_r касается круга $D_{r'}$, если и только если числа $r = \frac{p}{q}$ и $r' = \frac{p'}{q'}$ близки, то есть если $|pq' - p'q| = 1$.

Разумеется, условия 1 и 2 очень сильные и содержат всю информацию о касающихся кругах. Поэтому следующий результат очень важен.

ТЕОРЕМА 6.4. *Совершенные параметризации существуют и обладают дополнительным свойством: пусть v_0, v_1, v_2, v_3 — векторы*

в $\mathbb{R}^{1,3}$, соответствующие матрицам $A + C$, B , $A - C$, M . Тогда матрица Грама их скалярных произведений имеет вид

$$G = \|(v_i, v_j)\| = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad (6.4.1)$$

Первый основной пример: ковер-лента. Пусть $D = \{w \in \mathbb{C} : \text{Im } w \leq 0\}$, $D_\infty = \{w \in \mathbb{C} : \text{Im } w \geq 1\}$. Пусть D_0, D_1 — круги диаметра 1, касающиеся круга D в точках 0, 1, а круга D_∞ — в точках $i, i + 1$ (см. рис. 6.4).

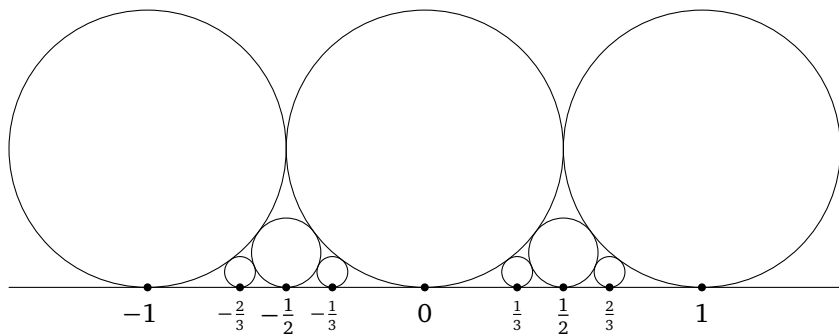


Рис. 6.4. Совершенная параметризация кругов в ковре-ленте

Тогда $\partial D = \overline{\mathbb{R}}$, $T = \overline{\mathbb{Q}}$. Тавтологическая параметризация T совершенна и справедливы равенства:

$$M = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad M_{p/q} = \begin{pmatrix} 2p^2 & -2pq - i \\ -2pq + i & 2q^2 \end{pmatrix}, \quad D_{p/q} : \left| w - \frac{2pq + i}{2q^2} \right| \leq \frac{1}{2q^2}.$$

Второй основной пример: прямоугольный ковер. Пусть $D = \{w \in \mathbb{C} : |w| \geq 1\}$ — дополнение к открытому единичному кругу, D_0 задан условием $\left| w - \frac{1}{2} \right| \leq \frac{1}{2}$, D_∞ — условием $\left| w + \frac{1}{2} \right| \leq \frac{1}{2}$, а D_1 — условием $\left| w - \frac{2i}{3} \right| \leq \frac{1}{3}$.

Здесь ∂D — единичная окружность, а T — рациональная единичная окружность. Совершенная параметризация имеет вид $t_r = \frac{p + iq}{p - iq}$

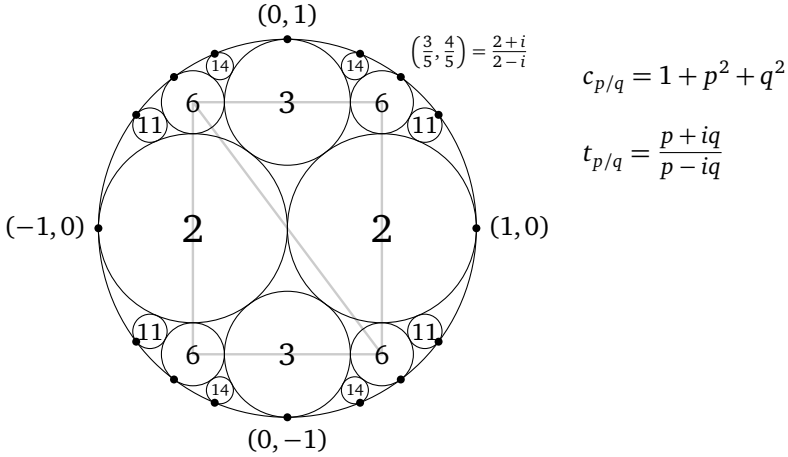


Рис. 6.5. Совершенная параметризация внешней окружности в прямоугольном ковре

для $r = \frac{p}{q}$, так что

$$M = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad M_r = \begin{pmatrix} p^2 + q^2 - 1 & -(p+iq)^2 \\ -(p-iq)^2 & p^2 + q^2 + 1 \end{pmatrix},$$

$$D_{p/q}: \left| w - \frac{(p+iq)^2}{p^2 + q^2 + 1} \right| \leq \frac{1}{p^2 + q^2 + 1}.$$

Доказательство теоремы 6.4. Пусть D_0, D_1, D_∞ — любые три круга из \mathcal{A} , касающиеся данного круга D и друг друга. Занумеруем числами 0, 1 и ∞ точки касания этих кругов с D в циклическом порядке, определенном ориентацией ∂D .

Тогда, в предположении, что совершенная параметризация существует, мы можем вычислить матрицы A, B, C из равенств

$$M_\infty = A - M, \quad M_0 = C - M, \quad M_1 = A + 2B + C - M.$$

Мы получим

$$A = M + M_\infty, \quad C = M + M_0, \quad B = \frac{1}{2}(M_1 - M - M_0 - M_\infty).$$

Теперь, используя свойства матриц M_0, M_1, M_∞ и M , мы можем проверить соотношения (6.4.1). Из них легко следует второе утверждение теоремы, если определить M_r согласно первому утверждению. \square

Практически совершенные параметризации определяются шаг за шагом. А именно, допустим, что для двух близких чисел r_1 и r_2 круги D_{r_1} и D_{r_2} уже построены так, что они касаются D и друг друга. Тогда мы определяем круг D_r , $r = r_1 \downarrow r_2$, как круг, касающийся D_{r_1} , D_{r_2} и D в точке, лежащей между t_{r_1} и t_{r_2} .

Следствие. Кривизна граничной окружности круга, касающегося D в точке с номером $r = \frac{p}{q}$ (несократимая дробь), выражается квадратным многочленом от p, q :

$$c(p, q) = (c_\infty + c) \cdot p^2 + (c_1 - c_0 - c_\infty - c) \cdot pq + (c_0 + c) \cdot q^2 - c, \quad (6.4.2)$$

где c_i (c) — кривизна граничной окружности круга D_i (D).

В частности, если четыре попарно касающиеся круга в ковре Аполлония имеют целочисленные кривизны граничных окружностей, то все круги этого ковра обладают тем же свойством.

УПРАЖНЕНИЕ 58. Для треугольного ковра Аполлония вычислите кривизны всех окружностей, которые касаются внешней окружности.

Ответ. $c(p, q) = \frac{2(p^2 - pq + q^2)}{\sqrt{3}} + 1$.

УПРАЖНЕНИЕ 59. Опишите совершенную параметризацию кругов, касающихся внешней окружности в треугольном ковре Аполлония.

Подсказка. Обозначьте через $0, 1, \infty$ точки касания трех максимальных кругов с внешней окружностью.

6.5. Целочисленные ковры Аполлония

Существует много вариантов ковра Аполлония на $\overline{\mathbb{R}^2}$, для которых кривизны всех окружностей — целые числа. Мы будем называть такие ковры *целочисленными*. Для каждого такого ковра существует (не обязательно единственная) четверка попарно касающихся кругов с кривизнами граничных окружностей ($c_1 \geq c_2 \geq c_3 \geq c_4$), для которой c_1 принимает наименьшее возможное значение. Мы назовем такую четверку *базисной*.

Лемма 6.5. Для базисной четверки справедливы соотношения

$$c_4 \leq 0, \quad |c_4| \leq c_3 \leq \left(1 + \frac{2}{\sqrt{3}}\right) |c_4| \approx 2,1547 \dots \cdot |c_4|.$$

Доказательство. Пусть D_i , $1 \leq i \leq 4$, — базисная четверка попарно касающихся кругов с кривизнами c_i , $1 \leq i \leq 4$, упорядоченными по убыванию.

Рассмотрим уравнение Декарта (4.1.3) как квадратное уравнение относительно c_1 с данными c_2, c_3, c_4 . Корни этого уравнения имеют вид

$$c_1 = c_2 + c_3 + c_4 \pm 2\sqrt{c_2c_3 + c_3c_4 + c_4c_2}. \quad (6.5.1)$$

Покажем, что для базисной четверки мы должны в формуле (6.5.1) выбрать для c_1 знак минус. В самом деле, в противном случае второй корень квадратного уравнения был бы меньше c_1 . Геометрический смысл второго корня — это кривизна другой окружности C , касающейся кругов D_2, D_3, D_4 (отличной от границы D_1). В этом случае мы могли бы заменить круг D_1 на круг с границей C и получить новую четверку с меньшим c_1 .

Неравенство $c_1 \geq c_2$ вместе с (6.5.1) дает

$$c_3 + c_4 \geq 2\sqrt{c_2c_3 + c_3c_4 + c_4c_2},$$

или $(c_3 - c_4)^2 \geq 4c_2(c_3 + c_4) \geq (c_3 + c_4)^2$. Это возможно лишь при $c_4 \leq 0$.

Наконец, для неположительного c_4 мы имеем

$$(c_3 - c_4)^2 \geq 4c_2(c_3 + c_4) \geq 4c_3(c_3 + c_4),$$

или $3c_3^2 + 6c_3c_4 + c_4^2 \leq 4c_4^2$. Это дает $\sqrt{3}(c_3 + c_4) \leq -2c_4$, следовательно,

$$c_3 \leq \frac{2 + \sqrt{3}}{\sqrt{3}}|c_4|. \quad \square$$

Мы приведем здесь список базисных четверок с малыми значениями кривизн, упорядоченный по возрастанию величины $|c_4|$:

$$\begin{aligned} c_4 = 0 & \quad (1, 1, 0, 0); \\ c_4 = -1 & \quad (3, 2, 2, -1); \\ c_4 = -2 & \quad (7, 6, 3, -2); \\ c_4 = -3 & \quad (13, 12, 4, -3), (8, 8, 5, -3); \\ c_4 = -4 & \quad (21, 20, 5, -4), (9, 9, 8, -4); \\ c_4 = -5 & \quad (31, 30, 6, -5), (18, 18, 7, -5); \\ c_4 = -6 & \quad (43, 42, 7, -6), (15, 14, 11, -6), (19, 15, 10, -6); \\ c_4 = -7 & \quad (57, 56, 8, -7), (20, 17, 12, -7), (32, 32, 9, -7); \\ c_4 = -8 & \quad (73, 72, 9, -8), (24, 21, 13, -8), (25, 25, 12, -8); \\ c_4 = -9 & \quad (91, 90, 10, -9), (50, 50, 11, -9), (22, 19, 18, -9), \\ & \quad (27, 26, 14, -9); \\ c_4 = -10 & \quad (111, 110, 11, -10), (62, 60, 12, -10), (39, 35, 14, -10), \\ & \quad (27, 23, 18, -10); \\ c_4 = -11 & \quad (133, 132, 12, -11), (72, 72, 13, -11), (37, 36, 16, -11), \\ & \quad (28, 24, 21, -11). \end{aligned}$$

Полное описание всех базисных четверок неизвестно. Я приведу здесь три формулы, каждая из которых позволяет построить бесконечное число базисных четверок (в том числе большинство из приведенных в списке).

$$\begin{aligned} c_4 &= -km && (k^2 + km + m^2, k(k+m), m(k+m), -km); \\ c_4 &= 1 - 2k && (2k^2, 2k^2, 2k+1, 1-2k); \\ c_4 &= -4k && ((2k+1)^2, (2k+1)^2, 4(k+1), -4k). \end{aligned}$$

Много других интересных фактов о целочисленных коврах и базисных четверках можно найти в работе [GLM⁺].

Схолия I. ФОРМУЛА ОБРАЩЕНИЯ МЁБИУСА

В теоретико-числовых подсчетах часто используется так называемая формула обращения Мёбиуса. Мы объясним здесь, откуда она происходит и как работает.

Пусть задано частично упорядоченное множество X со свойством: для каждого элемента $x \in X$ существует лишь конечное число элементов, меньших x . Пусть f — любая вещественная или комплексная функция на X . Определим новую функцию F на X формулой

$$F(x) = \sum_{y \leq x} f(y). \quad (6.5.1)$$

Вопрос: можно ли восстановить функцию f , зная F ?

Оказывается, что ответ на этот вопрос всегда положительный и дается простой формулой.

ПРЕДЛОЖЕНИЕ I.3. *Существует единственная функция $\tilde{\mu}$ на $X \times X$ со свойствами*

1) $\tilde{\mu}(x, y) = 0$, если $x \not\leq y$;

2) $\tilde{\mu}(x, x) = 1$;

3) если функция F получена из f по формуле (6.5.1), то f восстанавливается по F следующим образом:

$$f(x) = \sum_{y \leq x} \tilde{\mu}(x, y)F(y). \quad (6.5.2)$$

Часто в этой ситуации присутствует дополнительная структура. А именно, множество X является полугруппой положительных элементов A_+ в некоторой абелевой группе A , а отношение порядка инвариантно относительно групповых сдвигов: $x \leq y \Leftrightarrow a + x \leq a + y$ для всех $a \in A$. В этом случае функция $\tilde{\mu}$ также инвариантна относительно сдвигов и выражается через функцию одной переменной: $\tilde{\mu}(x, y) = \mu(x - y)$, где $\mu(x) = \tilde{\mu}(x, 0)$.

Формула обращения принимает вид

$$f(x) = \sum_{y \leq x} \mu(x-y)F(y) \quad (6.5.3)$$

(формула обращения Мёбиуса).

Мы оставляем доказательство заинтересованному читателю, а здесь рассмотрим только некоторые примеры, которые понадобятся дальше.

Пример 1. Пусть $A = \mathbb{Z}$ со стандартным порядком. Тогда формула (6.5.1) принимает вид $F(n) = \sum_{m \leq n} f(m)$, а формула обращения выглядит так: $f(n) = F(n) - F(n-1)$.

Мы видим, что предложение I.3 в этом случае справедливо и функция μ дается формулой

$$\mu(n) = \begin{cases} 1, & \text{если } n = 0, \\ -1, & \text{если } n = 1, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Пример 2. $A = A_1 \times A_2$ и порядок на A является произведением порядков на A_1 и на A_2 , то есть

$$(g_1, g_2) > (0, 0) \Leftrightarrow g_1 > 0 \ \& \ g_2 > 0.$$

Тогда функция μ для A — просто произведение функций μ для A_1 и A_2 . Отметим, что если A_1 и A_2 — упорядоченные группы, то $A = A_1 \times A_2$ является всего лишь частично упорядоченной.

Пример 3. $A = \mathbb{Q}^\times$ — мультипликативная группа ненулевых рациональных чисел. Частичный порядок определяется так: $r_1 \leq r_2$, если число $\frac{r_2}{r_1}$ целое. Таким образом, в этом случае $X = \mathbb{Z}_+$ с отношением порядка $m \leq n \Leftrightarrow m \mid n$ (m является делителем n).

Легко видеть¹, что эта частично упорядоченная группа изоморфна прямой сумме счетного множества копий \mathbb{Z} с обычным порядком. В самом деле, каждый элемент A единственным образом записывается в виде

$$r = \prod_{k \geq 1} p_k^{n_k},$$

где через p_k обозначено k -е простое число, $n_k \in \mathbb{Z}$ и только конечное множество чисел n_k отлично от нуля. Число r целое, если и только

¹ Стандартный оборот речи, который не всегда уместен. Читателю решать, как обстоит дело в этом случае.

если все n_k неотрицательны. Поэтому функция μ является произведением функций из примера 1. Точное определение таково:

$$\mu(n) = \begin{cases} 1, & \text{если } n = 1, \\ (-1)^k, & \text{если } n \text{ — произведение } k \text{ различных простых чисел,} \\ 0 & \text{в остальных случаях.} \end{cases}$$

Соотношение (6.5.3) в этом случае — это классическая формула обращения Мёбиуса:

$$f(n) = \sum_{d|n} \mu(d) F\left(\frac{n}{d}\right). \quad (6.5.4)$$

В качестве приложения мы выведем здесь формулу для φ -функции Эйлера.

Рассмотрим все натуральные числа $k \leq n$ и разобьем их на группы согласно величине $d = \text{НОД}(k, n)$. Ясно, что $\text{НОД}\left(\frac{k}{d}, \frac{n}{d}\right) = 1$. Отсюда следует, что в группе с номером d ровно $\varphi\left(\frac{n}{d}\right)$ чисел. Мы получаем тождество

$$n = \sum_{d|n} \varphi\left(\frac{n}{d}\right).$$

Применяя формулу обращения Мёбиуса, имеем:

$$\varphi(n) = \sum_{d|n} \mu(d) \cdot \frac{n}{d}, \quad \text{или} \quad \frac{\varphi(n)}{n} = \sum_{d|n} \frac{\mu(d)}{d}. \quad (6.5.5)$$

◇

Некоторые вычисления. Хорошо известной нерешенной задачей является вычисление размерности Хаусдорфа для ковра Аполлония, а также меры Хаусдорфа разных его модификаций (например, сферического или треугольного ковра). Нужно пояснить, что имеется в виду под термином «вычисление». В работе [МсМ] показано, что искомая размерность равна $d = 1,305688\dots$, но мы ничего не знаем о природе этого числа. Например, рационально оно или нет? Можно ли его выразить с помощью логарифмов натуральных чисел, как для канторова множества или ковра Серпинского? Имеет ли оно какие-нибудь арифметические свойства?

Другая нерешенная задача — найти сумму площадей всех кругов в ковре Аполлония, которые касаются данного круга D , например, внешнего обобщенного круга в прямоугольном или треугольном ковре.

Здесь, для примера, мы решим более простую задачу. Рассмотрим первый основной пример — ковер-ленту. Мы хотим сосчитать общую площадь всех кругов в ковре-ленте, которые касаются вещественной оси на отрезке $[0, 1]$. Более естественным вопросом, который к тому же имеет более красивый ответ, является вопрос о площади той части единичного квадрата с вершинами $0, 1, 1 + i, i$, которая покрыта кругами, касающимися нижней стороны квадрата.

Мы знаем, что диаметр круга с точкой касания $\frac{m}{n} \in [0, 1]$ равен $\frac{1}{2n^2}$. Поэтому площадь этого круга равна $\frac{\pi}{4n^4}$. Всего имеется $\varphi(n)$ кругов такого размера. Таким образом, интересующая нас площадь дается выражением:

$$A = \frac{\pi}{4} \cdot \sum_{n \geq 1} \frac{\varphi(n)}{n^4}. \quad (6.5.6)$$

Оказывается, можно выразить эту величину через значения ζ -функции Римана в точках 3 и 4.

Для этого мы воспользуемся формулой (6.5.5) для $\varphi(n)$, полученной в схолии I. Выражение (6.5.6) принимает вид

$$A = \frac{\pi}{4} \cdot \sum_{n \geq 1} \sum_{d|n} \mu(d) \frac{1}{n^3 d}.$$

Обозначим $\frac{n}{d}$ через m и произведем суммирование по d и m . Мы получим:

$$A = \frac{\pi}{4} \cdot \sum_{d \geq 1} \sum_{m \geq 1} \frac{\mu(d)}{m^3 d^4} = \frac{\pi}{4} \cdot \sum_{m \geq 1} \frac{1}{m^3} \cdot \sum_{d \geq 1} \frac{\mu(d)}{d^4}.$$

Сумма $\sum_{m \geq 1} \frac{1}{m^3}$ является по определению значением $\zeta(3)$. С другой стороны, сумма $\sum_{d \geq 1} \frac{\mu(d)}{d^4}$ может быть переписана как

$$\sum_{k \geq 0} (-1)^k \cdot \sum_{1 \leq i_1 < i_2 < \dots < i_k} (p_{i_1} p_{i_2} \dots p_{i_k})^{-4} = \prod_{i \geq 1} \left(1 - \frac{1}{p_i^4}\right) = \frac{1}{\sum_{n \geq 1} \frac{1}{n^4}} = \frac{1}{\zeta(4)}.$$

Окончательно мы получаем:

$$A = \frac{\pi}{4} \cdot \frac{\zeta(3)}{\zeta(4)} = \frac{45\zeta(3)}{2\pi^3} \approx 0,76.$$

Общая площадь всех кругов, касающихся внешней окружности в прямоугольном ковре, дается суммой

$$\frac{\pi}{2} \cdot \sum_{\text{НОД}(p,q)=1} \frac{1}{(p^2 + q^2 + 1)^2}.$$

Она может быть выражена через значения ζ -функции, связанной с гауссовским полем $\mathbb{Q}[i]$.

УПРАЖНЕНИЕ 60. Пусть Σ_m означает сумму $\sum_{\mathbb{Z}^2 \setminus \{(0,0)\}} \frac{1}{(k^2 + l^2)^m}$. Покажите, что

$$\sum_{\text{НОД}(p,q)=1} \frac{1}{(p^2 + q^2)^m} = \frac{\Sigma_m}{\zeta(2m)} \quad (6.5.7)$$

и

$$\sum_{\text{НОД}(p,q)=1} \frac{1}{(p^2 + q^2 + 1)^2} = \sum_{m=1}^{\infty} (-1)^{m-1} \frac{m \cdot \Sigma_{m+1}}{\zeta(2m+2)}. \quad (6.5.8)$$

Глава 7

ГЕОМЕТРИЧЕСКИЙ И ТЕОРЕТИКО-ГРУППОВОЙ ПОДХОДЫ

Схолия J. Плоскость Лобачевского (гиперболическая плоскость)

Гиперболическим называют пространство, которое удовлетворяет всем аксиомам евклидова пространства, кроме знаменитого 5-го постулата, утверждающего единственность параллельных линий. В гиперболическом пространстве существует много прямых линий, проходящих через данную точку и не пересекающих данной прямой. Такое пространство существует в каждой размерности, но мы в основном будем использовать двумерное гиперболическое пространство L , называемое также плоскостью Лобачевского. Существуют три наиболее удобных реализации L .

J.1. Первая модель Пуанкаре. Пусть \mathbb{C} — комплексная плоскость с комплексной координатой $z = x + iy$. Обозначим через H открытую верхнюю полуплоскость в \mathbb{C} , заданную условием $\text{Im } z > 0$. Первая модель Пуанкаре отождествляет L как множество с H . Группа \bar{G} конформных преобразований обоого рода (см. схолию F) действует на H и является, по определению, полной группой симметрий L . Согласно философии Клейна (высказанной в его Эрлангенской программе), геометрические свойства L — это те, которые инвариантны относительно группы \bar{G} .

В частности, расстояние $d(z_1, z_2)$ между двумя точками $z_1, z_2 \in H$ должно быть \bar{G} -инвариантным. Оказывается, это свойство, вместе с предположением, что L — риманово многообразие¹, определяет расстояние однозначно, с точностью до выбора единицы длины.

Чтобы получить явную формулу для расстояния, мы можем поступить так. Любой паре различных точек $p = (z_1, z_2)$ из H мы сопоставим четверку $q(p) = (z_1, z_2, \bar{z}_1, \bar{z}_2)$ точек на \mathbb{C} . Соответствие $p \mapsto q(p)$ инвариантно относительно группы $G = PSL(2, \mathbb{R})$, поскольку она дей-

¹ Это, в частности, значит, что в любой локальной системе координат выполняется

$$\text{равенство } d^2(x, 0) = \sum_{i,j=1}^n g_{ij} x_i x_j + o(|x|^2).$$

ствуется дробно-линейными преобразованиями с вещественными коэффициентами.

С другой стороны, хорошо известно, что для любой четверки $q = (z_1, z_2, z_3, z_4)$ точек из \mathbb{C} так называемое *двойное отношение*

$$\lambda(q) := \frac{z_4 - z_3}{z_4 - z_2} : \frac{z_1 - z_3}{z_1 - z_2}$$

не меняется при действии даже большей группы $PSL(2, \mathbb{C})$ всех дробно-линейных преобразований с комплексными коэффициентами.

Это следует из геометрического смысла двойного отношения: $\lambda(q)$ — это та точка, куда z_4 переходит под действием дробно-линейного преобразования, переводящего z_1 в 1, z_2 в ∞ и z_3 в 0.

Введем величину

$$\Delta(p) := \lambda(q(p)) = \frac{\bar{z}_2 - \bar{z}_1}{\bar{z}_2 - z_2} : \frac{z_1 - \bar{z}_1}{z_1 - z_2} = \frac{|z_1 - z_2|^2}{4 \operatorname{Im} z_1 \operatorname{Im} z_2}. \quad (7.0.1)$$

Эта функция от пары точек p в H положительна, симметрична и инвариантна относительно группы \bar{G} . Выясним, как она связана с искомым расстоянием на H . Для этого ограничим наше рассмотрение мнимой полуосью $T = \{ie^\tau, \tau \in \mathbb{R}\}$. Множество T инвариантно относительно однопараметрической группы подобий: $z \mapsto e^t z$ (или $\tau \mapsto \tau + t$) и на нем имеется единственное, с точностью до масштаба, расстояние, инвариантное относительно группы подобий и такое, что T является римановым многообразием относительно этой метрики. А именно,

$$d(ie^{\tau_1}, ie^{\tau_2}) = |\tau_1 - \tau_2|.$$

С другой стороны, мы имеем

$$\Delta(ie^{\tau_1}, ie^{\tau_2}) = \frac{(e^{\tau_1} - e^{\tau_2})^2}{4e^{\tau_1 + \tau_2}} = \frac{1}{4}(e^{\tau_1 - \tau_2} - 2 + e^{\tau_2 - \tau_1}) = \operatorname{sh}^2\left(\frac{\tau_1 - \tau_2}{2}\right).$$

Мы приходим к соотношению

$$\Delta(z_1, z_2) = \operatorname{sh}^2\left(\frac{d(z_1, z_2)}{2}\right), \quad (7.0.2)$$

которое справедливо на $T \times T$, причем обе части равенства G -инвариантны.

УПРАЖНЕНИЕ 61. Покажите, что $G \cdot (T \times T) = H \times H$. Более точно, любая пара точек (z_1, z_2) на H получается движением $g \in G$ из пары точек (i, ie^τ) на T при подходящем $\tau \in \mathbb{R}$.

Из упражнения 61 следует, что (7.0.2) верно повсюду. Простое вычисление приводит к окончательной формуле

$$d(z_1, z_2) = 2 \ln(\sqrt{\Delta(z_1, z_2)} + \sqrt{\Delta(z_1, z_2) + 1}). \quad (7.0.3)$$

Известно, что площадь области $\Omega \subset L$ и длина кривой $C \subset L$ даются интегралами

$$\text{площадь}(\Omega) = \int_{\Omega} \frac{dx \wedge dy}{y^2}, \quad \text{длина}(C) = \int_C \frac{\sqrt{(dx)^2 + (dy)^2}}{y}. \quad (7.0.4)$$

Первое подынтегральное выражение является единственной, с точностью до множителя, дифференциальной 2-формой на L , инвариантной относительно действия G . Эта форма не инвариантна относительно \bar{G} : конформное преобразование второго рода меняет знак формы.

Второе подынтегральное выражение — это квадратный корень из единственной, с точностью до множителя, дифференциальной квадратичной формы (то есть метрики) на L , инвариантной относительно действия G .

УПРАЖНЕНИЕ 62. Покажите, что геодезические, то есть кривые наименьшей длины, — это полуокружности, ортогональные к вещественной оси (включая предельные случаи вертикальных лучей).

Подсказка. Используйте тот факт, что любая пара точек p, q на L определяет единственную геодезическую, проходящую через эти точки. Поэтому любое преобразование $g \in G$, сохраняющее (неупорядоченную) пару p, q , переводит в себя и геодезическую, и середину ее отрезка $[p, q]$. Примените это соображение к паре $p = i, q = ir$ и преобразованию $s: z \mapsto -rz^{-1}$.

Существует замечательное соотношение между площадью треугольника, ограниченного дугами геодезических, и его углами:

$$\text{площадь}(ABC) = \pi - A - B - C. \quad (7.0.5)$$

УПРАЖНЕНИЕ 63. Проверьте формулу (7.0.5) для треугольника с тремя нулевыми углами, который задается неравенствами $-a \leq x \leq a$, $x^2 + y^2 \geq a^2$.

УПРАЖНЕНИЕ 64. Покажите, что круг на плоскости Лобачевского, то есть множество $B_r(a) = \{z \in L: d(z, a) \leq r\}$, — это обычный круг с центром a' и радиусом r' . Выразите a' и r' через a и r .

Ответ. $a' = \operatorname{Re} a + i \operatorname{ch} r \cdot \operatorname{Im} a$, $r' = \operatorname{sh} r \cdot \operatorname{Im} a$.

УПРАЖНЕНИЕ 65. Рассмотрим евклидов круг D на H , заданный неравенством $(x - a)^2 + (y - b)^2 \leq r^2$. Вычислите его диаметр d и его площадь A в смысле гиперболической геометрии¹.

$$\text{ОТВЕТ. } d = \ln \frac{b+r}{b-r}; A = 2\pi \left(\frac{b}{\sqrt{b^2 - r^2}} - 1 \right) = 4\pi \operatorname{sh}^2 \left(\frac{d}{4} \right).$$

J.2. Вторая модель Пуанкаре. Иногда удобнее рассматривать другой вариант модели Пуанкаре. Именно, преобразование Мёбиуса $h: w \mapsto \frac{w-i}{w+i}$ переводит вещественную ось в единичную окружность и верхнюю полуплоскость H во внутреннюю часть D^0 единичного круга $D: x^2 + y^2 \leq 1$. Все, что мы говорили выше об H , может быть повторено для D^0 *mutatis mutandis*.

В частности, группа \bar{G} , действующая на верхней полуплоскости, заменяется группой $\bar{G}' = h \cdot \bar{G} \cdot h^{-1}$, действующей на D^0 . Связная компонента единицы в \bar{G}' — это группа $h \cdot \operatorname{PSL}(2, \mathbb{R}) \cdot h^{-1} = \operatorname{PSU}(1, 1; \mathbb{C})$. Каждой паре точек $p' = (w_1, w_2) \in D^0 \times D^0$ мы ставим в соответствие (G' -инвариантным образом) четверку $q'(p') = (w_1, w_2, \bar{w}_1^{-1}, \bar{w}_2^{-1})$. Затем мы определяем функцию

$$\Delta'(p) := \lambda(q'(p')) = \frac{|w_1 - w_2|^2}{(1 - |w_1|^2)(1 - |w_2|^2)}. \quad (7.0.6)$$

Подгруппа растяжений H , состоящая из преобразований с матрицами $g_\tau = \begin{pmatrix} e^{\tau/2} & 0 \\ 0 & e^{-\tau/2} \end{pmatrix}$, переходит в подгруппу матриц вида

$$g'_\tau = h \cdot g_\tau \cdot h^{-1} = \begin{pmatrix} \operatorname{ch} \tau/2 & \operatorname{sh} \tau/2 \\ \operatorname{sh} \tau/2 & \operatorname{ch} \tau/2 \end{pmatrix}.$$

Эта подгруппа сохраняет интервал $h \cdot T = T' = (-1, 1) \subset D^0$. Если ввести на интервале T' локальный параметр t так, что $x = \operatorname{th} \frac{t}{2}$, то преобразование g'_τ примет простую форму $t \mapsto t + \tau$. Поэтому инвариантное расстояние на T' — это просто $d'(t_1, t_2) = |t_1 - t_2|$.

С другой стороны,

$$\Delta' \left(\operatorname{th} \frac{t_1}{2}, \operatorname{th} \frac{t_2}{2} \right) = \frac{\left(\operatorname{th} \frac{t_1}{2} - \operatorname{th} \frac{t_2}{2} \right)^2}{\left(1 - \operatorname{th}^2 \frac{t_1}{2} \right) \left(1 - \operatorname{th}^2 \frac{t_2}{2} \right)} = \operatorname{sh}^2 \left(\frac{t_1 - t_2}{2} \right).$$

¹ Напомним, что диаметром произвольного множества D называется максимальное расстояние между точками этого множества.

Тогда (7.0.2) и (7.0.3) принимают форму

$$\Delta'(w_1, w_2) = \operatorname{sh}^2\left(\frac{d'(w_1, w_2)}{2}\right), \quad (7.0.7)$$

$$d(w_1, w_2) = 2 \ln(\sqrt{\Delta'(w_1, w_2)} + \sqrt{\Delta'(w_1, w_2) + 1}). \quad (7.0.8)$$

Формулы (7.0.4) заменяются на

$$\text{площадь}(\Omega) = \int_{\Omega} \frac{4 \, dx \wedge dy}{(1-x^2-y^2)^2}, \quad \text{длина}(C) = \int_C \frac{2\sqrt{(dx)^2+(dy)^2}}{1-x^2-y^2}. \quad (7.0.9)$$

Геодезическими являются дуги окружностей, ортогональных к ∂D (включая диаметры D). Формула (7.0.5) остается в силе.

УПРАЖНЕНИЕ 66. Покажите, что множество точек

$$\{z \in L : d(z, a) \leq r\}$$

(гиперболический круг) является обычным кругом с центром a' и радиусом r' . Выразите a' и r' через a и r .

$$\text{ОТВЕТ. } a' = \frac{a}{\sqrt{a^2+1} \operatorname{ch} r}, \quad r' = \frac{(a+a^{-1})}{a \operatorname{th} r + a^{-1} \operatorname{cth} r}.$$

УПРАЖНЕНИЕ 67. Вычислите для круга $D_r(a, b)$ в D_0 , заданного неравенством $(x-a)^2 + (y-b)^2 \leq r^2$, диаметр d и площадь A .

$$\text{ОТВЕТ. } d = \ln \frac{b+r}{b-r}; \quad A = 2\pi \left(\frac{b}{\sqrt{b^2-r^2}} - 1 \right) = 4\pi \operatorname{sh}^2\left(\frac{d}{4}\right).$$

J.3. Модель Клейна. Расширенная группа Мёбиуса \bar{G} изоморфна группе $PO(2, 1, \mathbb{R}) \subset PGL(3, \mathbb{R})$ (см. схолию F). Поэтому существует еще одна модель плоскости Лобачевского L . Она называется моделью Клейна, и мы опишем ее здесь.

Группа $O(2, 1; \mathbb{R})$ по определению действует на вещественном векторном пространстве $\mathbb{R}^{2,1}$ с координатами X, Y, Z , сохраняя конус $X^2 + Y^2 = Z^2$. Рассмотрим вещественную проективную плоскость $P := P^2(\mathbb{R})$ с однородными координатами $(X : Y : Z)$ и локальными координатами $x = \frac{X}{Z}, y = \frac{Y}{Z}$. Соответствующее проективное действие группы $PO(2, 1; \mathbb{R})$ на P сохраняет окружность $x^2 + y^2 = 1$ и открытый круг $D^0: x^2 + y^2 < 1$. Это и есть модель Клейна.

Явная формула группового действия имеет вид

$$x \mapsto \frac{a'x + b'y + c'}{ax + by + c}, \quad y \mapsto \frac{a''x + b''y + c''}{ax + by + c}, \quad (7.0.10)$$

где

$$g = \begin{pmatrix} a' & b' & c' \\ a'' & b'' & c'' \\ a & b & c \end{pmatrix}$$

принадлежит $O(2, 1; \mathbb{R}) \subset GL(3, \mathbb{R})$.

Мы знаем, что $g \in O(2, 1; \mathbb{R})$, если и только если $g^t I g = I$, где $I = \text{diag}(1, 1, -1)$, или, подробнее,

$$\begin{aligned} (a')^2 + (a'')^2 &= a^2 + 1, & a'b' + a''b'' &= ab, \\ (b')^2 + (b'')^2 &= b^2 + 1, & b'c' + b''c'' &= bc, \\ (c')^2 + (c'')^2 &= c^2 - 1, & c'a' + c''a'' &= ca. \end{aligned} \quad (7.0.11)$$

УПРАЖНЕНИЕ 68. а) Покажите, что группа $O(2, 1; \mathbb{R})$ имеет четыре связных компоненты, различаемые знаками $\det g$ и c .

б) Покажите, что $PO(2, 1; \mathbb{R})$ имеет две связных компоненты, различаемые знаком величины $a'b'' - a''b'$: $PSO_+(2, 1; \mathbb{R})$ и $PSO_-(2, 1; \mathbb{R})$.

Отметим, что модель Клейна использует то же множество D^0 и ту же абстрактную группу $\bar{G} \simeq PO(2, 1; \mathbb{R})$, что и вторая модель Пуанкаре, но групповые действия различны.

Однако эти действия сопряжены в большей группе всех диффеоморфизмов D^0 . Более точно, существуют гладкое обратимое отображение $f: D^0 \rightarrow D^0$ и изоморфизм $\alpha: \bar{G} \rightarrow PO(2, 1; \mathbb{R})$ такие, что следующая диаграмма коммутативна:

$$\begin{array}{ccccc} \bar{G} & \times & D^0 & \xrightarrow{\text{конформное действие}} & D^0 \\ \alpha \downarrow & & \downarrow f & & \downarrow f \\ PO(2, 1; \mathbb{R}) & \times & D^0 & \xrightarrow{\text{проективное действие}} & D^0. \end{array}$$

Чтобы описать гомоморфизм α , рассмотрим сначала связную компоненту единицы $G \subset \bar{G}$, которую мы отождествим с группой $PSU(1, 1; \mathbb{C})$. Оказывается, ограничение гомоморфизма α на эту подгруппу индуцируется гомоморфизмом $\tilde{\alpha}: SU(1, 1; \mathbb{C}) \rightarrow SO_+(2, 1; \mathbb{R})$, который имеет вид

$$g = \begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix} \rightarrow \tilde{\alpha}(g) = \begin{pmatrix} \text{Re}(a^2 + b^2) & -\text{Im}(a^2 + b^2) & 2 \text{Re}(ab) \\ \text{Im}(a^2 + b^2) & \text{Re}(a^2 - b^2) & -2 \text{Im}(ab) \\ 2 \text{Re}(\bar{a}b) & -2 \text{Im}(\bar{a}b) & |a|^2 + |b|^2 \end{pmatrix}. \quad (7.0.12)$$

Вторая связная компонента группы \bar{G} является двусторонним классом смежности по подгруппе G , то есть определяется одним элементом $c \in \bar{G}$, в качестве которого можно взять отображение $w \mapsto \bar{w}$.

Но в группе $PO(2, 1; \mathbb{R})$ есть элемент c' , действующий по правилу $x \mapsto x, y \mapsto -y$. Он задается матрицей $\text{diag}(-1, 1, -1) \in SO_-(2, 1; \mathbb{R})$. Из соотношения $c \cdot g \cdot c = \bar{g}$ мы выводим, что $\alpha(c)$ совпадает с c' .

Далее, горизонтальный диаметр D^0 является множеством неподвижных точек для инволюции $c' = \alpha(c)$, следовательно, это геодезическая в модели Клейна. Разумеется, это верно и для всех остальных диаметров и, более общо, для всех образов этого диаметра под действием группы $PO(2, 1; \mathbb{R})$. Поскольку группа действует проективными преобразованиями, она переводит прямые в прямые. Мы приходим к замечательному свойству модели Клейна: все геодезические являются интервалами обычных прямых.

Теперь обратимся к отображению f . Чтобы его найти, мы воспользуемся следующими частными случаями формулы (7.0.12):

$$\tilde{\alpha}: \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \rightarrow \begin{pmatrix} \cos 2\theta & -\sin 2\theta & 0 \\ \sin 2\theta & \cos 2\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

и

$$\tilde{\alpha}: \begin{pmatrix} \text{ch } t & \text{sh } t \\ \text{sh } t & \text{ch } t \end{pmatrix} \rightarrow \begin{pmatrix} \text{ch } 2t & 0 & \text{sh } 2t \\ 0 & 1 & 0 \\ \text{sh } 2t & 0 & \text{ch } 2t \end{pmatrix}.$$

Мы видим, что повороту на угол 2θ в модели Пуанкаре соответствует такой же поворот в модели Клейна.

Напротив, группа сдвигов вдоль диаметра

$$x \mapsto \frac{x \text{ ch } t + \text{sh } t}{x \text{ sh } t + \text{ch } t} \quad (\text{или, если } x = \text{th } \tau, \tau \mapsto \tau + t)$$

переходит в ту же группу, но с удвоенным параметром:

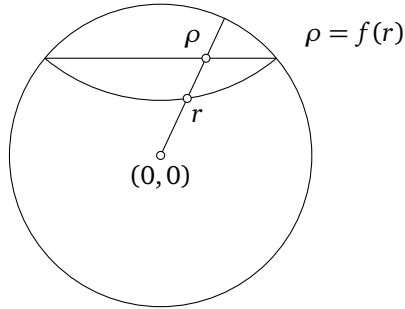
$$x \mapsto \frac{x \text{ ch } 2t + \text{sh } 2t}{x \text{ sh } 2t + \text{ch } 2t} \quad (\text{или } \tau \mapsto \tau + 2t).$$

Мы приходим к заключению, что искомый диффеоморфизм f в полярных координатах (r, α) для модели Пуанкаре и (ρ, θ) для модели Клейна имеет вид

$$f(r, \alpha) = (\rho, \theta), \quad \text{где } \theta = \alpha, \rho = \text{th}(2 \text{th}^{-1}(r)). \quad (7.0.13)$$

УПРАЖНЕНИЕ 69. Покажите, что соотношение (7.0.13) между r и ρ может быть записано также в форме равенств:

$$\text{а) } \frac{1+\rho}{1-\rho} = \left(\frac{1+r}{1-r}\right)^2; \quad \text{б) } \rho = \frac{2}{r+r^{-1}}. \quad (7.0.14)$$

Рис. 7.1. Диффеоморфизм f

Геометрически диффеоморфизм f означает «выпрямление» всех дуг окружностей, перпендикулярных границе, то есть замену их соответствующими хордами (см. рис. 7.1).

Модель Клейна наряду с описанным преимуществом (геодезические — отрезки прямых) имеет и недостатки. Во-первых, она не конформна: углы на модели не равны углам в геометрии Лобачевского. Во-вторых, формулы для расстояния, площади области и длины кривой выглядят сложнее, чем в модели Пуанкаре. Последние в полярных координатах имеют вид:

$$\begin{aligned} \text{площадь}(\Omega) &= \int_{\Omega} \frac{\rho^2 d\rho \wedge d\theta}{(1-\rho^2)\sqrt{1+\rho^2}}, \\ \text{длина}(C) &= \frac{1}{2} \int_C \frac{\sqrt{(d\rho)^2 + \rho^2(1-\rho^2)(d\theta)^2}}{1-\rho^2}. \end{aligned} \quad (7.0.15)$$

УПРАЖНЕНИЕ 70. Покажите, что переход от модели Клейна ко второй модели Пуанкаре может быть получен следующей геометрической конструкцией.

Пусть S_-^2 означает нижнюю (южную) полусферу единичной сферы, не включая экватора. Стереографическая проекция s переводит S_-^2 в открытый единичный круг D^0 , ограниченный экватором. Обозначим через p вертикальную проекцию S_-^2 на D^0 . Тогда отображение $s \circ p^{-1}: D^0 \rightarrow D^0$ задает изоморфизм между моделями Клейна и Пуанкаре.¹ \diamond

¹ Сказочное описание этого изоморфизма содержится в книге С. Боброва «Волшебный двурог» (М.: МЦНМО, 2006).

7.1. ДЕЙСТВИЕ ГРУППЫ \bar{G} И КОВРЫ АПОЛЛОНИЯ

Здесь мы рассмотрим подробнее действие расширенной группы Мёбиуса в связи с коврами Аполлония.

Под действием конформного преобразования (первого или второго рода) данный ковер Аполлония \mathcal{A} переходит в другой такой ковер \mathcal{A}' . Более того, мы знаем, что любой ковер Аполлония получается таким образом из одного фиксированного ковra. Значит, множество \mathbb{A} всех ковров Аполлония является однородным пространством относительно действия группы \bar{G} . Обозначим через $\overline{\text{Aut}}(\mathcal{A})$ подгруппу в \bar{G} , состоящую из преобразований, переводящих \mathcal{A} в себя. Через $\text{Aut}(\mathcal{A})$ мы обозначим пересечение $\overline{\text{Aut}}(\mathcal{A}) \cap G$. Тогда множество \mathbb{A} можно представить как

$$\mathbb{A} = \bar{G} / \overline{\text{Aut}}(\mathcal{A}) = G / \text{Aut}(\mathcal{A}).$$

ТЕОРЕМА 7.1. Подгруппы $\text{Aut}(\mathcal{A}) \subset G$ и $\overline{\text{Aut}}(\mathcal{A}) \subset \bar{G}$ дискретны.

Доказательство. Пусть D_1, D_2, D_3 — три попарно касающихся круга из \mathcal{A} . Выберем три внутренние точки $w_1 \in D_1, w_2 \in D_2, w_3 \in D_3$. Затем выберем окрестность единицы $U \subset G$, достаточно малую для того, чтобы для всех $g \in U$ выполнялось условие $g \cdot w_i \in D_i, i = 1, 2, 3$.

С другой стороны, если $g \in \text{Aut}(\mathcal{A})$, то круги $g \cdot D_1, g \cdot D_2, g \cdot D_3$ тоже принадлежат ковру \mathcal{A} . Но круги D_i и $g \cdot D_i$ имеют общую внутреннюю точку $g \cdot w_i$ и, следовательно, совпадают. Мы видим, что любой элемент $g \in U \cap \text{Aut}(\mathcal{A})$ сохраняет каждый из кругов D_1, D_2, D_3 . Поэтому он имеет три неподвижные точки — точки касания кругов и, значит, равен единице. Это и означает дискретность $\text{Aut}(\mathcal{A})$ в G .

Дискретность $\overline{\text{Aut}}(\mathcal{A})$ следует из того же доказательства, если мы выберем окрестность $U \subset G$ так, чтобы она не содержала конформных преобразований второго рода. \square

Мы хотим теперь исследовать строение групп $\text{Aut}(\mathcal{A})$ и $\overline{\text{Aut}}(\mathcal{A})$, которые являются стабилизаторами точки $\mathcal{A} \in \mathbb{A}$ в группах G и \bar{G} . Основное техническое средство для этого — рассмотрение множеств троек и четверок попарно касающихся кругов на $\bar{\mathbb{C}}$ как однородных пространств относительно группы \bar{G} . При этом полезно рассматривать как упорядоченные, так и неупорядоченные тройки и четверки. Введем соответствующие обозначения:

T (соответственно \tilde{T}) — множество неупорядоченных (соответственно упорядоченных) троек попарно касающихся кругов на $\bar{\mathbb{C}}$;

Q (соответственно \tilde{Q}) — множество неупорядоченных (соответственно упорядоченных) четверок попарно касающихся кругов на $\bar{\mathbb{C}}$.

Выберем начальную точку в \mathbb{A} , например, ковер-ленту, показанный на рис. 5.2, и обозначим его \mathcal{A}_0 . Кроме того, закрепим обозначения D_1, D_2, D_3, D_4 для следующих (обобщенных) кругов из \mathcal{A}_0 :

- D_1 — полуплоскость $\operatorname{Im} w \geq 1$,
- D_2 — полуплоскость $\operatorname{Im} w \leq -1$,
- D_3 — круг $|w - 1| \leq 1$,
- D_4 — круг $|w + 1| \leq 1$.

Назовем эту упорядоченную четверку кругов *базисной четверкой* в \mathcal{A}_0 и обозначим ее \tilde{q}_0 . Ту же четверку, рассматриваемую как неупорядоченную, мы обозначим q_0 .

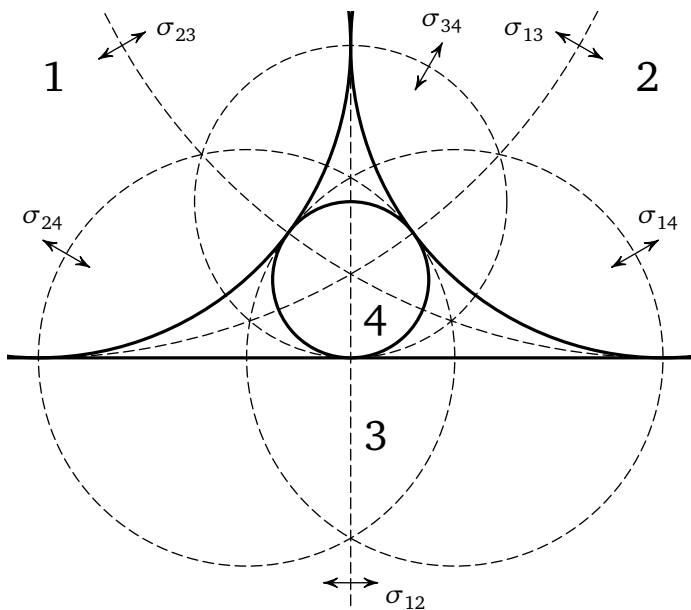
ТЕОРЕМА 7.2. *Группа \bar{G} действует просто транзитивно на множестве \tilde{Q} упорядоченных четверок.*

На множестве Q неупорядоченных четверок стабилизатором точки q_0 является группа, изоморфная S_4 и действующая на четверке q_0 перестановками кругов.

ДОКАЗАТЕЛЬСТВО. Пусть $\tilde{q}' = (D'_1, D'_2, D'_3, D'_4)$ — любая упорядоченная четверка. Существует единственный элемент $g \in G$, переводящий упорядоченную тройку $t_0 = (D_1, D_2, D_3)$ в тройку $t' = (D'_1, D'_2, D'_3)$, поскольку упорядоченная тройка кругов вполне определяется упорядоченной тройкой точек касания.

Круг $g(D_4)$ — это один из двух кругов, касающихся кругов из тройки t' . Второй круг, обладающий тем же свойством, получается из $g(D_4)$ отражением σ в зеркале, ортогональном ко всем кругам из тройки t' . (Аналогичное утверждение для тройки t_0 очевидно. Поэтому оно верно для всех троек.) Таким образом, ровно один из элементов $g, \sigma \circ g$ переводит \tilde{q}_0 в \tilde{q}' . Мы доказали транзитивность действия \bar{G} на \tilde{Q} . Покажем, что стабилизатор точки \tilde{q}_0 сводится к единице. Действительно, любой элемент этого стабилизатора имеет по крайней мере 6 неподвижных точек: $0, \infty, \pm 1 \pm i$ (точки касания четырех кругов). Но любой неединичный элемент G имеет не больше трех неподвижных точек. Что касается преобразований второго рода, заметим, что если такое преобразование имеет три неподвижные точки, то оно сохраняет зеркало (обобщенную окружность), проходящее через эти точки, и является отражением в этом зеркале. Это противоречит тому, что наши 6 точек не лежат на одной окружности.

Остается проверить, что стабилизатор q_0 в \bar{G} действует на этой четверке всеми возможными перестановками. Но мы уже знаем, что для любой перестановки $s \in S_4$ существует $g_s \in \bar{G}$, переводящий четверку \tilde{q}_0 в $(D_{s(1)}, D_{s(2)}, D_{s(3)}, D_{s(4)})$. Группа W , образуемая элементами

Рис. 7.2. Отражения из группы W

g_s , $s \in S_4$, изоморфна S_4 и порождена тремя отражениями σ_{12} , σ_{23} и σ_{34} (см. рис. 7.2). Теорема доказана. \square

Существует четыре четверки q_i , $1 \leq i \leq 4$, имеющие с базисной четверкой q_0 общую тройку $t_i = q_0 \setminus \{D_i\}$. Обозначим через D'_i тот круг в q_i , который не входит в q_0 и через s_i — отражение, переводящее D_i в D'_i и сохраняющее остальные круги в q_0 . См. рис. 7.3.

ТЕОРЕМА 7.3. *Группа Γ , порожденная базисными отражениями σ_i , $1 \leq i \leq 4$, изоморфна группе Γ_4 , введенной в схолии Н.*

СХЕМА ДОКАЗАТЕЛЬСТВА. Благодаря универсальному свойству Γ_4 мы имеем гомоморфизм $\alpha: \Gamma_4 \rightarrow \Gamma$, определенный на образующих правилом: $\alpha(s_i) = \sigma_i$, $1 \leq i \leq 4$. Ясно, что это — эпиморфизм. Посмотрим, каким может быть ядро α .

В схолии Н мы ввели нумерацию элементов группы Γ_4 словами в алфавите $\{1, 2, 3, 4\}$, не содержащими повторяющихся букв. Назовем такие слова *приведенными*.

Напомним также, что $l(w)$ означает длину слова, а $W^{(k)}$ означает множество всех приведенных слов длины k . Таким образом, $W^{(0)}$ содержит только пустое слово \emptyset , соответствующее единице группы;

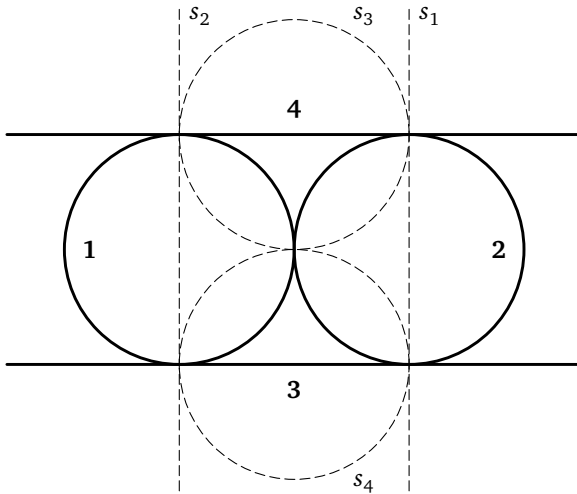


Рис. 7.3. Базисные зеркала группы Γ

множество $W^{(1)}$ содержит четыре слова $\{i\}$, $i = 1, 2, 3, 4$, соответствующих образующим s_i группы; множество $W^{(2)}$ содержит 6 слов $\{ij\}$, $i \neq j$, соответствующих произведениям $s_i s_j$, и т. д.

Для любого элемента $\gamma \in \Gamma_4$ вида $s_{i_1} s_{i_2} \dots s_{i_k}$, $k \geq 1$, мы обозначим через $D(\gamma)$ круг, в который переходит D_{i_1} под действием γ^{-1} , то есть

$$D(\gamma) := \sigma_{i_k} \sigma_{i_k} \cdot \sigma_{i_1} D_{i_1} = \sigma_{i_k} \sigma_{i_k} \cdot \sigma_{i_{k-1}} D'_{i_1}. \tag{7.1.1}$$

Ясно, что имеет место равенство

$$D(\gamma_1 \gamma_2) = \gamma_2^{-1} \cdot D(\gamma_1) \tag{7.1.2}$$

для всех $\gamma_1 \neq \gamma_2 \neq e$. Если неединичный элемент γ принадлежит ядру α , то, согласно (7.1.1), $D(\gamma) = \alpha(\gamma) D_{i_1} = D_{i_1}$.

Однако мы покажем ниже, что все круги $D(\gamma)$, $\gamma \neq e$, различны и лежат вне начальной четверки q_0 . Я рекомендую читателю попытаться найти собственное доказательство, поскольку приводимое ниже довольно длинно.

Разобьем группу Γ_4 на 5 частей: единицу и четыре подмножества $\Gamma_4^{(i)}$, $i = 1, 2, 3, 4$, в зависимости от первой буквы слова, соответствующего данному элементу. Для удобства изложения будем считать, что все неединичные элементы группы Γ_4 покрашены в четыре разные цвета.

Теперь покажем, что можно раскрасить все круги в \mathcal{A}_0 четырьмя цветами так, что в любой четверке попарно касающихся кругов встречаются все четыре цвета. Действительно, сначала раскрасим четырьмя цветами круги, входящие в начальную четверку q_0 . Далее, множество $S^2 \setminus q_0$ состоит из четырех треугольников, каждый из которых ограничен тремя кругами разного цвета. Мы можем продолжить раскраску, присвоив новому кругу, вписанному в треугольник, единственный цвет, отличный от цветов соседей. В новой картинке опять каждая четверка содержит все четыре цвета и мы можем опять продолжать раскраску.

Действие группы Γ_4 сохраняет раскраску, поскольку этим свойством обладают образующие элементы σ_i (проверьте!). Отметим также простое правило: цвет круга $D(\gamma)$ совпадает с первой буквой слова, соответствующего γ , и, стало быть, с цветом самого элемента γ .

Теперь мы построим нумерацию кругов в \mathcal{A}_0 . Именно, рассмотрим все конечные непустые слова в алфавите $\{1, 2, 3, 4\}$, не содержащие повторяющихся букв. Каждому слову $w = \{i_1 i_2 \dots i_k\}$ длины $k \geq 1$ мы поставим в соответствие элемент $\gamma_w = \sigma_{i_1} \sigma_{i_2} \dots \sigma_{i_k}$ длины k в Γ_4 и круг $D_w := D(\gamma_w)$ в \mathcal{A}_0 .

Наша цель — доказать, что все эти круги различны и лежат вне начальной четверки q_0 . Для этого мы сравним построенную здесь нумерацию с той, которая была построена в разделе 5.1. Оказывается, эти нумерации просто совпадают! Я не хочу приводить формальное доказательство, а просто предлагаю читателю прочитать параллельно описание обеих нумераций и убедиться, что они одинаковы. Отсюда, разумеется, вытекает нужное нам утверждение. \square

Продолжим изучение действия группы \bar{G} на кругах.

УПРАЖНЕНИЕ 71. а) Найдите все преобразования $g \in \bar{G}$, сохраняющие неупорядоченную тройку D_1, D_2, D_3 попарно касающихся кругов.

б) Тот же вопрос относительно неупорядоченной четверки D_1, D_2, D_3, D_4 .

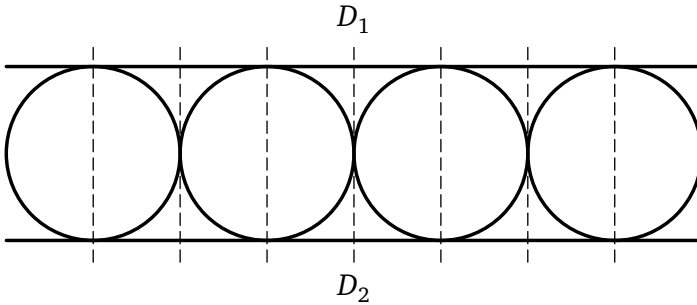
Подсказка. а) Рассмотрите точки касания: $1 \pm i$ и ∞ .

б) Посмотрите, какие из преобразований а) сохраняют круг D_4 .

Из упражнения 71 можно вывести следующий результат.

ТЕОРЕМА 7.4. а) Стабилизатор $S \subset \bar{G}$ любой неупорядоченной тройки попарно касающихся кругов в \mathcal{A} содержится в $\text{Aut}(\mathcal{A})$ и изоморфен группе S_3 : все перестановки кругов возможны.

б) Стабилизатор $\bar{S} \subset \bar{G}$ любой неупорядоченной тройки попарно касающихся кругов в \mathcal{A} содержится в $\text{Aut}(\mathcal{A})$ и изоморфен группе

Рис. 7.4. Стабилизатор пары $\{D_1, D_2\}$

$S_3 \times S_2$; центральный элемент, порождающий S_2 , является отражением в зеркале, ортогональном к $\partial D_1, \partial D_2, \partial D_3$.

в) Стабилизатор в G любой неупорядоченной четверки попарно касающихся кругов в \mathcal{A} содержится в $\text{Aut}(\mathcal{A})$ и изоморфен A_4 : все четные перестановки кругов возможны.

г) Стабилизатор в \bar{G} любой неупорядоченной четверки попарно касающихся кругов в \mathcal{A} содержится в $\overline{\text{Aut}}(\mathcal{A})$ и изоморфен S_4 : все перестановки кругов возможны.

д) Группа $\overline{\text{Aut}}(\mathcal{A})$ действует просто транзитивно на множестве всех упорядоченных четверок в \mathcal{A} . Относительно действия группы $\text{Aut}(\mathcal{A})$ множество всех упорядоченных четверок образует две орбиты.

Для упорядоченной тройки \tilde{t} стабилизатор в G тривиален, поэтому элемент $g \in G$ вполне определяется упорядоченной тройкой $g \cdot \tilde{t}$. По этой же причине элемент $g \in \bar{G}$ вполне определяется упорядоченной четверкой $g \cdot \tilde{q}$.

Теперь мы рассмотрим множество всех пар касающихся кругов в \mathcal{A}_0 . Они образуют однородное пространство относительно группы $\overline{\text{Aut}}(\mathcal{A})$. Стабилизатор пары $\{D_1, D_2\}$ изоморфен группе $\Gamma_2 \simeq \mathbb{Z}_2 * \mathbb{Z}_2 \simeq \text{Aff}(1, \mathbb{Z})$. В самом деле, искомый стабилизатор состоит из преобразований $w \rightarrow \pm w + k$, $k \in \mathbb{Z}$, и свободно порождается двумя отражениями $s_1(w) = -w$, $s_2(w) = 1 - w$ (см. рис. 7.4).

Наконец, рассмотрим стабилизатор в $\text{Aut}(\mathcal{A})$ круга $D_1 \in \mathcal{A}_0$. Удобно заменить ковер \mathcal{A}_0 аффинно эквивалентным ему ковром $\mathcal{A}_1 = \frac{1}{2}(\mathcal{A}_0 + 1 - i)$, так что D_1 становится верхней полуплоскостью, а точки касания D_1 с D_3 и D_4 будут 0 и 1.

ТЕОРЕМА 7.5. Стабилизатор D_1 в G совпадает с $\text{PSL}(2, \mathbb{Z})$. \square

7.2. ДЕЙСТВИЕ ГРУППЫ Γ НА КОВРЕ АПОЛЛОНИЯ

Пусть опять q_0 означает начальную четверку (см. текст перед теоремой 7.2), а s_i , $1 \leq i \leq 4$, означает отражение, сохраняющее три круга из q_0 , исключая D_i . Результат предыдущих рассуждений можно резюмировать в виде теоремы.

ТЕОРЕМА 7.6. а) Группа Γ , порожденная s_1, s_2, s_3, s_4 , изоморфна Γ_4 . Действие этой группы на множестве всех кругов из \mathcal{A} имеет четыре орбиты, каждая из которых содержит ровно один из кругов D_1, D_2, D_3, D_4 начальной четверки.

б) Стабилизатор круга D_1 в Γ порожден отражениями s_2, s_3, s_4 и изоморфен Γ_3 . Действие этого стабилизатора на множестве кругов из \mathcal{A} , касающихся D_1 , имеет три орбиты, каждая из которых содержит ровно один из кругов D_2, D_3, D_4 .

в) Стабилизатор пары D_1, D_2 в Γ порожден отражениями s_3, s_4 и изоморфен Γ_2 . Действие этого стабилизатора на множестве кругов из \mathcal{A} , касающихся D_1, D_2 , имеет две орбиты, каждая из которых содержит ровно один из кругов D_3, D_4 .

г) Стабилизатор тройки D_1, D_2, D_3 в Γ порожден отражением s_4 и изоморфен Γ_1 . Действие этого стабилизатора на множестве кругов из \mathcal{A} , касающихся D_1, D_2, D_3 , имеет одну орбиту, состоящую из кругов D_4, D'_4 .

Иллюстрацией к первому утверждению теоремы служит рис. 7.5, где четыре Γ -орбиты раскрашены разными цветами.

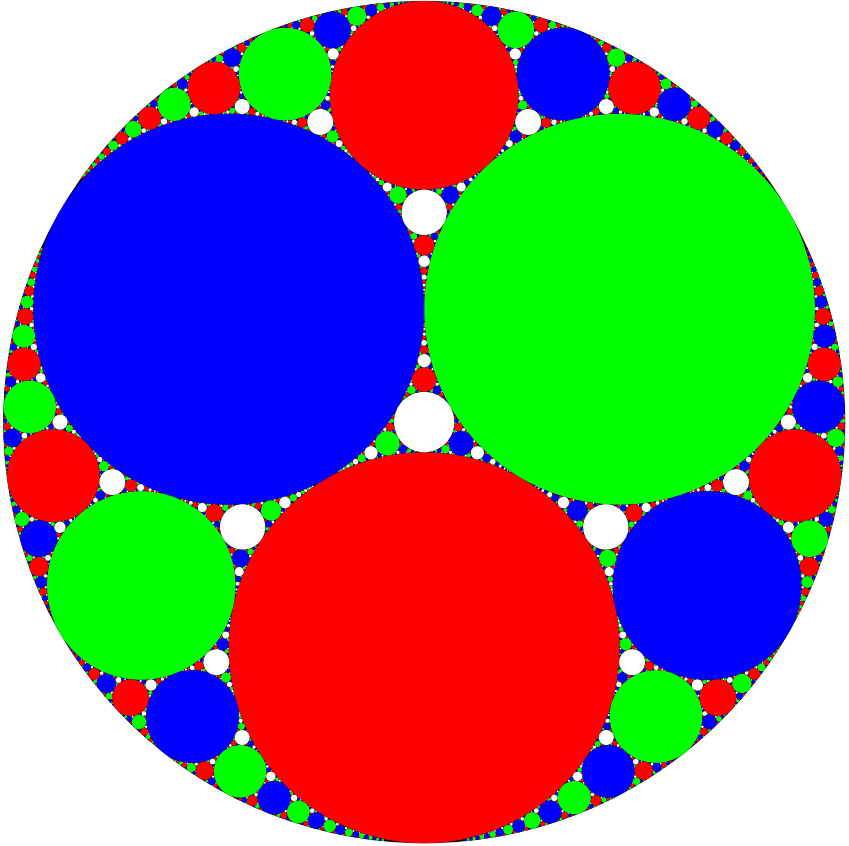
Группа $\Gamma \simeq \Gamma_4$ имеет еще одно действие на множестве всех упорядоченных четверок попарно касающихся кругов из \mathcal{A} . А именно, определим действие образующей s_i , $1 \leq i \leq 4$, на четверке $q = \{D'_1, D'_2, D'_3, D'_4\}$ как отражение q в зеркале, ортогональном окружностям из $\{\partial D'_1, \partial D'_2, \partial D'_3, \partial D'_4\}$, за исключением $\partial D'_i$. Ясно, что это преобразование инволютивно: $s(s(q)) = q$. (Проверьте, что $s_1(q) = \{D''_1, D'_2, D'_3, D'_4\}$, где D''_1 — второй круг (кроме D'_1), касающийся D'_2, D'_3, D'_4 .)

В силу универсального свойства Γ_4 существует единственный гомоморфизм $\beta: \Gamma_4 \rightarrow \overline{\text{Aut}(\mathcal{A})}$, продолжающий действие образующих.

Это действие обладает следующим замечательным свойством.

ЛЕММА 7.1. Допустим, что кривизны граничных окружностей для четверки кругов q равны c_1, c_2, c_3, c_4 . Тогда для четверки $s_1(q)$ соответствующие кривизны равны

$$c'_1, c_2, c_3, c_4, \quad \text{где } c'_1 = 2c_2 + 2c_3 + 2c_4 - c_1. \quad (7.2.1)$$

Рис. 7.5. Орбиты группы Γ

Доказательство. Согласно уравнению Декарта, мы имеем:

$$(c_1 + c_2 + c_3 + c_4)^2 = 2(c_1^2 + c_2^2 + c_3^2 + c_4^2),$$

$$(c'_1 + c_2 + c_3 + c_4)^2 = 2(c_1'^2 + c_2^2 + c_3^2 + c_4^2).$$

Вычитая из первого равенства второе, получаем:

$$(c'_1 - c_1)(c_1 + c'_1 + 2c_2 + 2c_3 + 2c_4) = 2(c_1^2 - c_1'^2),$$

или

$$c_1 + c'_1 + 2c_2 + 2c_3 + 2c_4 = 2(c_1 + c'_1),$$

что и утверждалось. \square

Таким образом, группа $\Gamma \simeq \Gamma_4$ действует линейно в пространстве \mathbb{R}^4 по правилу

$$\begin{aligned} \pi(s_1) &= \begin{pmatrix} -1 & 2 & 2 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & \pi(s_2) &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & -1 & 2 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \\ \pi(s_3) &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 2 & -1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & \pi(s_4) &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 2 & 2 & 2 & -1 \end{pmatrix}. \end{aligned} \tag{7.2.2}$$

Замечательное свойство, о котором говорилось выше, состоит в том, что это действие согласовано с действием $\beta(\Gamma)$ коммутативной диаграммой:

$$\begin{array}{ccc} \tilde{q} & \xrightarrow{\beta(\gamma)} & \tilde{q} \\ c \downarrow & & \downarrow c \\ \mathbb{R}^4 & \xrightarrow{\pi(\gamma)} & \mathbb{R}^4, \end{array}$$

где c означает отображение, ставящее в соответствие упорядоченной четверке $q \in \tilde{q}$ точку в \mathbb{R}^4 с координатами (c_1, c_2, c_3, c_4) , равными кривизнам окружностей, ограничивающих круги из q .

Задача 13. а) Описать образ группы Γ при линейном представлении π .

б) Показать, что π — точное представление (то есть его ядро состоит только из единичного элемента).

Есть еще одна интересная группа, порожденная отражениями и действующая на ковре Аполлония. А именно, пусть h_{ij} — отражение в зеркале, которое проходит через точку касания t_{ij} кругов D_i и D_j и ортогонально двум другим кругам начальной четверки. Ясно, что это отражение переставляет D_i и D_j и сохраняет два остальных круга. Обозначим через H группу, порожденную шестью отражениями h_{ij} . Мы оставляем читателю доказать, что эта группа совпадает с группой W , определенной в конце пункта 7.1.

ТЕОРЕМА 7.7. Полная группа $\text{Aut}(A)$ конформных автоморфизмов (обоого рода) ковра Аполлония \mathcal{A}_0 является полупрямым произведением $H \rtimes \Gamma$ подгруппы H и нормального делителя Γ .

Схема доказательства. По определению, группа H переставляет круги из начальной четверки. Поэтому сопряжение элементами

$h \in H$ переставляет образующие группы Γ . Следовательно, H нормализует подгруппу Γ . Далее, согласно теореме 7.2 группа Γ может перевести любую неупорядоченную четверку q в начальную четверку q_0 . С помощью подходящего элемента H можно произвольно менять порядок кругов в q_0 . Таким образом, группа $H \rtimes \Gamma$ действует транзитивно на множестве упорядоченных четверок из \mathcal{A}_0 . Пусть теперь g — любой элемент группы $\text{Aut}(A)$. Он переводит q_0 в некоторую четверку q . С помощью подходящего элемента $g' \in H \rtimes \Gamma$ мы можем вернуть q в q_0 . Композиция $g' \circ g$ сохраняет q_0 , следовательно, является единицей. Поэтому $g = (g')^{-1}$ принадлежит $H \rtimes \Gamma$. \square

УПРАЖНЕНИЕ 72. Пусть \mathcal{M} означает множество всех зеркал, отражение в которых сохраняет \mathcal{A}_0 . Является ли \mathcal{M} однородным пространством:

- а) для группы Γ ;
- б) для группы $\text{Aut}(\mathcal{A})$;
- в) для группы $\text{Aut}(\mathcal{A})$?

Рассмотрим подробнее линейное представление π группы Γ , определенное формулами (7.2.2).

Лемма 7.2. Преобразования $\pi(s_i)$ являются отражениями и сохраняют квадратичную форму

$$Q(c) = \frac{(c_0 + c_1 + c_2 + c_3)^2}{2} - (c_0^2 + c_1^2 + c_2^2 + c_3^2);$$

следовательно, они переводят каждое решение уравнения Декарта в другое (или то же) решение.

Доказательство. Гиперплоскость M_i , заданная уравнением $c_i = \sum_{j \neq i} c_j$, инвариантна относительно $\pi(s_i)$, поскольку для точек этой гиперплоскости мы имеем $c'_i = 2c_i - c_i = c_i$ (см. лемму 7.1). Значит, $\pi(s_i)$ является отражением в M_i в направлении i -й координатной оси. \square

Напомним, что выше мы использовали замену координат (6.1.1), которая переводит целочисленные решения уравнения Декарта в целочисленные световые векторы пространства Минковского $\mathbb{R}^{1,3}$ с координатами t, x, y, z . Таким образом, представление π эквивалентно некоторому представлению Φ в $\mathbb{R}^{1,3}$ псевдоортогональными преобразованиями.

УПРАЖНЕНИЕ 73. Покажите, что $\Phi(s_i)$ действуют в $\mathbb{R}^{1,3}$ по формулам

$$s_i(v) = v - \frac{2(v, \xi_i)}{(\xi_i, \xi_i)} \xi_i, \quad (7.2.3)$$

где ξ_i , $0 \leq i \leq 3$, — вектор-столбцы матрицы

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

которая приводит форму $Q(c)$ к диагональному виду.

Подсказка. Используйте тот факт, что преобразования (6.1.1) в пространстве \mathbb{R}^4 с координатами (c_0, c_1, c_2, c_3) являются отражениями.

Скажем еще немного о строении группы Γ .

Лемма 7.3. *Группа Γ изоморфна полупрямому произведению $\mathbb{Z}_2 \times F_3$, где F_3 — свободная группа с тремя образующими, а нетривиальный элемент подгруппы \mathbb{Z}_2 действует на F_3 внешним автоморфизмом, обращающим все генераторы.*

Доказательство. В самом деле, мы знаем, что

$$\Gamma = \langle s_0, s_1, s_2, s_3 \mid s_i^2 = 1 \rangle.$$

Введем новые образующие: $s := s_0$ и $\tau_i := s_0 s_i$, $i = 1, 2, 3$. Тогда $s^2 = 1$, $s \tau_i s = \tau_i^{-1}$ и мы должны только проверить, что τ_i не связаны никакими соотношениями. Это можно сделать, используя явную реализацию Γ , построенную выше. \square

В терминах новых образующих представление Φ имеет вид

$$\begin{aligned} \Phi(\tau_1) &= \begin{pmatrix} 5 & -4 & 2 & 2 \\ 4 & -3 & 2 & 2 \\ 2 & -2 & 1 & 0 \\ 2 & -2 & 0 & 1 \end{pmatrix}, & \Phi(\tau_2) &= \begin{pmatrix} 5 & 2 & -4 & 2 \\ 2 & 1 & -2 & 0 \\ 4 & 2 & -3 & 2 \\ 2 & 0 & -2 & 1 \end{pmatrix}, \\ \Phi(\tau_3) &= \begin{pmatrix} 5 & 2 & 2 & -4 \\ 2 & 1 & 0 & -2 \\ 2 & 0 & 1 & -2 \\ 4 & 2 & 2 & -3 \end{pmatrix}, & \Phi(s) &= \begin{pmatrix} 2 & -1 & -1 & -1 \\ 1 & 0 & -1 & -1 \\ 1 & -1 & 0 & -1 \\ 1 & -1 & -1 & 0 \end{pmatrix}. \end{aligned} \tag{7.2.4}$$

Первые три матрицы унитарны с жордановой структурой $(3, 1)$. Интересно было бы найти прямое геометрическое доказательство дискретности этой группы и найти ее фундаментальную область (см., например, [СМ]).

Глава 8

МНОГОМЕРНЫЕ КОВРЫ АПОЛЛОНИЯ

8.1. ОБЩИЕ СООБРАЖЕНИЯ

Рассмотрим следующий аналог проблемы Декарта: найти соотношение между радиусами $n + 2$ попарно касающихся шаров в \mathbb{R}^n .

Здесь опять лучше перейти от радиусов к кривизнам, а также расширить \mathbb{R}^n , добавив бесконечно удаленную точку ∞ . Расширенное пространство $\overline{\mathbb{R}^n}$ топологически эквивалентно единичной сфере S^n , заданной уравнением

$$\sum_{k=0}^n \alpha_k^2 = 1$$

в пространстве R^{n+1} с координатами $\alpha_0, \dots, \alpha_n$.

Пусть \mathcal{B}_n означает совокупность всех обобщенных¹ шаров в $\overline{\mathbb{R}^n}$. Мы дадим здесь несколько параметризаций множества \mathcal{B}_n . Поучительно сравнить этот общий результат со случаем $n = 2$, изучавшимся в предыдущих разделах.

Первая параметризация. Пусть $\mathbb{R}^{1,n+1}$ означает $(n + 2)$ -мерное вещественное векторное пространство с координатами (p^0, \dots, p^{n+1}) , снабженное квадратичной формой

$$|p|^2 := (p^0)^2 - (p^1)^2 - (p^2)^2 - \dots - (p^{n+1})^2. \quad (8.1.1)$$

Каждому вектору $p \in \mathbb{R}^{1,n+1}$ мы поставим в соответствие полупространство $H_p \subset \mathbb{R}^{n+1}$, заданное неравенством

$$H_p := \left\{ \alpha \in \mathbb{R}^{n+1} : p^0 + \sum_{k=1}^{n+1} p^k \alpha_k \leq 0 \right\}. \quad (8.1.2)$$

УПРАЖНЕНИЕ 74. Покажите, что пересечение $S^n \cap H_p$

- для $|p|^2 > 0$ — пусто;
- для $|p|^2 = 0$ — или вся сфера, или единственная точка (какая?);
- для $|p|^2 < 0$ — замкнутый шар, который мы обозначим B_p .

¹ Обобщенный шар — это либо обычный замкнутый шар положительного радиуса, либо замкнутое полупространство, либо внешность непустого открытого шара.

Подсказка. Рассмотрите проекцию \mathbb{R}^{n+1} на прямую, ортогональную к H_p .

Ясно, что для $c > 0$ полупространства H_p и H_{cp} совпадают. Значит, $B_p = B_{cp}$. Поэтому мы можем нормализовать p условием $|p|^2 = -1$. Таким образом, множество \mathcal{B}_n всех шаров в S^n параметризуется точками гиперлоида $|p|^2 = -1$ в $\mathbb{R}^{1,n+1}$.

ВТОРАЯ ПАРАМЕТРИЗАЦИЯ. Определим стереографическую проекцию $s: S^n \rightarrow \overline{\mathbb{R}^n}$, как в схолии F. Это отображение устанавливает биекцию S^n на $\overline{\mathbb{R}^n}$ и переводит шары в S^n в обобщенные шары в \mathbb{R}^n .

Неравенство из (8.1.2) переходит в неравенство

$$a + (b, x) + c(x, x) \leq 0, \quad (8.1.3)$$

где $x = (x_1, \dots, x_n)$, $b = (p^1, \dots, p^n)$, $a = p^0 - p^{n+1}$, $c = p^0 + p^{n+1}$ и выполняется условие $ac - |b|^2 < 0$.

Таким образом, \mathcal{B}_n параметризуется тройками $(a, b, c) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}$. Как и раньше, мы нормализуем вектор (p^0, \dots, p^{n+1}) (или тройку (a, b, c)) условием $|p|^2 = ac - |b|^2 = -1$.

УПРАЖНЕНИЕ 75. Шары B_{p_1} и B_{p_2} касаются друг друга, если и только если $|p_1 + p_2|^2 = 0$.

УПРАЖНЕНИЕ 76. Предположим, что ∂B_{p_1} и ∂B_{p_2} имеют общую точку x . Найдите угол α между радиусами B_{p_1} и B_{p_2} в точке x .

Подсказка. Используйте тот факт, что ответ практически не зависит от размерности n : достаточно рассмотреть двумерную плоскость, проходящую через центры шаров и через точку x . Поэтому общий случай сводится к случаю $n = 2$.

ОТВЕТ.

$$\cos \alpha = -(p_1, p_2). \quad (8.1.4)$$

Пусть теперь B_{p_k} , $k = 1, 2, \dots, n + 2$, — попарно касающиеся шары в $\overline{\mathbb{R}^n}$. Тогда, как и в разделе 5.4, мы имеем

$$(p_i, p_j) = 1 - 2\delta_{ij}.$$

Это следует также из (8.1.4), так как для касающихся шаров $\cos \alpha = \cos \pi = -1$.

Введем матрицу Грама G попарных скалярных произведений: $G_{ij} = (p_i, p_j)$. Число -2 является собственным значением кратности $n + 1$, так как ранг матрицы $G + 2 \cdot 1$ равен 1. Учитывая, что след матрицы Грама равен $-(n + 2)$, мы видим, что $(n + 2)$ -е собственное значение равно n . Поэтому матрица G невырождена и векторы p_k , $1 \leq k \leq n + 2$, линейно независимы. Следовательно, они образуют базис в $\mathbb{R}^{1,n+1}$.

Далее, для любого вектора $v \in \mathbb{R}^{1,n+1}$ мы введем два набора координат: так называемые *ковариантные* координаты $v_k = (v, p_k)$ и *контравариантные* координаты v^k , определяемые равенством $v = \sum v^k p_k$.

Соотношения между двумя типами координат выводятся в точности так же, как для двумерного случая. Они выглядят так:

$$v_j = \sum_i v^i - 2v^j, \quad v^i = \frac{1}{2n} \sum_j v_j - \frac{1}{2} v_i. \quad (8.1.5)$$

Основная квадратичная форма в этих координатах дается формулами

$$|v|^2 = \left(\sum_i v^i \right)^2 - 2 \sum_i (v^i)^2 = \frac{1}{2n} \left(\left(\sum_i v_i \right)^2 - n \sum_i (v_i)^2 \right). \quad (8.1.6)$$

Рассмотрим вектор v с ковариантными координатами $(1, -1, 0, \dots, 0, 0)$; тогда $v_k = (v, p_k) = p_k^{n+1} + p_k^0 = c_k$. Напомним, что c_k — это кривизна границы шара B_{p_k} . Поскольку $|v|^2 = 0$, мы получаем уравнение

$$\left(\sum_k c_k \right)^2 = n \cdot \sum_k c_k^2, \quad (8.1.7)$$

которое является n -мерным аналогом уравнения Декарта.

УПРАЖНЕНИЕ 77*. Докажите следующий n -мерный аналог обобщенного уравнения Декарта:

$$\Sigma_1^2 = n \cdot \Sigma_2 - 2n^2 \cdot 1, \quad (8.1.8)$$

где

$$\Sigma_1 = \sum_{i=0}^{n+1} M_i, \quad \Sigma_2 = \sum_{i=0}^{n+1} M_i^2 \quad (8.1.9)$$

и M_i , $0 \leq i \leq n+1$, — матрицы, соответствующие $n+2$ попарно касающимся шарам в \mathbb{R}^n .

Чтобы понять ситуацию в общем случае, рассмотрим следующую конфигурацию. Пусть $\{B_k^0\}_{1 \leq k \leq n}$ — семейство попарно касающихся шаров в \mathbb{R}^n . Попробуем описать все двусторонние последовательности шаров $\{B_j\}_{j \in \mathbb{Z}}$, обладающие такими свойствами: B_j касается шара $B_{j \pm 1}$ и всех шаров $\{B_k^0\}_{1 \leq k \leq n}$. Пусть d_k означает кривизну границы B_k^0 , а c_j — кривизну границы B_j .

Из (8.1.7) следуют уравнения:

$$(c_j + c_{j \pm 1} + d_1 + \dots + d_n)^2 = n \cdot (c_j^2 + c_{j \pm 1}^2 + d_1^2 + \dots + d_n^2).$$

Вычитая одно уравнение из другого, получаем:

$$2c_j + c_{j+1} + c_{j-1} + 2d_1 + \dots + 2d_n = n(c_{j+1} + c_{j-1}),$$

или

$$(n-1)(c_{j+1} + c_{j-1}) - 2c_j = 2(d_1 + \dots + d_n). \quad (8.1.10)$$

Это неоднородное рекуррентное уравнение для последовательности $\{c_j\}$. Сравнивая два таких уравнения для соседних значений j , можно получить однородное уравнение

$$(n-1)c_{j+1} - (n+1)c_j + (n+1)c_{j-1} - (n-1)c_{j-2} = 0. \quad (8.1.11)$$

Характеристический многочлен этого уравнения равен

$$(n-1)\lambda^3 - (n+1)\lambda^2 + (n+1)\lambda - (n-1) \quad (8.1.12)$$

и его корни имеют вид

$$\lambda_0 = 1, \quad \lambda_{\pm 1} = \frac{1 \pm \sqrt{n(2-n)}}{n-1}. \quad (8.1.13)$$

Мы видим, что структура множества корней и, следовательно, поведение последовательности $\{c_j\}$ различны в случаях $n=2$, $n=3$ и $n>3$.

Когда $n=2$, характеристический многочлен имеет тройной корень $\lambda=1$. Поэтому последовательность $\{c_j\}$ квадратична по j . (Левая часть (8.1.11) — это третья разностная производная от $\{c_j\}$.)

В случае $n=3$ характеристический многочлен имеет корни $\lambda_k = -e^{2k\pi i/3}$, $k=-1, 0, 1$, то есть три корня шестой степени из 1, которые не являются кубическими корнями из 1. Поэтому последовательность $\{c_j\}$ имеет период 6. Более того, не только последовательность кривизн, но и сама последовательность шаров имеет период 6. Есть еще одно обстоятельство: поскольку только три из шести корней из 1 являются корнями характеристического многочлена, последовательность $\{c_j\}$ обладает дополнительным свойством: $c_j + c_{j+3}$ не зависит от j . Эти красивые геометрические факты были известны уже в Древней Греции (см. подробности в [Sod]).

УПРАЖНЕНИЕ 78. Пусть B_1, B_2, B_3 — три попарно касающихся единичных шара в \mathbb{R}^3 . Опишите шестерку шаров, касающихся всех B_k , $k=1, 2, 3$.

Подсказка. Соответствующие кривизны равны 0, 0, 3, 6, 6, 3.

При $n > 3$ ситуация совершенно другая. Характеристический многочлен имеет один вещественный корень $\lambda_0 = 1$ и два комплексных корня $\lambda_{\pm 1} = \frac{1 \pm i\sqrt{n^2 - 2n}}{n - 1}$ на единичной окружности. Запишем их в форме $\lambda_{\pm 1} = e^{\pm i\alpha}$. Тогда $\cos \alpha = \frac{1}{n - 1}$.

Предложение 8.1. Все целочисленные положительные решения m, n уравнения $\cos \frac{2\pi}{m} = \frac{1}{n}$ исчерпываются следующими двумя: $m = n = 1$; $m = 6, n = 2$.

Отсюда следует, что при $n > 3$ последовательность шаров $\{B_j\}_{j \in \mathbb{Z}}$ в \mathbb{R}^3 имеет квазипериодический характер и самопересекается бесконечно много раз. Кроме того, из рекуррентного соотношения

$$c_{j+1} = \frac{2}{n-1}c_j - c_{j-1} \quad (8.1.14)$$

мы заключаем, что при $n > 3$ кривизны границ B_j не могут быть целочисленными для всех j .

8.2. ТРЕХМЕРНЫЙ КОВЕР АПОЛЛОНИЯ

Как мы видели выше, случай $n = 3$ является исключительным. Из каждого решения (c_1, \dots, c_5) уравнения Декарта мы можем изготовить пять новых решений, применяя операции $s_i, i = 1, \dots, 5$. Именно, операция s_i заменяет c_i на $\sum_{j \neq i} c_j - c_i$, оставляя остальные кривизны неизменными. Операции s_i , как и ранее, являются инволюциями: $s_i^2 = \text{Id}$, но, кроме того, они удовлетворяют соотношениям $(s_i s_j)^3 = \text{Id}$ для $i \neq j$. Поэтому каждая пара $(s_i, s_j), i \neq j$, порождает группу, изоморфную S_3 , — группу Вейля алгебры Ли типа A_2 .

Еще более интересно, что любые три отражения (s_i, s_j, s_k) порождают так называемую аффинную группу Вейля для A_2 , изоморфную полупрямому произведению $S_3 \times \mathbb{Z}^2$.

Предложение 8.2. Для любых трех попарно касающихся шаров V_1, V_2, V_3 в $\overline{\mathbb{R}^3}$ множество центров шаров, касающихся всех трех V_i , топологически представляет собой окружность. Можно так параметризовать эту окружность точками $T = \mathbb{R}/2\pi\mathbb{Z}$, что шары $V(\alpha)$ и $V(\beta)$ с центрами в α и β будут касаться тогда и только тогда, когда $\alpha - \beta = \pm \frac{\pi}{3} \pmod{2\pi\mathbb{Z}}$.

Предложение 8.3. Для любых двух касающихся шаров V_1, V_2 в $\overline{\mathbb{R}^3}$ множество центров шаров, касающихся V_1 и V_2 , топологически является двумерной сферой. Можно так параметризовать эту сферу

точками $\overline{\mathbb{R}^2}$, что шары $B(\alpha)$ и $B(\beta)$ с центрами в α и β будут касаться тогда и только тогда, когда $|\alpha - \beta| = 1$.

Мы оставляем читателю доказать эти предложения и связать их со структурой подгрупп $\langle s_i, s_j \rangle$ и $\langle s_i, s_j, s_k \rangle$.

Задача 14. Определить структуру группы $\Gamma = \langle s_1, s_2, s_3, s_4, s_5 \rangle$ и ее подгруппы $\langle s_i, s_j, s_k, s_l \rangle$.

Много полезной информации об этой задаче можно найти в книге [EGM]. См. также в [Con] очень интересное введение в теорию квадратичных форм.

Понятие совершенной параметризации кругов может быть обобщено на 3-мерный случай. Рассмотрим поле алгебраических чисел $K = \mathbb{Q}[\varepsilon]$, где $\varepsilon = e^{2\pi i/3}$ — кубический корень из единицы. Общий элемент этого поля имеет вид $k = \alpha\varepsilon + \beta\bar{\varepsilon}$, где $\alpha, \beta \in \mathbb{Q}$, а черта означает комплексное сопряжение. Можно рассматривать K как двумерное пространство над полем \mathbb{Q} с базисом ε и $\bar{\varepsilon}$. Операция умножения на k в этом базисе записывается матрицей $M(k) = \begin{pmatrix} -\beta & \alpha - \beta \\ \beta - \alpha & -\alpha \end{pmatrix}$. Определим след и норму числа $k \in K$ как след и определитель $M(k)$ соответственно:

$$\operatorname{tr} k = k + \bar{k} = \alpha + \beta, \quad \|k\|_K^2 = |k|^2 = k\bar{k} = \alpha^2 - \alpha\beta + \beta^2. \quad (8.2.1)$$

Через E мы обозначим подмножество K , выделяемое условиями $\alpha \in \mathbb{Z}, \beta \in \mathbb{Z}$. Элементы E — это целые числа поля K . Существует шесть обратимых целых чисел: $\pm 1, \pm\varepsilon, \pm\bar{\varepsilon}$. Они характеризуются условием $\|k\| = 1$ и называются *единицами* поля K . Известно, что каждое целое число однозначно (по модулю единиц) разлагается в произведение *простых* чисел, то есть таких, которые делятся только на себя и на единицы. Что касается простых чисел, они бывают двух сортов: обычные простые числа вида $p = 3m - 1$ и такие комплексные числа $k = a\varepsilon + b\bar{\varepsilon}$ с целыми a, b , для которых квадрат нормы равен 3 или обычному простому числу вида $3m + 1$.

Начало списка простых чисел поля K имеет вид

$$2, \varepsilon - \bar{\varepsilon}, 5, 2\varepsilon - \bar{\varepsilon}, 11, 3\varepsilon - \bar{\varepsilon}, 17, 3\varepsilon - 2\bar{\varepsilon}, 23, 29, 5\varepsilon - \bar{\varepsilon}, 4\varepsilon - 3\bar{\varepsilon}, \dots$$

Из сказанного следует, что каждый элемент $k \in K$ может быть однозначно (по модулю единиц) записан в виде несократимой дроби $k = \frac{p}{q}$, где $p, q \in E$ не имеют общих множителей, не считая единиц. Можно также записать k в виде $k = \frac{l\varepsilon + m\bar{\varepsilon}}{n}$, где l, m, n — обычные целые числа с $\operatorname{НОД}(l, m, n) = 1$.

ОПРЕДЕЛЕНИЕ 8.1. Пусть D — трехмерный шар в целочисленном \mathbb{R}^2 3-мерном ковре Аполлония. Параметризация множества ∂D точками $\overline{\mathbb{R}^2}$ называется *совершенной*, если точки касания D с остальными шарами в точности соответствуют точкам $\overline{K} \subset \overline{\mathbb{R}^2}$.

Пусть $D_k \in \mathcal{A}$ означает шар, касающийся D в точке $k = \frac{p}{q} \in \overline{K}$. (Если $q = 0$, $k = \infty \in \overline{K}$.)

ТЕОРЕМА 8.1. Совершенные параметризации существуют и обладают следующими свойствами:

а) Пусть $K \ni k = \frac{p}{q}$. Кривизна c_k границы шара D_k дается формулой

$$c_k = \alpha|p|^2 + \beta p\bar{q} + \bar{\beta}\bar{p}q + \gamma|q|^2 + \delta, \quad (8.2.2)$$

где $\alpha, \gamma \in \mathbb{R}$, $\beta \in \mathbb{C}$.

б) Существует такая координатная система (x_1, x_2, x_3) в окружающем \mathbb{R}^3 , что

$$x_i = \frac{\alpha_i|p|^2 + \beta_i p\bar{q} + \bar{\beta}_i \bar{p}q + \gamma_i|q|^2 + \delta_i}{\alpha|p|^2 + \beta p\bar{q} + \bar{\beta}\bar{p}q + \gamma|q|^2 + \delta}. \quad (8.2.3)$$

в) Пусть $k_i = \frac{p_i}{q_i}$, $i = 1, 2$. Шары D_{k_1} и D_{k_2} касаются тогда и только тогда, когда

$$|k_1 - k_2| = \frac{1}{|q_1 q_2|}. \quad (8.2.4)$$

Мы оставляем читателю в качестве нелегкой самостоятельной работы доказать сформулированную теорему и найти ее матричный аналог.

В заключение мы проиллюстрируем теорему 8.1 двумя примерами совершенных параметризаций в 3-мерном ковре Аполлония.

Введем в \mathbb{R}^3 ортогональный базис $1, i, j$. Шару с центром $x + iy + jz$ и радиусом $r > 0$ мы поставим в соответствие эрмитову кватернионную матрицу $\begin{pmatrix} a & b \\ \bar{b} & c \end{pmatrix}$, где $c = \frac{1}{r}$, $b = \frac{x + iy + jz}{r}$, $\bar{b} = \frac{x - iy - jz}{r}$, $a = \frac{x^2 + y^2 + z^2 - r^2}{r}$.

Наш ковер \mathcal{A} будет 3-мерным аналогом ковра-ленты на плоскости. В него входят два обобщенных шара, один из которых — полупространство $z \geq 1$, а другой — полупространство $-z \geq 1$. Они соответствуют матрицам $M_{\pm} = \begin{pmatrix} 2 & \mp j \\ \pm j & 0 \end{pmatrix}$. Далее, наш ковер содержит счетное множество единичных шаров, соответствующих матрицам

$\begin{pmatrix} |v|^2 - 1 & v \\ -\bar{v} & 1 \end{pmatrix}$, где вектор v пробегает решетку $V \subset \mathbb{C}$, порожденную векторами 2ε и $2\bar{\varepsilon}$.

Первый пример совершенной параметризации — это параметризация всех шаров, касающихся плоскости $z = 1$, элементами \bar{K} . А именно, точке $k = \frac{p}{q} \in \bar{K}$ мы ставим в соответствие матрицу

$$M_k = \begin{pmatrix} 4|p|^2 + |q|^2 - 2 & 2p\bar{q} + (1 - |q|^2)j \\ 2\bar{p}q - (1 - |q|^2)j & |q|^2 \end{pmatrix}. \quad (8.2.5)$$

Соответствующий шар касается плоскости в точке

$$t_k = -2\frac{p}{q} + \left(1 - \frac{1}{|q|^2}\right)j$$

и имеет радиус $r = \frac{1}{|q|^2}$.

Второй пример — параметризация всех шаров, касающихся единичного шара с матрицей $M = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$. Здесь формула выглядит так:

$$M_k = \begin{pmatrix} |p|^2 + |q|^2 + 1 & 2p\bar{q} + (|p|^2 - |q|^2)j \\ 2\bar{p}q + (|q|^2 - |p|^2)j & |p|^2 + |q|^2 - 1 \end{pmatrix}. \quad (8.2.6)$$

Соответствующий шар касается единичного шара в точке

$$t_k = \frac{-2p\bar{q} + (|q|^2 - |p|^2)j}{|p|^2 + |q|^2}$$

и имеет радиус $r = \frac{1}{|p|^2 + |q|^2 - 1}$.

ЛИТЕРАТУРА

А. ПОПУЛЯРНЫЕ КНИГИ, ЛЕКЦИИ И ОБЗОРЫ

- [Bar] M. Barnsley, *Fractals everywhere*, Academic Press Inc., Boston, MA, 1988.
- [LGRE] N. Lesmair-Gordon, W. Rood, and R. Edney, *Introducing fractal geometry*, Icon Books, UK, Totem books, USA, 2000.
- [Sod] F. Soddy, *The kiss precise*, *Nature* **137** (1936), 1021.
- [Ste] K. Stephenson, *Circle packing: a mathematical tale*, *Notices Amer. Math. Soc.* **50** (2003), no. 11, 1376–1388.
- [Str1] R. S. Strichartz, *Analysis on fractals*, *Notices Amer. Math. Soc.* **46** (1999), no. 10, 1199–1208.

В. КНИГИ

- [Bea] A. F. Beardon, *The geometry of discrete groups*, Graduate Texts in Mathematics, vol. 91, Springer-Verlag, New York, 1983.
Русский перевод: Бердон А. Геометрия дискретных групп. М.: Наука, 1986.
- [Con] J. H. Conway, *The sensual (quadratic) form*, With the assistance of Francis Y. C. Fung, Carus Mathematical Monographs, vol. 26, Mathematical Association of America, Washington, DC, 1997.
Русский перевод: Конвей Дж. Квадратичные формы, данные нам в ощущениях. М.: МЦНМО, 2008.
- [Cox] H. S. M. Coxeter, *Introduction to geometry*, John Wiley & Sons Inc., New York, 1969.
Русский перевод: Коксетер Г. С. М. Введение в геометрию. М.: Наука, 1966.
- [CH] R. Courant and D. Hilbert, *Methods of mathematical physics. Vol. II*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1989.
Русский перевод: Курант Р., Гильберт Д. Методы математической физики. Т. 2. М.-Л.: ГТТИ, 1951.
- [CM] H. S. M. Coxeter and W. O. J. Moser, *Generators and relations for discrete groups*, 2nd ed., *Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, Band 14*, Springer-Verlag, Berlin, 1965.
Русский перевод: Коксетер Г. С. М., Мозер У. О. Дж. Порождающие элементы и определяющие соотношения дискретных групп. М.: Наука, 1980.
- [Edg] G. A. Edgar, *Measure, topology, and fractal geometry*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1990.
- [EGM] J. Elstrodt, F. Grunewald, and J. Mennicke, *Groups acting on hyperbolic space*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 1998.
Русский перевод: Груневальд Ф., Меннике Й., Эльстродт Ю. Группы, действующие на гиперболическом пространстве. М.: МЦНМО, 2003.

- [Kig] Jun Kigami, *Analysis on fractals*, Cambridge Tracts in Mathematics, vol. 143, Cambridge University Press, Cambridge, 2001.
- [HL] Christina Q. He and Michel L. Lapidus, *Generalized Minkowski content, spectrum of fractal drums, fractal strings and the Riemann zeta-function*, Mem. Amer. Math. Soc. **127** (1997), no. 608.
- [Man] Benoit B. Mandelbrot, *The fractal geometry of nature*, W. H. Freeman and Co., San Francisco, Calif., 1982.
Русский перевод: Мандельброт Б. Фрактальная геометрия природы. М.: Институт компьютерных исследований, 2002.
- [Thu] William P. Thurston, *Three-dimensional geometry and topology. Vol. 1*, Princeton Mathematical Series, vol. 35, Princeton University Press, Princeton, NJ, 1997. Edited by Silvio Levy.
Русский перевод: Тёрстон У. Трехмерная геометрия и топология. М.: МЦНМО, 2001.

C. НАУЧНЫЕ СТАТЬИ

- [AS] D. Aharonov and K. Stephenson, *Geometric sequences of discs in the Apollonian packing*, Алгебра и анализ **9** (1997), no. 3, 104–140; English transl., St. Petersburg Math. J. **9** (1998), 509–545.
- [Bar] M. T. Barlow, *Harmonic analysis on fractal spaces*, Astérisque (1992), no. 206, Exp. No. 755, 5, 345–368. Séminaire Bourbaki, Vol. 1991/92.
- [Bro] R. Brooks, *The spectral geometry of the Apollonian packing*, Comm. Pure Appl. Math. **38** (1985), no. 4, 359–366.
- [BL] A. F. Beardon and L. Lorentzen, *Continued fractions and restrained sequences of Möbius maps*, Rocky Mountain J. Math. **34** (2004), no. 2, 441–466.
- [DK] R. L. Dobrushin and S. Kusuoka, *Statistical mechanics and fractals*, Lecture Notes in Mathematics, vol. 1567, Springer-Verlag, Berlin, 1993, pp. 39–98.
- [FS] M. Fukushima and T. Shima, *On a spectral analysis for the Sierpiński gasket*, Potential Anal. **1** (1992), no. 1, 1–35.
- [GLM⁺] R. L. Graham, J. C. Lagarias, C. L. Mallows, A. R. Wilks, and C. H. Yan, *Apollonian circle packings: number theory*, J. Number Theory **100** (2003), no. 1, 1–45.
- [KS] E. Kasner and F. Supnick, *The Apollonian packing of circles*, Proc. Nat. Acad. Sci. U.S.A. **29** (1943), 378–384.
- [Max] G. Maxwell, *Sphere packings and hyperbolic reflection groups*, J. Algebra **79** (1982), no. 1, 78–97.
- [McM] C. T. McMullen, *Hausdorff dimension and conformal dynamics. III. Computation of dimension*, Amer. J. Math. **120** (1998), no. 4, 691–721.
- [MT] L. Malozemov and A. Teplyaev, *Pure point spectrum of the Laplacians on fractal graphs*, J. Funct. Anal. **129** (1995), no. 2, 390–405.
- [Nev] E. H. Neville, *The structure of Farey series*, Proc. London Math. Soc. (2) **51** (1949), 132–144.
- [Ram] R. Rammal, *Spectrum of harmonic excitations on fractals*, J. Physique **45** (1984), no. 2, 191–206.

- [dR1] G. de Rham, *Sur les courbes limites de polygones obtenus par trisection*, Enseignement Math. (2) **5** (1959), 29–43.
- [dR2] G. de Rham, *Sur une courbe plane*, J. Math. Pures Appl. (9) **35** (1956), 25–42.
- [dR3] G. de Rham, *Un peu de mathématiques à propos d'une courbe plane*, Elemente der Math. **2** (1947), 73–76, 89–97.
- [dR4] G. de Rham, *Sur quelques courbes définies par des équations fonctionnelles*, Univ. e Politec. Torino. Rend. Sem. Mat. **16** (1956/1957), 101–113.
- [Sal] R. Salem, *On some singular monotonic functions which are strictly increasing*, Trans. Amer. Math. Soc. **53** (1943), 427–439.
- [Str2] R. S. Strichartz, *Taylor approximations on Sierpinski gasket type fractals*, J. Funct. Anal. **174** (2000), no. 1, 76–127.
- [Теп] A. V. Teplyaev, *Gradients on fractals*, J. Funct. Anal. **174** (2000), no. 1, 128–154.

D. МАТЕРИАЛЫ В ИНТЕРНЕТЕ

<http://en.wikipedia.org/wiki/fractal>

<http://classes.yale.edu/fractals/>

<http://www.faqs.org/faqs/fractal-faq/>

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	5
----------	---

Часть I. КОВЕР СЕРПИНСКОГО

Глава 1. ОПРЕДЕЛЕНИЕ И ОСНОВНЫЕ СВОЙСТВА	11
1.1. Возникновение и наивное определение	11
<i>Схолия А.</i> Метрические пространства	14
А.1. Определение метрического пространства (14). А.2. Сжимающие отображения (16). А.3. Компактные пространства (17).	
1.2. Самоподобные фракталы	18
<i>Схолия В.</i> Хаусдорфова мера и хаусдорфова размерность	24
Глава 2. ОПЕРАТОР ЛАПЛАСА НА КОВРЕ СЕРПИНСКОГО	29
<i>Схолия С.</i> Оператор Лапласа и гармонические функции	30
С.1. Анализ на римановом многообразии (30). С.2. Оператор как отношение двух квадратичных форм (33). С.3. Оператор Лапласа на римановом многообразии (34).	
2.1. Оператор Лапласа на \mathcal{S}_n	36
2.2. Сравнение спектров Δ_n и Δ_{n-1}	39
2.3. Спектр оператора Лапласа на \mathcal{S}_n	40
Глава 3. ГАРМОНИЧЕСКИЕ ФУНКЦИИ НА КОВРЕ СЕРПИНСКОГО	41
3.1. Основные свойства гармонических функций	41
3.2. Базисные функции χ, φ, ψ, ξ	44
3.3. Продолжение и вычисление функций $\chi(t)$ и $\psi(t)$	48
<i>Схолия D.</i> Производные и интегралы дробного порядка	51
3.4. Некоторые арифметические свойства основных функций	53
3.5. Функции $x(t), y(t)$ и $y(x)$	56
3.6. Гармонический образ ковра \mathcal{S}	60
3.7. Многомерные аналоги ковра \mathcal{S}	61

<i>Схолия Е.</i> Числовые системы	63
Е.1. Стандартные системы (63). Е.2. Нестандартные системы (65). Е.3. Непрерывные дроби (66). Е.4. Обобщенные числовые системы (68).	
3.8. Применения обобщенных числовых систем	69

Часть II. КОВЕР АПОЛЛОНИЯ

Глава 4. КРУГИ НА СФЕРАХ	75
4.1. Теорема Декарта.	75
<i>Схолия F.</i> Конформная группа и стереографическая проекция.	83
F.1. Стереографическая проекция (83). F.2. Конформная группа (85). F.3. Малые размерности (88).	
Глава 5. СТРОГОЕ ОПРЕДЕЛЕНИЕ КОВРА АПОЛЛОНИЯ	93
5.1. Основные факты	93
<i>Схолия G.</i> Числа Фибоначчи	99
5.2. Ковры с неограниченными размерами кругов.	101
5.3. Три интерпретации множества \mathcal{D}	106
5.4. Обобщенная теорема Декарта.	110
Глава 6. АРИФМЕТИЧЕСКИЕ СВОЙСТВА КОВРОВ АПОЛЛОНИЯ	115
6.1. Целочисленные решения уравнения Декарта.	115
<i>Схолия H.</i> Структура некоторых групп, порожденных отражениями	117
6.2. Структура множества $\overline{\mathbb{Q}}$	121
6.3. Рациональная параметризация окружности.	125
6.4. Совершенные параметризации кругов, касающихся данного круга.	134
6.5. Целочисленные ковры Аполлония	137
<i>Схолия I.</i> Формула обращения Мёбиуса	139
Глава 7. ГЕОМЕТРИЧЕСКИЙ И ТЕОРЕТИКО-ГРУППОВОЙ ПОДХОДЫ	145
<i>Схолия J.</i> Плоскость Лобачевского (гиперболическая плоскость)	145
J.1. Первая модель Пуанкаре (145). J.2. Вторая модель Пуанкаре (148). J.3. Модель Клейна (149).	

<i>Оглавление</i>	179
7.1. Действие группы \bar{G} и ковры Аполлония.....	153
7.2. Действие группы Γ на ковре Аполлония	159
Глава 8. Многомерные ковры Аполлония	165
8.1. Общие соображения	165
8.2. Трехмерный ковер Аполлония.....	169
ЛИТЕРАТУРА	173

Александр Александрович Кириллов

ПОВЕСТЬ О ДВУХ ФРАКТАЛАХ

Подписано в печать 10.07.2009 г. Формат 60×90/16. Бумага офсетная №1.

Печать офсетная. Печ. л. 12. Тираж 1000 экз. Заказ №

Издательство Московского центра
непрерывного математического образования
119002, Москва, Бол. Власьевский пер., 11. Тел. (499) 241-74-83

Отпечатано с готовых диапозитивов в ООО «Типография „Сарма“»

Книги издательства МЦНМО можно приобрести в магазине «Математическая книга»
по адресу: Бол. Власьевский пер., д. 11. тел. (499) 241-72-85. E-mail: biblio@mcsme.ru
