

ANALYSIS OF TROPICAL CLIMATE

THEORY AND 3 TROPICAL CLIMATE MECHANIX SYSTEM

THEORY AND 3 TROPICAL
CLIMATE MECHANIX
SYSTEM

А. Б. КАТОК
Б. ХАССЕЛБЛАТ

ВВЕДЕНИЕ В ТЕОРИЮ ДИНАМИЧЕСКИХ СИСТЕМ

С ОБЗОРОМ ПОСЛЕДНИХ ДОСТИЖЕНИЙ

*Перевод с английского
под редакцией А. С. Городецкого*

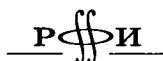
Москва
Издательство МЦНМО
2005

УДК 531.01+517.9

ББК 22.21

К29

Издание осуществлено
при поддержке РФФИ
(издательский проект № 02-01-14088).



A FIRST COURSE IN DYNAMICS

*with a Panorama of Recent
Developments*

BORIS HASSELBLATT

Tufts University

ANATOLE KATOK

The Pennsylvania State University



Каток А. Б., Хасселблат Б.

К29 Введение в теорию динамических систем с обзором последних достижений.— М.: МЦНМО, 2005.— 464 с.: ил.

ISBN 5-94057-063-1

Книга представляет собой введение в теорию динамических систем. Авторы объясняют фундаментальные понятия этой теории и рассматривают их на многочисленных примерах.

Книга предназначена студентам, аспирантам и научным сотрудникам физико-математических специальностей. Представляет большой интерес для специалистов в области нелинейной физики и теории хаоса.

ББК 22.21

Translation from the English language edition: Boris Hasselblat, Anatole Katok, *A first course in dynamics: with a panorama of recent developments*, Cambridge University Press, 2003. ISBN 0-521-58304-7

ISBN 5-94057-063-1



9 785940 570639 >

© Каток А. Б., Хасселблат Б., 2005

© МЦНМО, перевод на рус. яз., 2005

ОГЛАВЛЕНИЕ

Предисловие	6
Глава 1. Введение	8
§ 1.1. Динамика	8
§ 1.2. Динамика в природе	11
§ 1.3. Динамика в математике	28

Часть I

Динамические системы: от простого к сложному

Глава 2. Системы с устойчивым асимптотическим поведением	41
§ 2.1. Линейные отображения и линеаризация	41
§ 2.2. Сжимающие отображения в евклидовом пространстве	42
§ 2.3. Неубывающие отображения интервала и бифуркации	57
§ 2.4. Дифференциальные уравнения	61
§ 2.5. Квадратичные отображения	69
§ 2.6. Метрические пространства	73
§ 2.7. Фракталы	82
Глава 3. Линейные отображения и линейные дифференциальные уравнения	87
§ 3.1. Линейные отображения на плоскости	87
§ 3.2. Линейные дифференциальные уравнения на плоскости	99
§ 3.3. Линейные отображения и дифференциальные уравнения в более высоких размерностях	103
Глава 4. Рекуррентность и равномерное распределение на окруж- ности	108
§ 4.1. Повороты окружности	108
§ 4.2. Некоторые приложения плотности и равномерного распределения	121
§ 4.3. Обратимые отображения окружности	136
§ 4.4. Феномен Кантора	149
Глава 5. Рекуррентность и равномерное распределение в высших размерностях	157
§ 5.1. Сдвиги и линейные потоки на торе	157
§ 5.2. Приложения сдвигов и линейных потоков	166

Глава 6. Консервативные системы	169
§ 6.1. Сохранение фазового объема и рекуррентность	169
§ 6.2. Ньютоновы системы классической механики	176
§ 6.3. Биллиард: определение и примеры	191
§ 6.4. Выпуклые биллиарды	200
Глава 7. Простые системы со сложной структурой орбит	210
§ 7.1. Рост числа периодических точек	210
§ 7.2. Топологическая транзитивность и хаос	220
§ 7.3. Кодирование	227
§ 7.4. Дальнейшие примеры кодирования	237
§ 7.5. Равномерное распределение	247
§ 7.6. Независимость, энтропия, перемешивание	253
Глава 8. Энтропия и хаос	261
§ 8.1. Размерность компактного пространства	261
§ 8.2. Топологическая энтропия	264
§ 8.3. Приложения и обобщения	270

Часть II

Динамические системы: обзор современных результатов

Глава 9. Простая динамика как средство анализа	279
§ 9.1. Введение	279
§ 9.2. Теоремы о неявной и об обратной функции в евклидовом пространстве	280
§ 9.3. Устойчивость трансверсальных неподвижных точек	286
§ 9.4. Решения дифференциальных уравнений	288
§ 9.5. Гиперболичность	295
Глава 10. Гиперболическая динамика	302
§ 10.1. Гиперболические множества	302
§ 10.2. Структура и рост орбит	308
§ 10.3. Кодирование и перемешивание	315
§ 10.4. Статистические свойства	319
§ 10.5. Неравномерно гиперболические динамические системы	324
Глава 11. Квадратичные отображения	325
§ 11.1. Предварительные замечания	325
§ 11.2. Эволюция простого поведения после первой бифуркации	330
§ 11.3. Сложные динамические явления	334
§ 11.4. Гиперболическое и стохастическое поведение	342

Глава 12. Гомоклиническая картина	347
§ 12.1. Нелинейные подковы	347
§ 12.2. Гомоклинические точки	349
§ 12.3. Возникновение подков	351
§ 12.4. Применение подков	353
§ 12.5. Обнаружение гомоклинических пересечений: метод Пуанкаре— Мельникова	357
§ 12.6. Гомоклинические касания	359
Глава 13. Странные аттракторы	361
§ 13.1. Знакомые аттракторы	361
§ 13.2. Соленоид	363
§ 13.3. Аттрактор Лоренца	365
Глава 14. Вариационные методы, закручивающие отображения и замкнутые геодезические	373
§ 14.1. Вариационный метод и биркгофовы периодические орбиты для бильярдов	373
§ 14.2. Биркгофовы периодические орбиты и теория Обри—Мазера для закручивающих отображений	377
§ 14.3. Инвариантные окружности и области неустойчивости	390
§ 14.4. Периодические точки для отображений цилиндра	394
§ 14.5. Геодезические на сфере	397
Глава 15. Динамика, теория чисел и диофантовы приближения	400
§ 15.1. Равномерная распределенность дробных частей для многочленов	400
§ 15.2. Цепные дроби и рациональные приближения	404
§ 15.3. Отображение Гаусса	410
§ 15.4. Однородная динамика, геометрия и теория чисел	414
§ 15.5. Квадратичные формы от трех переменных	420
Литература	423
Приложение	426
§ А.1. Метрические пространства	426
§ А.2. Дифференцируемость	438
§ А.3. Интегрирование по Риману в метрических пространствах	440
Ответы и указания	446
Решения	452
Предметный указатель	457
Указатель имён	463

ПРЕДИСЛОВИЕ

Первая часть данной книги представляет собой введение в теорию динамических систем для студентов старших курсов, а вторая часть содержит обзор некоторых более современных результатов, связанных с динамическими системами. Эти результаты можно рассматривать как приложения и развитие идей, изложенных в первой части. Несмотря на значительные различия в подходе к изложению материала, обе части книги довольно тесно связаны друг с другом. Однако они могут рассматриваться независимо: курс теории может читаться без обзора результатов, а обзор результатов не требует каких-то специальных знаний из курса. Любая из этих частей может быть полезна для ученых и инженеров.

Исправления и другая дополнительная информация могут быть найдены на интернет-странице второго автора.

Введение. Книга начинается с введения, цель которого — заинтересовать читателя предметом книги и представить примеры научных и математических задач, которые могут быть решены с использованием динамики. Эта глава помогает читателю понять, для чего же нужна динамика, хотя и не является необходимой для дальнейшего изложения.

Основной курс. Курс динамики предполагает только знание линейной алгебры, дифференциального исчисления функций многих переменных и интеграла Римана с отдельными доказательствами. Некоторые необходимые результаты обсуждаются в гл. 9 и добавлении.

Отдельные части текста, которые рассматриваются авторами как более трудные для первого прочтения, набраны мелким шрифтом, что означает, что их можно пропустить без ущерба для понимания остального материала. Однако и выделенные части текста не предполагают каких-либо предварительных знаний.

Динамика позволяет описывать долгосрочное поведение в системах, которые развиваются во времени. В данном курсе идеи динамики излагаются последовательно начиная с простых понятий, переходя к сложным вопросам и результатам с доказательствами. В курсе рассматриваются как топологический подход, так и статистический. Нам неизвестны другие пособия, в которых оба подхода были бы представлены на доступном уровне.

Обзор результатов. Обзор динамических систем предполагает использование более серьезного математического аппарата, но это компенсируется меньшей строгостью доказательств, которые являются скорее схемами рассуждений с указанием основных шагов и не включают детальных выкладок. В этой части книги приведены примеры применения идей из курса, связанные с текущими исследованиями в этой области, подкрепленные большим количеством ссылок в тексте.

Дополнительная литература. Наиболее естественное продолжение настоящего курса и обзора результатов представлено в нашей книге «Введение в современную теорию динамических систем»¹, чтение которой позволит читателю углубить свои знания в данной области. Другие источники указаны в конце книги.

Благодарности. Многие из рисунков были выполнены Б. Каток, С. Ферлегер и Р. Гунеш. Данная книга стала такой, какая она есть, во многом благодаря программе MASS (Mathematics Advanced Studies Semesters) в университете штата Пенсильвания осенью 1996 г., когда были опробованы ранние варианты многих частей текста и разработаны многие упражнения. Второй автор также выражает благодарность Центру динамических систем университета штата Пенсильвания за существенную финансовую поддержку на протяжении почти всего времени работы над текстом. Для нас было большой радостью и удовольствием работать с нашим редактором в Cambridge University Press, Лорен Каулз. Ее работа над рукописью отличалась идеальным сочетанием терпения и настойчивости, а полезные замечания из предварительных отзывов, которые она тщательно собрала, помогли определить порядок нашей работы на протяжении последнего года и позволили существенно улучшить содержание.

В заключение мы хотим поблагодарить Кэтлин Хасселблат и Светлану Каток за их поддержку и безграничное терпение.

¹ А. Каток, Б. Хасселблат. Введение в современную теорию динамических систем. М.: Факториал, 1999.

ВВЕДЕНИЕ

Настоящая глава является прелюдией ко всей книге. Она в общих чертах описывает, о чем идет речь в теории динамических систем. Введение содержит большое количество примеров. Некоторые из задач, рассматриваемых в книге позже, впервые появляются именно здесь.

§ 1.1. Динамика

Что такое динамическая система? Динамичность означает, что в системе что-то происходит, что-то изменяется с течением времени. Как вещи изменяются в природе? Работы Галилео Галилея и Исаака Ньютона сыграли ключевую роль в научной революции, центральный принцип которой заключался в следующем: *Природа подчиняется неизменным законам, которые могут быть выражены математически.* Окружающие нас предметы ведут себя и развиваются по определенным раз и навсегда фиксированным правилам. Предыстория динамики — это описание законов механики, интерес к точным наукам и окончательное становление классической и небесной механики. Ньютонская революция основана на том предположении, что законы природы могут быть выражены в математических терминах и физические события могут быть предсказаны и смоделированы с математической точностью. После того как были сформулированы математические законы механики, электричества, магнетизма и термодинамики, другие естественные науки также последовали этому примеру, и даже в общественных науках количественные детерминированные закономерности заняли свое место.

1.1.1. Детерминизм против предсказуемости. Ключевое слово в этом пункте — детерминизм: природа подчиняется неизменным законам. Регулярность движения небесных тел была всегда основным примером порядка в природе:

И сказал Бог: да будут светила на тверди небесной для отделения дня от ночи, и для знамений, и времен, и дней, и годов.

Успехи классической и в особенности небесной механики в XVIII и XIX столетиях выглядели безграничными, и Пьер Симон де Лаплас имел основания сказать следующее (во вступительном абзаце, который он добавил к своему «Философскому сочинению по вероятности» 1812 г.):

Таким образом, мы должны рассматривать настоящее состояние Вселенной как следствие ее предыдущего состояния и как причину последующего.

Ум, которому были бы известны для какого-либо данного момента все силы, одушевляющие природу, и относительное положение всех ее составных частей, если бы вдобавок он оказался достаточно обширным, чтобы подчинить эти данные анализу, обнял бы в одной формуле движения величайших тел Вселенной наравне

с движениями легчайших атомов: не осталось бы ничего, что было бы для него недостоверно, и будущее, также как и прошедшее, предстало бы перед его взором.¹

Энтузиазм, характерный для этой «увертюры 1812 года», хорошо понятен, и это выразительное описание детерминизма является хорошей отправной точкой для понимания одного из основных аспектов теории динамических систем. Кроме того, колоссальная работа Лапласа в области небесной механики на протяжении всей его жизни дала ему право делать такие смелые заявления. Однако в этом утверждении все же имеются некоторые спорные моменты, и центральная задача теории динамических систем и этой книги состоит в том, чтобы установить соотношение между детерминизмом и предсказуемостью, о котором ничего не говорится в утверждении Лапласа. История современной теории динамических систем начинается с Анри Пуанкаре в конце XIX в. Почти через 100 лет после Лапласа он ответил на его утверждение следующим образом:

Если бы мы знали точно законы природы и состояние Вселенной в начальный момент, то мы могли бы точно предсказать состояние Вселенной в любой последующий момент. Но даже и в том случае, если бы законы природы не представляли собой никакой тайны, мы могли бы знать первоначальное состояние только приближенно. Если это нам позволяет предвидеть дальнейшее ее состояние с тем же приближением, то это все, что нам нужно. Мы говорим, что явление было предвидено, что оно управляется законами. Но дело не всегда обстоит так; иногда небольшая разница в первоначальном состоянии вызывает большое различие в окончательном явлении. Небольшая погрешность в первом вызвала бы огромную ошибку в последнем. Предсказание становится невозможным, мы имеем перед собой явление случайное.²

Рассуждения Пуанкаре приводят к принятию точки зрения, которая лежит в основе изучения динамики в настоящее время и которую мы излагаем в этой книге: изучение долгосрочного асимптотического поведения, и особенно поведения качественных характеристик, требует прямых методов, которые не основываются на знании решений в явном виде. И в дополнение к качественному (геометрическому) изучению динамической системы определенную роль играют также вероятностные явления.

Главной причиной интереса к изучению динамических систем послужило их все более возрастающее значение при исследовании процессов, связанных с окружающим нас миром. Многие системы развиваются непрерывно с течением времени (например, системы, встречающиеся в механике), однако в природе имеются и системы, которые развиваются дискретными шагами. Например, мы вскоре опишем модели популяций бабочек, которые развиваются в соответствии с природными циклами. Бабочки живут летом, и мы рассмотрим законы, позволяющие определить размер популяции следующим летом на основе знания размера

¹ *P. S. Laplace. Philosophical Essay on Probabilities. New York: Springer-Verlag, 1995. P. 2.* Имеется перевод: *П. С. Лаплас. Опыт философии теории вероятностей. М., 1908. С. 9.*

² *H. J. Poincaré. Science et méthode. Section IV.II. Flammarion, 1908. См. также: H. J. Poincaré. The Foundations of Science. Science and hypothesis. The value of science. Science and method. Lancaster, PA: The Science Press, 1946. P. 397f.* Имеется перевод: *А. Пуанкаре. Наука и метод // О науке. М.: Наука, 1990. С. 417.*

популяции текущим летом. Можно также изучать системы с непрерывным временем, рассматривая их как системы с дискретным временем. Например, можно отмечать положение луны каждые 24 часа, или же можно следить за тем, где она восходит каждый день. Поэтому мы допускаем, что динамические системы могут развиваться дискретными шагами, когда одно и то же правило последовательно применяется к результату каждого предшествующего шага.

Это допущение важно еще и по другой причине. Такие пошаговые процессы происходят не только в окружающем нас мире, но также и в наших умах. Это случается всякий раз, когда мы проходим одни и те же этапы на пути к получению точного решения. Динамика, примененная к таким процедурам, дает результаты и предлагает методы, которые оказываются полезными в анализе. Мы покажем в этой книге, что важные результаты математического анализа являются следствиями результатов динамики, в том числе даже некоторых довольно простых, например, принцип сжимающих отображений (предложение 2.2.8, предложение 2.2.10, предложение 2.6.10) приводит к теореме об обратной функции (теорема 9.2.2) и теореме о неявной функции (теорема 9.2.3). Методы динамики в таких ситуациях позволяют решать различные задачи с помощью итерационной процедуры последовательной аппроксимации ответа. Динамика естественным образом позволяет понять, к какому результату приводит подобная процедура.

1.1.2. Динамика в анализе. Всякий раз, когда систематическая процедура используется для уточнения решения, точное решение задачи с большой вероятностью может быть найдено с использованием динамики. Чтобы представить себе широчайшие возможности этого подхода, важно понять, что итеративные процессы, к которым может быть применима динамика, не обязательно должны быть связаны с числами. Речь может идти о работе с весьма сложными классами объектов: числа, точки в евклидовом пространстве, кривые, функции, последовательности, отображения и т.д. Возможности многообразны, и во всех этих случаях мы можем с успехом применить динамику. Мы применяем итеративные схемы к функциям в § 9.4, к отображениям в п. 9.2.1 и к последовательностям в § 9.5. Красота этих приложений заключается в элегантности, широком спектре возможностей использования и простоте решений и результатов, полученных с их помощью.

1.1.3. Динамика в математике. В предыдущем пункте были приведены лишь некоторые примеры использования динамических систем для понимания математических структур. Имеются и другие примеры, позволяющие получить результаты в различных разделах математики, используя предположение, что в основе рассматриваемой структуры лежит некоторая динамическая система, которая легко может быть проанализирована или же которая уже была проанализирована. Широта возможностей приложения динамических идей поистине удивительна, поскольку динамика часто позволяет описать очень тонкие и разнообразные явления. Красота применения динамических систем в математике заключается в многообразии поведения изучаемых объектов, в удивительном порядке, который обнаруживается даже в чрезвычайно сложных явлениях, и в открытии неожиданной согласованности между различными областями матема-

тики. Чуть позже в этой вводной главе мы приведем несколько простых примеров приложений такого рода.

Упражнения

В этих упражнениях читателю предлагается поработать с некоторыми простыми итерационными процессами с использованием калькулятора. Выбор этих процессов не случаен — со временем мы вернемся к некоторым из них. В каждом упражнении задана функция f и число x_0 . Задача состоит в том, чтобы рассмотреть последовательность, определенную рекуррентно данным начальным значением и правилом $x_{n+1} = f(x_n)$. Вычислите столько членов последовательности, чтобы можно было предсказать ее дальнейшее поведение. Если последовательность сходится, запишите предполагаемый предел и постарайтесь определить формулу для него. Отметьте, сколько шагов вам потребовалось, чтобы определить характер поведения или получить хорошую аппроксимацию предела.

1.1.1. $f(x) = \sqrt{2+x}$, $x_0 = 1$.

1.1.2. $f(x) = \sin x$, $x_0 = 1$. Используйте *градусную* меру углов на вашем калькуляторе; это означает, что в радианах функция имеет вид $f(x) = \sin(\pi x/180)$.

1.1.3. $f(x) = \sin x$, $x_0 = 1$. Здесь и далее используйте *радианную* меру.

1.1.4. $f(x) = \cos x$, $x_0 = 1$.

1.1.5. $f(x) = \frac{x \sin x + \cos x}{1 + \sin x}$, $x_0 = 3/4$.

1.1.6. $f(x) = \{10x\} = 10x - \lfloor 10x \rfloor$ (дробная часть), $x_0 = \sqrt{1/2}$.

1.1.7. $f(x) = \{2x\}$, $x_0 = \sqrt{1/2}$.

1.1.8. $f(x) = \frac{5+x^2}{2x}$, $x_0 = 2$.

1.1.9. $f(x) = x - \operatorname{tg} x$, $x_0 = 1$.

1.1.10. $f(x) = kx(1-x)$, $x_0 = 1/2$, $k = 1/2; 1; 2; 3,1; 3,5; 3,83; 3,99; 4$.

1.1.11. $f(x) = x + e^{-x}$, $x_0 = 1$.

§ 1.2. Динамика в природе

1.2.1. Кролики-антиподы. Кролики не обитали изначально в Австралии, но приблизительно в 1860 г. 24 диких европейских кролика были ввезены неким Томасом Остином в местечко близ Джилонга в Южной Виктории, что привело к весьма неприятным последствиям. В течение одного десятилетия они уже распространились по всей Виктории, а через 20 лет многомиллионная популяция кроликов опустошила землю этой провинции до такой степени, что была объявлена награда в 25000 фунтов стерлингов за решение возникшей проблемы. К 1910 г. потомки этих кроликов распространились почти по всему континенту. Экологический эффект оказался глубоким, затронул почти всю страну и был даже назван национальной трагедией. Ежегодный ущерб сельскому хозяйству

оценивается в 600 миллионов австралийских долларов. Беспрепятственный рост популяции кроликов приводит к интересному примеру динамической системы.

При моделировании прироста численности мы делаем несколько предположений. Большой размер численности позволяет считать кроликов миллионами, и если x обозначает количество миллионов кроликов, тогда x вовсе не обязательно целое число. Начальное значение x — 0,000024 миллионов кроликов. Итак, мы измеряем размер популяции вещественным числом x . Что касается времени, то в благоприятном климате, как известно, кролики воспроизводятся непрерывно. (Для некоторых популяций это не так, скажем, жизненный цикл и воспроизводство для бабочек строго сезонны; см. п. 1.2.9.) Поэтому мы можем считать переменную времени также вещественным числом t . Таким образом, мы хотим описать число кроликов как функцию времени $x(t)$.

Чтобы понять эту зависимость от времени, посмотрим, чем занимаются кролики: они едят и воспроизводятся. Австралия — большая страна, так что еды для кроликов достаточно, и в течение любого заданного периода времени Δt фиксированный процент от (женской) популяции производит потомство и (меньший) процент умирает от старости (мы считаем, что природных врагов у кроликов нет). Поэтому приращение $x(t + \Delta t) - x(t)$ пропорционально $x(t)\Delta t$ (где коэффициент пропорциональности есть разница между процентом рождаемости и процентом смертности). Переходя к пределу при $\Delta t \rightarrow 0$, мы получаем, что

$$\frac{dx}{dt} = kx, \quad (1.2.1)$$

где k — (фиксированный) относительный коэффициент прироста численности кроликов. Иначе можно записать $\dot{x} = kx$, где точка обозначает дифференцирование по t . Эта модель должна быть хорошо известна из курса математического анализа.

Неизменная окружающая среда (и биология) обеспечивает справедливость этого закона эволюции и позволяет описать рост популяции с помощью динамической системы того вида, который мы изучаем. Дифференциальное уравнение (1.2.1), которое связывает x и коэффициент изменения, легко разрешимо: разделение переменных $\frac{1}{x} dx = k dt$ и интегрирование по t приводит к равенству

$$\ln |x| = \int \frac{1}{x} dx = \int \frac{1}{x} \frac{dx}{dt} dt = \int k dt = kt + C.$$

Таким образом, $|x(t)| = e^C e^{kt}$ с $e^C = |x(0)|$, и мы получаем, что

$$x(t) = x(0)e^{kt}. \quad (1.2.2)$$

Упражнение 1.2.1. Объясните исчезновение знака модуля в вычислениях.

Упражнение 1.2.2. Если $x(0) = 3$ и $x(4) = 6$, вычислите $x(2)$, $x(6)$ и $x(8)$.

1.2.2. Кролики из Пизы. В 1202 г. Леонардо Пизанский рассмотрел менее глобальную задачу о кроликах, которую мы исследуем в примере 2.2.9 и п. 3.1.9. Главное отличие от крупномасштабной австралийской модели, описанной выше, состоит в том, что размер его городского сада ограничивал число кроликов, а для

такого малого числа кроликов прирост популяции происходит не непрерывно, а в существенно различные дискретные периоды времени. Вот как сам Леонардо сформулировал свою задачу:

Сколько пар кроликов могут быть выведены из одной пары за один год?

Некоторый человек имеет пару кроликов, которые содержатся на участке, со всех сторон окруженном стеной. Мы хотим знать, сколько пар можно получить от этой пары за один год, если пара кроликов каждый месяц производит одну новую пару, которая в свою очередь становится способной к воспроизводству на второй месяц после рождения. Допустим, что первая пара производит еще одну пару в первом месяце, тогда через месяц будут иметься 2 пары. Из этих пар одна, а именно первая, производит новую пару на второй месяц, так что через 2 месяца будут иметься уже 3 пары. Две из них в следующем месяце смогут произвести потомство, так что на третий месяц родятся еще 2 пары кроликов. Таким образом, через три месяца будет иметься уже 5 пар. Из них в следующем месяце 3 пары смогут произвести потомство, так что через 4 месяца будет иметься 8 пар и т.д. [Мы получили этот результат], складывая первое число со вторым ($1 + 2$), второе с третьим, третье с четвертым и т.д.³

Другими словами, Леонардо Пизанский нашел рекуррентную формулу для последовательности чисел (пар кроликов) $b_{n+1} = b_{n+1} + b_{n-1}$ и, взяв исходные значения $b_0 = b_1 = 1$, получил последовательность 1, 1, 2, 3, 5, 8, 13, ... Не правда ли, эта последовательность выглядит знакомой? (Подсказка: отца Леонардо звали Боначчи, поэтому сам он был известен как *filius Bonacci*, что означает «сын с хорошим характером», или кратко — Фибоначчи.) Сформулируем вопрос, на который можно легко ответить с минимальным использованием динамики. Как модель Фибоначчи соотносится с непрерывной моделью экспоненциального роста, рассмотренной выше?

Согласно закону экспоненциального роста следует ожидать, что при достаточно больших членах последовательности мы всегда будем иметь $a_{n+1} \approx aa_n$ для некоторой постоянной величины a , не зависящей от n . Если предположить, что на самом деле имеется точное равенство, то рекуррентная формула приведет к соотношениям

$$a^2 b_n = ab_{n+1} = b_{n+2} = b_{n+1} + b_n = (a + 1)b_n$$

и мы получим $a^2 = a + 1$. Тогда значение постоянной прироста a можно будет вычислить с помощью формулы для корней квадратного уравнения.

Упражнение 1.2.3. Вычислите значение a .

Заметим, однако, что мы показали только то, что *если* рост становится в конечном счете экспоненциальным, тогда постоянная прироста — это значение a , определяемое из уравнения $a^2 = a + 1$, а не то, что рост в конечном счете станет экспоненциальным. (Если мы *предположим*, что рекуррентная формула $b_{n+1} = 1$ приводит к экспоненциальному росту, мы можем найти параметр

³Leonardo of Pisa. Liber abaci (1202). Опубликовано в книге: Scritti di Leonardo Pisano. Rome: B. Boncompagni, 1857; см. также: D.J. Struik. A source Book in Mathematics. Princeton, NJ: Princeton University Press, 1986. С. 3. (О задаче Фибоначчи см.: Д.Я. Стройк. Краткий очерк истории математики. 5-е изд. М.: Наука, 1990. — Прим. ред.)

роста при условии, что проделаем вычисления раньше, чем обратим внимание на противоречие.) Динамика предоставляет в наше распоряжение средства, которые позволяют легко проверить свойство экспоненциальности роста различными способами (пример 2.2.9 и п. 3.1.9). В предложении 3.1.11 мы найдем явную формулу для элементов этой рекуррентно заданной последовательности.

Значение этого асимптотического отношения было известно Иоганну Кеплеру. Это число называется золотым сечением или божественной пропорцией. В своей книге 1619 года «Гармония мира» он написал (с. 273):

... имеется отношение, которое никогда полностью не может быть выражено числом и не может быть показано с помощью чисел любым другим способом, кроме как длинной последовательностью чисел, постепенно приближающих его: его точное выражение называется *золотым сечением*, и оно правит повсюду. Следующие отношения являются первыми приближениями божественного соотношения: 1 : 2, 2 : 3, 3 : 5, 5 : 8. Наименее точно оно представлено отношением 1 : 2, более точно — отношением 5 : 8 и еще точнее — если поделить 8 на сумму чисел 5 и 8, т. е. 13.⁴

В примере 15.2.5 мы заметим, что эти дроби Фибоначчи являются наилучшими рациональными приближениями золотого сечения.

Упражнение 1.2.4. Выразите $1 + 1 + 2 + 3 + \dots + b_n$ через b_{n+2} .

1.2.3. Изысканный ужин. Когда-то омары настолько изобиловали в водах Новой Англии, что они служили пищей преимущественно для бедных людей. Случалось даже, что заключенные в штате Мэн бунтовали, требуя в качестве питания что-то другое вместо омаров. В настоящее время омары менее распространены и стали считаться изысканной пищей. Одна (оптимистическая?) модель для объяснения уменьшения уловов предполагает, что улов в любой заданный год должен равняться среднему улову за два предыдущих года.

Обозначая через a_n число омаров, пойманных в n -м году, мы можем описать эту модель простым рекуррентным соотношением

$$a_{n+1} = a_{n-1}/2 + a_n/2. \quad (1.2.3)$$

В качестве начальных условий можно брать уловы штата Мэн 1996 и 1997 гг., которые составляли 16,435 и 20,871 (метрических) тонн соответственно. Эта рекуррентная формула схожа с рекуррентной формулой для чисел Фибоначчи, но в данном случае нельзя ожидать экспоненциального роста. Из вида формулы видно, что все будущие уловы должны находиться между двумя начальными данными. Действительно, 1997 год был рекордным. В предложении 3.1.13 мы получим явную формулу для будущих уловов, в которых улов выражается как функция от n .

Эта задача, так же как и задача Фибоначчи о кроликах, является примером задачи с дискретным временем. Имеются много других примеров, где подобные модели возникают естественно. Такой сценарий из биологии популяций обсуждается в п. 1.2.9. Другие биологические примеры относятся к области ге-

⁴J. Kepler. Epitome of Copernican Astronomy. Harmonies of the world. Prometheus Books, 1995.

нетики (частота генов) или эпидемиологии. Социологи также используют модели с дискретным временем (цены товаров, скорость распространения слухов, теории обучения, моделирующие количество информации, приобретенной в течение данного времени).

1.2.4. Появление нового листа. Слово *филлотаксис* (phyllotaxis) происходит от слов *phyllo* — лист и *taxis* — порядок или расположение. Этот термин обозначает порядок расположения листьев на стеблях или других частей растения на более крупных его частях. В качестве примеров можно привести семена подсолнечника и сосновой шишки. Красивое описание подобных моделей дается Гарольдом Скоттом Макдональдом Кокстером в его «Введении в геометрию». Правильные модели встречаются часто, как можно видеть на примерах подсолнечника и ананасов.

Для некоторых видов деревьев листья на ветках также располагаются закономерно. Структура изменяется в зависимости от вида дерева. Самая простая структура — листья, чередующиеся на противоположных сторонах ветки. Это называется (1, 2)-филлотаксисом: последовательные листья отделены друг от друга половиной оборота вокруг ветки. Этой закономерности подчиняются листья вяза и орешника⁵. Оборот между смежными листьями может также составлять $\frac{2}{3}$ полного оборота вокруг ветки, что в соответствии с нашей терминологией называется (2, 3)-филлотаксисом. Эта модель встречается у бука. Дубы имеют модель (3, 5), тополи — (5, 8), а ивы — (8, 13). Разумеется, на практике рост листьев не всегда в точности соответствует модели, и на некоторых растениях по мере их роста происходят переходы от одной модели к другой.

Семена подсолнечника, имеющие форму ромба, упакованы плотно и регулярно. В их расположении отчетливо просматривается спиралевидный рисунок, даже 2 спирали, идущие в противоположных направлениях. Количества витков в двух спиралях являются последовательными числами Фибоначчи. Семена еловой шишки также располагаются по спирали, но скорее на конусе, а не на плоскости. Для них также можно различить 2 спирали, количества витков для которых снова являются последовательными числами Фибоначчи.

На поверхности ананасов также можно наблюдать спиралевидный рисунок, и, поскольку их поверхность составлена из ячеек, имеющих форму, близкую к шестиугольнику, существуют три возможных направления, в которых можно наблюдать спирали. Соответственно в более сложных моделях можно было бы встретить 5, 8 и 13 спиралей, например 5 спиралей, имеющих наклон слегка вправо, 8 — наклон влево и 13 спиралей, имеющих сильный наклон вправо.

Эти красивые закономерности были замечены в XIX в. Однако объяснение того, почему возникают именно такие закономерности, появилось не сразу. На самом деле исследования не закончены полностью и сейчас.

Имеется модель, которая ведет к объяснению того, как возникает явление филлотаксиса. Процесс роста почек листьев или зародышей семян происходит следующим образом: они растут из центра и затем перемещаются дальше от

⁵По которому второй автор этой книги должен быть экспертом. (Фамилия Hasselblat означает по-немецки «лист орешника». — *Прим. перев.*)

него согласно трем правилам, предложенным в 1868 г. ботаником Вильгельмом Фридрихом Бенедиктом Хофмейстером, который в то время был профессором и директором ботанического сада в Гейдельберге.

1. Новые почки формируются на одинаковых расстояниях друг от друга и как можно дальше от предыдущих.

2. Почки двигаются по радиальному направлению от центра.

3. Коэффициент роста уменьшается по мере продвижения наружу.

Физический эксперимент, разработанный для воспроизведения трех *правил Хофмейстера*, приводит к спиральным структурам типа Фибоначчи, так что можно предположить, что из этих правил следует вывод о наличии спиральных структур. Это недавно было выполнено с использованием методов, описываемых в данной книге⁶.

Покажем, как при этом может помочь динамика. Построим модель развития листьев с учетом правил Хофмейстера. Для этого рассмотрим семейство $N + 1$ концентрических окружностей радиуса r^k ($k = 0, \dots, N$), где r — коэффициент роста, и поместим почку на каждую окружность. Угол (относительно начала координат) между соседними почками обозначим через θ_k . Возможные модели теперь параметризуются углами $(\theta_0, \dots, \theta_N)$. Это означает, что «пространство растений» является *тором*, см. п. 2.6.4. Когда новая почка появляется на единичной окружности, все другие почки перемещаются на одну окружность дальше от центра. Угол между новой почкой и предшествующей ей почкой зависит от всех предыдущих углов, так что мы получаем отображение, связывающее старые углы θ_k и новые углы Θ_k :

$$\Theta_0 = f(\theta_0, \dots, \theta_N), \quad \Theta_1 = \theta_0, \dots, \Theta_N = \theta_{N-1}.$$

Теперь необходимо задать функцию f , чтобы иметь возможность учесть в модели первое правило Хофмейстера. Один из способов сделать это состоит в задании естественной потенциальной энергии, отражающей «отталкивание» между почками, и в выборе $f(\theta_0, \dots, \theta_N)$ как ее минимума. Естественный потенциал равен

$$W(\Theta) = \sum_{k=0}^N U(\|r^k e^{i\theta_k} - e^{i\Theta}\|),$$

где $U(x) = 1/x^s$ для некоторого $s > 0$. Более простой потенциал, который приводит к тому же качественному поведению, это $W(\Theta) = \max_{0 \leq k \leq N} U(\|r^k e^{i\theta_k} - e^{i\Theta}\|)$.

В любом случае можно показать, что правильные спирали (т. е. $\theta_0 = \dots = \theta_N$) являются притягивающими неподвижными точками (см. п. 2.2.7) этого отображения. Это означает, что спирали появляются естественным образом. Кроме того, из результатов анализа следует появление чисел Фибоначчи.

1.2.5. Разновидности экспоненциального роста. В примере с популяцией кроликов п. 1.2.1 естественно ожидать, что коэффициент роста k в уравнении $\dot{x} = kx$ будет положительным. Этот коэффициент, однако, является разностью

⁶ P. Atela, C. Golé, S. Holton. A dynamical system for plant pattern formation: A rigorous analysis // Journal of Nonlinear Science. 2002. V. 12, no. 6. P. 641—676.

между коэффициентами рождаемости и смертности. Для населения некоторых западных стран коэффициент рождаемости ниже показателя смертности. И в этой ситуации применяется та же самая модель, но при $k < 0$, а решение $x(t) = x(0)e^{kt}$ описывает экспоненциально убывающую популяцию.

То же самое дифференциальное уравнение $\dot{x} = kx$ используется в многочисленных простых моделях, поскольку оно представляет собой самое простое дифференциальное уравнение с одной переменной.

Еще один известный пример — процесс радиоактивного распада. Экспериментально установлено, что определенная часть радиоактивного вещества распадется за фиксированное время. Как и выше, это обстоятельство приводит к уравнению $\dot{x} = kx$, где $k < 0$. В этой ситуации постоянная k часто выражается через *период полураспада*, т. е. такой период времени T , что $x(t + T) = x(t)/2$. В зависимости от характера вещества этот период времени может составлять доли секунды или тысячи лет. Эта модель играет важную роль при решении проблем утилизации радиоактивных отходов, которые часто имеют большой период полураспада, и радиоактивного заражения. Биологические лаборатории используют радиоактивный фосфор, который имеет период полураспада несколько дней, как маркер. Пролитый радиоактивный фосфор на лабораторном столе обычно закрывают плексигласом примерно на две недели, после чего излучение становится достаточно незначительным. С другой стороны, положительным эффектом радиоактивного распада является возможность применения радиоизотопного анализа для определения возраста органических или геологических пород. В отличие от биологии популяций экспоненциальная модель радиоактивного распада не нуждается ни в каких усовершенствованиях, чтобы более адекватно описывать реальные явления. Она является точным законом природы.

Упражнение 1.2.5. Выразите период полураспада через k и наоборот.

Значение этого простого дифференциального уравнения простирается далеко за пределы моделей, в которых оно непосредственно появляется. Это уравнение также возникает при изучении более сложных дифференциальных уравнений в качестве аппроксимации, которая может пролить свет на поведение в более сложных ситуациях. Такой подход, называемый *линеаризацией*, очень важен в динамических системах.

1.2.6. Судный день. Теперь вернемся к задаче прироста населения. Фактические данные о численности населения показывают, что население земли растет с увеличивающейся скоростью. Поэтому следует несколько изменить основную модель, с тем чтобы принять во внимание прогресс цивилизации. Предположим, что с ростом населения растущее число ученых может постепенно уменьшать показатель смертности и увеличивать показатель рождаемости. При грубом предположении, что эти изменения делают относительный коэффициент прироста населения равным некоторой малой степени x^ϵ существующей численности населения x (а не постоянной k), мы приходим к уравнению

$$\frac{dx}{dt} = x^{1+\epsilon}.$$

Как и прежде, это уравнение легко решается разделением переменных:

$$t + C = \int x^{-1-\epsilon} dx = -x^{-\epsilon}/\epsilon,$$

где $C = -x(0)^{-\epsilon}/\epsilon$, так что $x(t) = (x(0)^{-\epsilon} - \epsilon t)^{-1/\epsilon}$, а это решение становится бесконечным для $t = 1/(\epsilon x(0)^\epsilon)$. Иначе говоря, мы получаем настоящий демографический взрыв!

Отсюда видно, что в реальной биологической ситуации эту модель следует изменить. Ясно, например, что наше предположение об увеличивающемся коэффициенте роста населения было нереальным (в конечном счете, все ресурсы ограничены). Однако как пример дифференциального уравнения эта модель интересна: мы видим, что бывают разумно выглядящие дифференциальные уравнения, которые тем не менее имеют расходящиеся решения.

1.2.7. Хищники. Причина, по которой кролики не распространились по Европейскому континенту, кроется в том, что в Европе всегда были хищники, которые убивали кроликов. Это обстоятельство влияет на динамику популяций интересным образом, поскольку популяции хищников и их жертв взаимодействуют: малое число кроликов обрекает популяцию хищников на голодание, что в свою очередь увеличивает популяцию кроликов. Таким образом, можно ожидать устойчивого равновесия — или, может быть, колебаний численности.

Множество моделей взаимодействующих популяций хищников и их жертв были предложены независимо Альфредом Лоткой и Вито Вольтерра. Одна из простых моделей — *уравнение Лотки—Вольтерра*

$$\frac{dx}{dt} = a_1x + c_1xy,$$

$$\frac{dy}{dt} = a_2x + c_2xy,$$

где $a_1, c_2 > 0$ и $a_2, c_1 < 0$, т. е. x обозначает численность популяции жертв, которая росла бы сама по себе ($a_1 > 0$), но уменьшается за счет хищников ($c_1 < 0$), а y обозначает численность популяции хищников, которая вырождалась бы сама по себе ($a_2 < 0$), но растет за счет жертв ($c_2 > 0$). Естественно считать x и y положительными величинами. Эта модель предполагает, что нет никакой задержки времени между причинами и результатами, вызванной периодом беременности или высиживания яиц. Это предположение разумно, если рассматриваемый период времени не слишком мал. Кроме того, выбор непрерывного времени лучше всего соответствует существенному пересечению поколений. Популяции с непрерывающимися поколениями будут рассмотрены ниже.

Для уравнения Лотки—Вольтерра имеется равновесие видов в точке $(a_2/c_2, a_1/c_1)$. Для любых других начальных значений численностей популяций мы получаем колебания числа хищников и их жертв. Чтобы убедиться в этом, можно показать с использованием дифференцирования сложной функции, что функция

$$E(x, y) := x^{-a_2} e^{-c_2 x} y^{a_1} e^{c_1 y}$$

постоянна вдоль орбит, т. е. $\frac{d}{dt}E(x(t), y(t)) = 0$. Это означает, что решения уравнения Лотки—Вольтерра лежат на кривых $E(x, y) = \text{const}$. Все эти кривые замкнуты.

1.2.8. Ужас вакуума. Изучение уравнения Лотки—Вольтерра наводит на мысль совершить небольшое отступление и рассмотреть физическую систему, которая демонстрирует иной вид колебательного поведения. Нелинейные колебания в этой системе вызвали большой интерес, и она оказалась важной для ряда исследований в динамике.

Голландский инженер Балтазар Ван дер Пол из научной лаборатории фирмы Филипс на фабрике по выпуску электроламп в Эйндховене смоделировал вакуумную трубу дифференциальным уравнением

$$\frac{d^2x}{dt^2} + \epsilon(x^2 - 1)\frac{dx}{dt} + x = 0,$$

которое может быть переписано (с заменой $y = \frac{dx}{dt}$) в виде

$$\frac{dx}{dt} = y,$$

$$\frac{dy}{dt} = \epsilon(1 - x^2)y - x.$$

Если $\epsilon = 1$, то начало координат является отталкивающей неподвижной точкой, или репеллером (определение 2.3.6). Однако решения не растут неограниченно, поскольку имеется периодическое решение, соответствующее вращению вокруг начала координат. Действительно, для $\epsilon = 0$ имеются только такие решения, а для $\epsilon = 1$ одна из этих окружностей сохраняется в несколько искаженной форме, а все другие решения приближаются к этому решению при $t \rightarrow +\infty$, что ясно видно из рисунка (рис. 1.2.1), полученного с использованием численных методов. Эта предельная кривая называется *предельным циклом*.

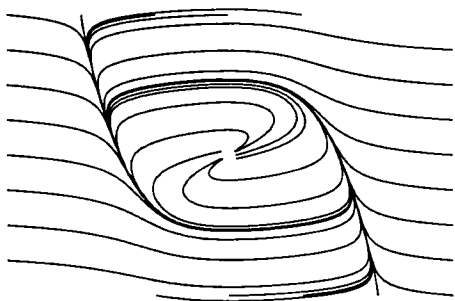


Рис. 1.2.1. Уравнение Ван дер Пола

Немного отступив от темы, заметим, что изучение вакуумной трубы позволяет получить модели необычайной сложности. В 1927 г. Ван дер Пол и Дж. Ван дер Марк сообщили об экспериментах с излучателем переменного напряжения, сконструированным из конденсатора и неоновой лампы (нелинейный элемент), на который периодически подается напряжение. (Это напряжение соответствует

периодическому члену в правой части уравнения Ван дер Пола.) Исследователей заинтересовало то обстоятельство, что в отличие от линейного осциллятора (такого как скрипичная струна), который дает частоты, кратные основной частоте, эти колебания были делителями основной частоты, т. е. половина, третья часть и т. д. до $1/40$ -й части основной частоты, по мере увеличения напряжения. Они получили эти частоты, прослушивая «определенным образом подключенный к системе телефон», и описали свои результаты следующим образом:

Часто в телефонной трубке перед скачком частоты на следующий, более низкий уровень слышен нерегулярный шум. Однако это побочный эффект, а главный результат эксперимента — регулярно уменьшающееся значение частоты колебаний, кратные основной частоте.

Этот нерегулярный шум был одним из первых экспериментальных примеров явления, позднее получившего название хаос, но время для его исследования тогда еще не пришло⁷.

1.2.9. Еще один эффект бабочек⁸. Динамика популяций естественно описывается моделями с дискретным временем, если различные поколения не пересекаются друг с другом. Это было в некотором смысле искусственным ограничением в задаче, изложенной Леонардо Пизанским (п. 1.2.2). Для многих популяций это естественное явление, особенно для популяций насекомых в умеренных поясах, включая насекомых-вредителей. Относительно удобный пример такого рода — колония бабочек в изолированной местности с постоянным сезонным циклом (неизменные законы и отсутствие внешнего воздействия). Пересечений между текущим поколением (поколением этого лета) и следующим (следующим летом) нет вообще. Нам хотелось бы знать, как изменяется размер популяций от лета к лету. На самом деле существует множество факторов окружающей среды, которые воздействуют на популяцию, но, считая законы неизменными, мы можем гарантировать, что численность популяции следующим летом зависит только от ее численности этим летом и эта зависимость одна и та же каждый год. Это означает, что единственный параметр в этой модели — численность популяции. Поэтому с точностью до выбора фиксированных констант закон эволюции системы должен точно определять размер популяции следующим летом как функцию размера популяции только этим летом. Сам закон эволюции системы будет следовать из той модели, которую мы изберем в соответствии с нашим пониманием связанных с популяцией биологических процессов.

1. Экспоненциальный рост. Например, вполне правдоподобно, что большая совокупность должна откладывать больше яиц и давать большее потомство в следующем году, пропорциональное текущей популяции. Обозначая текущую численность популяции через x , мы получаем, что численность в следующем году равна $f(x) = kx$, где k — положительное число, равное среднему количеству потомков для одной бабочки. Если обозначить численность популяции в i -м году

⁷ B. van der Pol, J. van der Mark. Frequency demultiplication // Nature. 1927. V. 120. P. 363—364.

⁸ Это намек на утверждение Эдварда Лоренса (см. § 13.3), согласно которому взмах крыла бабочки в Рио-де-Жанейро может вызвать через неделю тайфун в Токио.

через x_i , то легко видеть, что $x_{i+1} = f(x_i) = kx_i$ и, в частности, что $x_1 = kx_0$, $x_2 = kx_1 = k^2x_0$, и т. д., т. е. $x_i = k^i x_0$. Иными словами, численность популяции растет экспоненциально. Это очень напоминает экспоненциальную модель роста, которая была проанализирована нами в случае непрерывного времени.

2. Конкуренция. Следует отметить, что экспоненциальная модель роста довольно быстро ведет к большим популяциям. Если подходить к проблеме более реалистично, следует принять во внимание то обстоятельство, что большая популяция столкнется с проблемой ограниченных ресурсов питания. Недоедание или голодание приведет к уменьшению числа бабочек, способных в определенное время к кладке яиц. Как результат мы получим относительно малую популяцию бабочек на следующий год.

Самый простой закон, учитывающий подобные разумные качественные свойства популяций, задается формулой $f(x) = k(1 - \alpha x)x$, где x обозначает текущее число бабочек. Это правило самое простое, поскольку к коэффициенту роста k добавлена только линейная поправка. Здесь коэффициент α описывает скорость уменьшения численности потомства из-за конкуренции. Другими словами, $1/\alpha$ — это максимально возможное число бабочек, т. е. если в этом году имеется $1/\alpha$ бабочек, то они съедят всю доступную пищу еще перед тем, как отложить яйца, а следовательно, вынуждены будут голодать, что приведет к вымиранию бабочек в следующем году. Таким образом, если вновь обозначить через x_i размер популяции бабочек в i -м году, начиная с $i = 0$, тогда закон эволюции системы задается формулой $x_{i+1} = kx_i(1 - \alpha x_i) =: f(x_i)$. Это детерминированная математическая модель, в которой каждое будущее состояние (размер колонии бабочек) может быть вычислено исходя из размера колонии в текущем году. Один недостаток этой модели состоит в том, что колонии численностью больше чем $1/\alpha$, как оказывается, дают отрицательные размеры популяции на следующий год, чего можно было бы избежать, используя, скажем, модель $x_{i+1} = x_i e^{k(1-x_i)}$. Но легкость анализа делает более простую модель более популярной, и это сыграло значительную роль в понимании того, что простые модели могут иметь сложное долгосрочное поведение⁹.

Одна особенность, напоминающая о модели с экспоненциальным ростом, состоит в том, что если численность популяции намного меньше предельного значения, то ее рост является по существу экспоненциальным: если $\alpha x \ll 1$, то $1 - \alpha x \approx 1$ и, таким образом, $x_{i+1} \approx kx_i$, следовательно, $x_n \approx k^n x_0$, но это приближенное значение выполняется, только пока размер популяции остается малым. В этом имеется интуитивный смысл: малая популяция, в отличие от многочисленной, не страдает от борьбы за пищу.

Заметим, что в предыдущем абзаце мы допустили некоторую вольность речи: последовательность $x_n \approx k^n x_0$ экспоненциально возрастает, если $k > 1$. Если

⁹Как видно из названия, эта мысль была четко выражена в оказавшей большое влияние на математиков статье: R. M. May. Simple Mathematical Models with Very Complicated Dynamics // Nature. 1976. V. 261. P. 455—467. В этой статье также была предложена для изучения квадратичная модель. Хорошее представление об использовании моделей в разных разделах биологии может дать работа: J. Gleik. Chaos, Making a New Science. Viking Press, New York, 1987. P. 78ff.

это не так, колония бабочек становится вымирающей. Выбор между этими возможностями определяется интересным взаимодействием между коэффициентом рождаемости и предельно возможной численностью популяции.

3. *Замена переменных.* Для упрощения анализа этой системы удобно сделать простую замену переменных, которая устранил параметр α . Мы описываем здесь эту операцию достаточно подробно, потому что замена переменных является важным средством в динамике.

Запишем закон эволюции системы в виде $x' = kx(1 - \alpha x)$, где x — размер популяции в некотором году, а x' — размер популяции в следующем году. Производя замену координат $y = \alpha x$, мы получаем

$$y' = \alpha x' = \alpha kx(1 - \alpha x) = ky(1 - y).$$

Другими словами, теперь нам требуется проитерировать отображение $g(y) = ky(1 - y)$. Отображения f и g связаны соотношением $g(y) = h^{-1}(f(h(y)))$, где $h(y) = y/\alpha = x$. Это можно прочесть так: «перейти от новой переменной к старой, применить старое отображение, а затем вернуться к новой переменной».

В результате этой замены переменной коэффициент α , отражающий влияние конкуренции, заменяется на 1. Поскольку мы не связаны с выбором определенных единиц измерения, можно снова обозначить переменную через x , а отображение через f .

4. *Логистическое уравнение.* Таким образом, мы получили модель рассматриваемой системы, которая определяется итерациями отображения

$$f(x) = kx(1 - x).$$

Это отображение f называется *логистическим отображением* (или логистическим семейством, поскольку в нем имеется параметр), а уравнение $x' = kx(1 - x)$ называется логистическим уравнением. Термин «логистический» происходит от французского слова *logistique*, которое в свою очередь происходит от *logement*, расквартирование солдат. Мы также называем эту совокупность отображений *квадратичным семейством*. Она была впервые рассмотрена в 1845 г. бельгийским социологом и математиком Верхалстом¹⁰.

Как уже отмечалось, случай $k \leq 1$ приводит к неизбежному вымиранию. Для $k < 1$ это очевидно, потому что $kx(1 - x) < kx$, и для $k = 1$ это также нетрудно проверить, хотя в этом случае скорость вымирания не является экспоненциальной. Большие значения k , наоборот, должны были бы приводить к большой популяции. Но, быть может, это и не так. Дело в том, что слишком большая популяция должна будет произвести на свет не столь многочисленное поколение. Можно ожидать, что через некоторое время популяция должна достигнуть определенного разумного размера, при котором имеется баланс между способностью к воспроизведению потомства и конкуренцией.

Упражнение 1.2.6. Докажите, что случай $k = 1$ приводит к вымиранию.

¹⁰ P.-F. Verhulst. Recherches mathématiques sur la loi d'accroissement de la population // Nouvelles Mémoires l'Académie Royale des Sciences et Belles-lettres de Bruxelles. 1845. V. 18. P. 1—38.

Заметим, что, в отличие от более простой модели экспоненциального роста, в данном случае мы не стали выписывать явную формулу для x_n в терминах x_0 . Эта формула содержит многочлены степени 2^n . Даже если и удастся написать такие формулы для каких-либо разумных значений n , они не будут информативны. Но все же мы сможем в свое время весьма детально описать поведение этой модели. Сейчас же имеет смысл приступить к исследованию возможных типов поведения. Пока не ясно, имеет ли значение начальный размер популяции. Однако изменение параметра k , конечно же, должно приводить к другому результату, или во всяком случае на это можно надеяться, потому что было бы довольно печально, если бы модель всегда предсказывала вымирание. Разумные границы для параметра k — от 0 до 4. (Для $k > 4$ модель предсказывает, что популяция размером $1/2$ через два года станет отрицательной, что не имеет никакого смысла с точки зрения биологии. Это наводит на мысль, что придется внести несколько более точные (нелинейные) поправки.)

5. Эксперименты. При увеличении коэффициента k становится более вероятным, что популяция стабилизируется, т. е. не произойдет вымирания. Так что мы приступим к исследованию модели для некоторого $k > 1$. Для определенности начнем со значения $k = 2$, как раз посередине между 0 и 4.

Упражнение 1.2.7. Проведите несколько итераций функции $2x(1 - x)$, начиная с $x = 0,01$, пока не станет ясен характер поведения модели.

Начиная с малой популяции, мы получаем устойчивый рост и в конечном счете приходим к популяции численностью $1/2$. Именно такое поведение следует ожидать от хорошей модели. Заметим, что устойчивые состояния удовлетворяют уравнению $x = 2x(1 - x)$, единственными решениями которого являются 0 и $1/2$.

Упражнение 1.2.8. Проведите несколько итераций функций $1,9x(1 - x)$ и $2,1x(1 - x)$, начиная с $x = 0,01$, пока не станет ясен характер поведения модели.

Если k лишь немного меньше 2, поведение практически такое же, а если k слегка больше 2, оно очень сходно, за исключением того что значения немного осциллируют, приближаясь к устойчивому состоянию.

Упражнение 1.2.9. Проведите несколько итераций функций $3x(1 - x)$ и $2,9x(1 - x)$, начиная с $x = 0,01$, пока не станет ясен характер поведения модели.

Для $k = 3$ поведение почти такое же, однако популяция приходит к устойчивому состоянию несколько иначе. Имеются довольно существенные колебания слишком больших и слишком малых популяций, которые затихают медленно, в то время как для значений k , близких к 2, имелся лишь намек на такое поведение и колебания быстро утихали. Однако в пределе устойчивое состояние все еще преобладает.

Упражнение 1.2.10. Проведите несколько итераций функции $3,1x(1 - x)$, начиная с $x = 0,01$, пока не станет ясен характер поведения модели.

Для $k = 3,1$ мы, как и прежде, встречаем колебания слишком больших и слишком малых популяций. Они на самом деле становятся немного меньше, но на сей

раз полностью не затихают. Имея простую программу, можно провести достаточно большое число итераций и убедиться, что устойчивое состояние не достигается.

Упражнение 1.2.11. Проведите несколько итераций функций $3,1x(1 - x)$, начиная с $x = 0,66$, пока не станет ясен характер поведения модели.

В предыдущем случае, вообще говоря, имеется возможность того, что колебания все-таки утихают, но настолько медленно, что численные эксперименты не могут этого показать. Поэтому для контроля мы проведем те же самые итерации, начиная со среднего двух значений. Если наша гипотеза верна, итерации должны подходить к устойчивому состоянию. Но это не так. Мы наблюдаем колебания, которые растут до тех пор, пока их размер не станет таким же, как и прежде.

Эти колебания устойчивы! Это первая из наших моделей популяций, которая имеет устойчивое не монотонное поведение. Независимо от того, с какого значения мы начинаем, популяция с рождаемостью 3,1 лишь немного «слишком плодородна» и каждый следующий год продолжает сталкиваться с перенаселением, небольшим, но систематическим.

Посмотрев, какие значения k уже были исследованы, мы можем заключить, что осталось изучить поведение модели только при значениях, близких к $k = 4$. Но чтобы получить более полную картину, лучше еще раз попробовать провести итерации для значений k , расположенных ближе к 3. По крайней мере будет интересно увидеть, становятся ли колебания больше с ростом k . Этого следует ожидать. Но насколько вырастают колебания, мы пока не знаем.

Упражнение 1.2.12. Проведите несколько итераций функций $3,45x(1 - x)$ и $3,5x(1 - x)$, начиная с $x = 0,66$, пока не увидите ясной закономерности.

Поведение становится более сложным вблизи $k = 3,45$. Вместо простых колебаний между двумя значениями здесь появляется еще одно движение вокруг каждого из этих значений. Колебания теперь включают четыре соответствующих размера популяции: «Большая, малая, большая, Малая», повторяющихся в 4-цикле. Период колебаний удвоился.

Упражнение 1.2.13. Проведите аналогичные вычисления, выбирая значения параметра немного больше чем 3,5.

Экспериментатор, хорошо владеющий численными методами, увидит здесь определенную модель поведения: при довольно небольшом увеличении параметра период снова удваивается и появляются восемь размеров популяции, через которые модель неуклонно циркулирует. Еще менее значительное увеличение параметра приводит к периоду 16, и поведение продолжает усложняться. Продолжая незначительно увеличивать параметр, мы наблюдаем удвоение периодов, которые становятся степенями числа 2. Это похоже на то, что мы наблюдаем в случае линейного осциллятора (струна скрипки, струя воздуха в духовых инструментах или трубах органа): имеется частота с высокими гармониками, которые вдвое, втрое, в четыре раза больше основной частоты. Здесь же частота последовательно разделяется, и мы получаем *субгармоники*, что по существу есть нелинейное явление.

Возникает вопрос, продолжает ли период удваиваться вплоть до $k = 4$.

Упражнение 1.2.14. Проведите несколько итераций функции $3,83x(1 - x)$, начиная с $x = 0,5$, пока не станет ясен характер поведения модели.

При исследовании значения $k = 3,83$ мы замечаем несколько другое поведение: снова имеется периодическая модель, к которой мы, кажется, уже привыкли. Но период равен 3, т. е. не является степенью числа 2. Так что эта модель принципиально иная. Мы не наблюдаем периодов, являющихся степенями числа 2, так что они, видимо, исчезли еще при меньших значениях числа k .

Упражнение 1.2.15. Попробуйте провести итерации для $k = 3,828$.

Никакого очевидного образца поведения здесь найти не удастся.

Упражнение 1.2.16. Попробуйте провести итерации для $k = 4$.

И здесь не видно явной закономерности.

6. Перспективы. Попробовав провести эксперименты для нескольких значений параметра в самой простой из возможных нелинейных моделей популяций, мы столкнулись с поведением, которое сильно различается для разных значений параметра. Даже в случаях, когда поведение относительно очевидно, мы не можем объяснить, почему оно имеет именно такой вид. Почему периоды лишь некоторое время удваиваются? При каком значении параметра появляются колебания периода 3? И наконец, в экспериментах с бесчисленным множеством других значений параметра мы сталкиваемся с поведением, которое не можем эффективно описать просто из-за недостатка слов. Пока мы мало что можем сказать, кроме того, что в этих случаях числа, кажется, могут получаться какими угодно.

Мы вернемся к этой модели позже (§2.5, п. 7.1.2, п. 7.4.3 и гл. 11), чтобы объяснить некоторые основные механизмы, которые вызывают подобные виды поведения в случае квадратичного семейства $f_k(x) = kx(1 - x)$. Мы не будем приводить исчерпывающий анализ, который охватывал бы все значения параметра, но динамика этих отображений станет довольно понятной. В книге будут рассмотрены важные понятия, которые необходимы для описания сложного поведения, наблюдаемого в этой ситуации, а также во многих других важных случаях.

Но уже сейчас это чисто численное исследование дает нам определенную информацию. Первый вывод, который можно сделать, состоит в том, что *простые системы могут демонстрировать сложное долгосрочное поведение*. Мы получили рассмотренный пример из линейного, произведя самую незначительную замену, и немедленно столкнулись с таким сложным поведением, которое не поддается описанию. Поэтому, хотя это может показаться неожиданным, такое сложное поведение должно быть весьма распространенным.

Другой вывод состоит в том, что необходимо искать способы понимания, описания и объяснения такого богатого и сложного поведения. Действительно, важные результаты, описанные в этой книге, относятся к системам, для которых явное вычисление невозможно или бесполезно. Мы видим, что даже при отсутствии идеально рассчитанных результатов для всего промежутка времени можно сформулировать точные и полезные качественные и количественные утверждения

относительно таких динамических систем. В книге изучаются понятия, подходящие для описания таких же сложных явлений как те, с которыми мы встретились в этом примере. К подробному изучению этого частного примера мы приступаем в §2.5, где исследуются простые виды поведения, которые имеются для малых значений параметра. В п. 7.1.2 и 7.4.3 изучаются большие значения параметра. Для них асимптотическое поведение наиболее хаотично. В гл. 11 мы излагаем некоторые идеи, которые позволяют понять, при каких промежуточных значениях параметра происходят переходы к максимальной сложности.

В качестве интересного замечания укажем, что аналогичная модель популяции с непрерывным временем (которая может применяться для других видов) не обладает таким сложным поведением (см. п. 2.4.2).

1.2.10. Искра вдохновения. Еще один пример динамики в природе — светлячки. Вот, возможно, самое раннее описание этого замечательного явления, которое встречается в заметках экспедиции сэра Фрэнсиса Дрейка, датированных 1577 г.:

Наш командующий ... причалил к некоему маленькому острову к югу от Целебеса, ... обильно поросшему густым высоким лесом ... каждую ночь среди деревьев по всему острову показывались несметные стаи крутящихся в воздухе огненных червячков, размером не превосходящих обычных английских мух, которые так ярко светились, что каждая веточка дерева была подобна пылающей свече.¹¹

Более ясное описание необычайных свойств светлячков было дано Энгельбертом Кампфером, доктором из восточной Вестфалии, который совершил десятилетнее путешествие по России, Персии, юго-восточной Азии и Японии. Проплывая вниз по реке Чао Прая (Мейнам) из Бангкока, 6 июля 1690 г. он наблюдал следующее явление:

Жуки-светляки (cicindelaе) представляют еще одно замечательное зрелище. Они располагаются на некоторых деревьях, подобно огненному облаку, и, что самое удивительное, целая стая этих насекомых, занявшая одно дерево и усеявшая все его ветви, иногда одновременно прекращает свечение, а в следующий момент вновь начинает светиться, причем это явление происходит необычайно регулярно, как будто они находятся в вечной систоле и диастоле.¹²

Итак, в некоторых районах множество светлячков определенного вида, расположившись на одном или нескольких кустарниках, синхронизируют свечение, превращая свой древесный дом в замечательную картину, похожую на Рожде-

¹¹ *R. Hakluyt*. A selection of the principal voyages, traffiques and discoveries of the English nation / Edited by Laurence Irving. New York: Knopf, 1926.

¹² *E. Kämpfer*. The history of Japan / J. G. Scheuchzer, ed. London: Scheuchzer, 1727. Перевод не очень хороший. Немецкий оригинал, очевидно, оставался неопубликованным в течение столетий: «Einen zweiten sehr angenehmen Anblick geben die Lichtmücken (cicindelaе) welche einige Bäume am Ufer mit einer Menge, wie eine brennende Wolke, beziehn. Es war mir besonders hiebei merkwürdig, daß die ganze Schaar dieser Vögel, so viel sich ihrer auf einem Baume verbunden, und durch alle Aeste desselben verbreitet haben, alle zugleich und in einem Augenblick ihr Licht verbergen und wieder von sich geben, und dies mit einer solchen Harmonie, als wenn der Baum selbst in einer beständigen Systole und Diastole begriffen wäre». (Geschichte und Beschreibung von Japan (1677—79). Internet-Edition by Wolfgang Michel // Engelbert-Kaempfer-Forum. Kyushu University, 1999.).

стенскую елку. Это настолько поразительное явление, что в течение долгого времени оно считалось почти таким же недостоверным, как рассказы про драконов и морских монстров. Еще до 1938 г. описания светлячков не принимались всерьез биологами. Только с появлением новых возможностей для путешествий и увеличением скорости передвижения те, кто сомневался, смогли наконец увидеть это явление своими глазами¹³. Даже когда ученые убедились, что это явление действительно происходит в природе, потребовалось не одно десятилетие, чтобы понять его механизм. Первоначально высказывалось предположение, что такое единообразное поведение вызывается некоторым необнаруженным периодическим воздействием внешней среды, но на самом деле эти светлячки естественным образом мерцают примерно в одном и том же ритме, что вместе с тенденцией к гармонии с соседями и приводит к тому, что вся колония вспыхивает совершенно синхронно.

Аналогичная ситуация наблюдается при изучении наших суточных ритмов, при которых периодические изменения в нашем теле (цикл сна) синхронизируются со внешними сигналами о смене дня и ночи. В отсутствии часов и каких-либо других показателей времени суток биологические часы человека возвращаются к естественному периоду, который для большинства людей несколько больше 24 часов. Внешние сигналы воздействуют на нейроны и гормоны, которые составляют наш сложный внутренний осциллятор, и слегка ускоряют его. В этом случае скорость приспособления к новому режиму довольно велика. Даже самое большое нарушение суточного ритма организма при смене часовых поясов обычно проходит за несколько дней, т. е. за несколько циклов.

Подобные системы служат примерами спаренных осцилляторов, которые встречаются и в других многочисленных обличьях. Пара Земля-Луна также может рассматриваться как подобная система, если искать ответ на вопрос, почему мы всегда видим одну и ту же сторону Луны, т. е. почему вращения Луны вокруг своей оси и вокруг Земли синхронизированы. Здесь простое приливное трение послужило механизмом, который веками приводит вращение Луны вокруг своей оси в соответствие с вращением вокруг Земли и в конце концов синхронизирует также и вращение Земли, так что день будет длиться месяц или месяц будет длиться день, что сделает Луну геостационарным спутником. Интересно, что в некоторый момент времени удлинившиеся дни будут соответствовать нашим внутренним часам, так что можно считать, что сейчас эволюция человека опередила свое время.

Мы еще вернемся к системам, составленным из двух простых осцилляторов, в п. 4.4.5, где с помощью относительно простых соображений придем к выводу, что такого сорта синхронизация в некотором смысле типична¹⁴.

¹³Об этом говорится в статьях: *J. Buck*. Synchronous rhythmic flashing of fireflies // *Quarterly review of Biology*. 1938. V. 13, no. 3. P. 301—314; II // *Quarterly review of Biology*. 1988. V. 63, no. 3. P. 265—289. Эти статьи содержат приведенные здесь цитаты и много других сообщений о светлячках, поступающих с разных континентов.

¹⁴Мы не будем детально рассматривать спаренные линейные осцилляторы. Вопрос о светлячках рассмотрен в работе: *R. Mirollo, S. Strogatz*. Synchronization of Pulse-Coupled Biological Oscillators // *SIAM Journal of Applied Mathematics*. 1990. V. 50, no. 6. P. 1645—1662.

Упражнения

1.2.17. В 1900 г. общее население Земли составляло 1,65 миллиардов человек, а в 1950 г. оно уже было равно 2,52 миллиардам человек. Используя экспоненциальную модель (уравнение 1.2.2), попробуйте предсказать численность населения в 1990 г. и год, в который население достигнет 6 миллиардов. (Фактическая численность населения в 1990 г. составила приблизительно 5,3 миллиарда, а в июле 1999 г. оно достигло 6 миллиардов. Таким образом, рост мирового населения ускоряется.)

1.2.18. Обозначим через a_n число последовательностей из нулей и единиц длины n , которые не имеют двух последовательных нулей. Покажите, что $a_{n+1} = a_n + a_{n-1}$. (Заметим, что это та же самая рекуррентная формула, что и для чисел Фибоначчи, и что $a_1 = 2$ и $a_2 = 3$.)

1.2.19. Определите $\lim_{n \rightarrow \infty} a_n$ в (1.2.3), если $a_0 = 1$ и $a_1 = 0$.

§ 1.3. Динамика в математике

Здесь мы приведем несколько примеров математических задач, для которых знание динамических систем позволяет получить интересные результаты.

1.3.1. Геронические попытки вычисления вавилонских корней. Около 250 г. до н. э. в своем учебнике *Metrica* Герон Александрийский при вычислении площади треугольника со сторонами 7, 8, и 9 впервые получил формулу $S^2 = s(s-a)(s-b)(s-c)$, где a, b, c — стороны треугольника и $2s = a + b + c$. Для того чтобы вычислить получающийся квадратный корень из $12 \cdot 5 \cdot 4 \cdot 3 = 720$, он воспользовался следующим методом, который, возможно, был известен вавилонянам за 2000 лет до этого:

Поскольку $[z = 720$ не имеет рациональной стороны [т. е. 720 не является полным квадратом], мы можем получить его сторону с весьма незначительной погрешностью следующим образом. Так как ближайший следующий квадрат есть 729, а его сторона $[x = 27$, разделим 720 на 27. Мы получим $[y = 26\frac{2}{3}$. Прибавляя к этому числу 27, получим $53\frac{2}{3}$, а половина этой величины будет составлять $[x' = \frac{1}{2}(x + y) = 26\frac{1}{2}\frac{1}{3}$. Таким образом, сторона квадрата 720 приближенно равна $26\frac{1}{2}\frac{1}{3} \dots$. Если мы хотим уменьшить погрешность, ... следует взять $[x' = \frac{1}{2}(x + y) = 26\frac{1}{2}\frac{1}{3} = 26\frac{5}{6}$ вместо $[x = 27$, и, проделав подобные выкладки, мы получим, что погрешность стала еще меньше...¹⁵

Герон использовал этот метод для нахождения квадратного корня из z . Для этого достаточно найти квадрат с площадью z , тогда его стороны будут иметь длину $\sqrt{z}$. Геометрическое описание этой процедуры таково: в качестве первого

¹⁵Т. L. Heath. History of Greek Mathematics: from Aristarchus to Diophantus. Dover, 1981. P. 324. Эта последовательность приближений встречается также в вавилонских текстах, как отмечается в книге: Б. Л. Ван дер Варден. Пробуждающаяся наука. М.: ГИФМЛ, 1959, где приводится геометрическая интерпретация. Некоторый вариант данного метода был известен Архимеду.

приближения искомого квадрата мы берем прямоугольник со сторонами x и y , где x — какое-либо хорошее приближение ответа и $xy = z$. (Если z не так велико, как число из задачи Герона, можно просто взять $x = 1$, $y = z$.) Чтобы перейти от этого прямоугольника площади z к другому прямоугольнику той же площади, стороны которого меньше отличаются от искомого значения, следует заменить стороны x и y на $(x + y)/2$ (среднее арифметическое) и, чтобы получить ту же площадь, $2xy/(x + y)$ (это число называется средним гармоническим значений x и y). Построение можно описать как повторное применение функции

$$f(x, y) = \left(\frac{x + y}{2}, \frac{2xy}{x + y} \right) \quad (1.3.1)$$

двух переменных, начиная с $(x_0, y_0) = (z, 1)$ (или $(x_0, y_0) = (27, 26\frac{2}{3})$, как в примере Герона). По-видимому, Архимед использовал один из вариантов этого метода. Описанная процедура удобна тем, что пары чисел, получаемые на каждом шаге, находятся по разные стороны от точного результата (поскольку на каждом шаге сохраняется соотношение $xy = z$), так что имеется возможность контролировать ошибку в явном виде. Даже перед началом применения этой процедуры первое приближение Герона заключает результат в пределы между $26\frac{2}{3}$ и 27.

Упражнение 1.3.1. Найдите приближенное значение $\sqrt{4}$. Для этого вычислите числа (x_i, y_i) , $0 \leq i \leq 4$, используя описанный метод, начиная с $(1, 4)$ и полагая расстояние равным 2.

Упражнение 1.3.2. Прделайте еще один шаг приближенного вычисления значения $\sqrt{720}$ по методу Герона и, используя калькулятор, определите точность этого приближения.

Упражнение 1.3.3. Выбирая в качестве начального значения 1, определите, сколько шагов этого метода потребуется для получения приближения значения $\sqrt{720}$, лучшего, чем начальное предположение Герона 27.

После того как эта процедура применяется несколько раз, числа x_n и y_n , получающиеся в результате, становятся почти равными и поэтому близкими к $\sqrt{z}$.

С начальным значением Герона его первое приближение было достаточно хорошим (отклонение числа $26\frac{5}{6}$ от $\sqrt{720}$ составляет не более 0,002%), и, по всей видимости, он никогда не выполнял повторные приближения, которые сам предложил. Этот метод замечателен не только тем, что он работает, но и тем, что работает так быстро. Однако почему он работает? И почему он работает так быстро? И насколько быстро он работает? На эти вопросы мы сможем легко ответить, как только приступим к изучению динамических систем (п. 2.2.8).

1.3.2. Поиск корней. Многие задачи, требующие численного решения, несложно и удобно переформулировать как задачи нахождения решения уравнения вида $f(x) = 0$ для некоторой подходящей функции f . Мы опишем два хорошо известных метода, в которых используется функция одной переменной.

1. Бинарный поиск. Имеется ситуация, в которой мы можем быть уверены, что решение существует: теорема о промежуточном значении из математического

анализа утверждает, что если $f: [a, b] \rightarrow \mathbb{R}$ — такая непрерывная функция, что $f(a) < 0 < f(b)$ или $f(b) < 0 < f(a)$, т. е. $f(a)f(b) < 0$, то существует такое $c \in (a, b)$, что $f(c) = 0$.

Упражнение 1.3.4. Покажите, что приведенная формулировка теоремы эквивалентна стандартной формулировке (непрерывная функция на отрезке принимает все промежуточные значения).

Однако уверенности в существовании решения недостаточно, чтобы найти само решение или, по крайней мере, определить с достаточной точностью, где оно расположено. Тем не менее, имеется простой метод нахождения корня. Учитывая, что $f(a) < 0 < f(b)$, рассмотрим середину отрезка $z = (a + b)/2$.

Случай 0. Если $f(z) = 0$, мы нашли корень. Если $f(z) \neq 0$, имеются следующие возможности.

Случай 1. Если $f(z) > 0$, заменим отрезок $[a, b]$ отрезком $[a, z]$, который в два раза короче исходного, но все еще содержит корень по теореме о промежуточном значении, поскольку $f(a) < 0 < f(z)$. Повторим ту же самую процедуру на новом отрезке.

Случай 2. Если $f(z) < 0$, заменим $[a, b]$ на отрезок $[z, b]$, который также в два раза короче исходного, и применим процедуру на этом новом отрезке.

В результате бинарного поиска мы получаем последовательность вложенных отрезков, длины которых уменьшаются в два раза на каждом шаге. Каждый из этих отрезков содержит решение, так что мы получаем все более точные приближения, и предел правых (или левых) конечных точек является решением уравнения.

Заметим, что, хотя эта процедура и является итерационной, она все же не определяет динамическую систему, во всяком случае систему, действующую на числах. Можно было бы считать, что полученная динамическая система действует на отрезках, в концах которых функция f принимает значения разных знаков.

Упражнение 1.3.5. Выполните три шага этой процедуры для функции $f(x) = x - \cos x$ на $[0, 1]$. Получите приближенное решение и определите его точность.

Этот метод надежен: он позволяет получать все более точные приближения решения с гарантированной скоростью. Более того, эта скорость велика, и погрешность легко вычисляется. Например, после 9 шагов погрешность составляет меньше чем $(b - a)/1000$.

2. Метод Ньютона. Метод Ньютона (или метод Ньютона—Рафсона) также применяется для решения задачи нахождения нулей функций. Это более интересный метод, чем бинарный поиск: он оригинален и может работать довольно быстро, но не всегда применим.

Чтобы применять этот метод, нужно предположить, что функция f , нуль которой мы ищем, дифференцируема и, конечно, что этот нуль существует. Как обычно, мы начинаем с некоторого приближения корня x_0 . Каким взять это приближение, зависит от предпочтений исследователя и от поставленной задачи. Можно использовать схематичный график, а в некоторых ситуациях помогает бинарный поиск. В последнем случае несколько шагов дают превосходное начальное значение.

Метод Ньютона обычно позволяет получать значительно более точное приближение решения. Если функция линейна, то, учитывая, что тангенс угла наклона графика к оси абсцисс равен производной, мы из начального значения x_0 немедленно получим точное решение. Если функция f дифференцируема, то ее график хорошо приближается касательными. Поэтому мы можем записать уравнение касательной к графику функции в точке с абсциссой x_0 , где x_0 — начальное значение, и найти точку пересечения этой касательной с осью абсцисс, которую обозначим через x_1 . Это и есть улучшенное приближение. Проведенные построения можно описать следующей формулой:

$$x_1 = F(x_0) := x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Упражнение 1.3.6. Проверьте, что эта формула соответствует приведенным выше геометрическим построениям.

Упражнение 1.3.7. Примените этот метод четыре раза к уравнению $x^2 - 4 = 0$ с начальным приближением 1. Сравните с упражнением 1.3.1. (См. также упражнения 1.3.18 и 1.3.19.)

Эта простая процедура может применяться неоднократно как итерирование функции F . В результате мы получаем последовательность все более точных приближений (во всяком случае, мы можем надеяться на это). В п. 2.2.8 будет предложен простой критерий, гарантирующий, что этот метод приведет к успеху.

Упражнение 1.3.8. В нескольких упражнениях из § 1.1 встречаются примеры применения метода Ньютона. Найдите эти примеры и определите, для решения каких уравнений в них используется данный метод.

Так как метод Ньютона определяет динамическую систему, он и изучается как динамическая система. Это в значительной степени обусловлено тем, что некоторые начальные приближения приводят к ситуациям со сложным асимптотическим поведением. В некоторых случаях, особенно при работе с комплексными числами, можно получить красивые картины с помощью приближенных вычислений. Важное обобщение метода Ньютона заключается в его применении к точкам в функциональных пространствах. Эта процедура обычно называется методом Колмогорова—Арнольда—Мозера (или методом КАМ). Этот метод позволил получить новые результаты на пути к решению вопроса о стабильности нашей солнечной системы. Это великолепный пример того, как знание простой асимптотики динамической системы в некотором вспомогательном пространстве позволяет глубже понять структуру другой динамической системы.

1.3.3. Замкнутые геодезические. Если бы пилот держал штурвал самолета прямо, имея при этом достаточно горючего, самолет летал бы вокруг земли по большой окружности, все время возвращаясь в точности к исходной точке. Такой же эксперимент можно было бы провести с наземным транспортным средством, но здесь возникли бы сложности с из-за наличия гор, океанов, тропических лесов и т. п. Идеализированной моделью такого эксперимента является свободное движение материальной точки по поверхности сферы. Поскольку внешние силы

отсутствуют (и мы предполагаем нулевое трение), такая материальная точка движется с постоянной скоростью без изменения направления. Вполне очевидно, что она будет периодически возвращаться в исходное положение на сфере. Итак, имеется бесконечно много путей, по которым свободное движение осуществляется периодически.

Но что произойдет, если сфера из нашего эксперимента не будет такой круглой, как идеальная сфера? Она может иметь небольшие выбоины или даже быть сильно искаженной. Можно было бы украсить ее грибоподобным выростом или даже придать ей форму штанги. Не позволено только делать каких-либо разрезов или склеиваний поверхности. Итак, мы рассматриваем гладкую, но не шарообразную «сферу». В этом случае не имеется никаких очевидных причин для того, чтобы свободно двигающаяся материальная точка возвращалась в исходное положение. Почти любая деформация сферы порождает множество непериодических движений. Можно задать вопрос, который оказывается довольно трудным: имеется ли все еще бесконечно много путей на данной деформированной сфере, свободное движение вдоль которых является периодическим?

Замечательная особенность свободного движения материальной точки заключается в том, что путь движения всегда является самой короткой кривой между любыми двумя точками на поверхности, которые находятся не слишком далеко друг от друга. (Очевидно, что замкнутый путь не может быть самой короткой кривой от точки до нее самой.) Это свойство очевидно для круглой сферы, когда такие кривые являются большими окружностями, однако оно выполняется и в общем случае, и такие кривые называются *геодезическими*. Поэтому вопрос, указанный выше, можно также сформулировать в терминах геодезических: на любой ли сфере, независимо от того, как она деформирована, имеется бесконечно много замкнутых геодезических?

Этот вопрос относится к геометрии и был сформулирован довольно давно. Он был решен (не так давно) с использованием теории динамических систем. В п. 6.2.8 мы объясняем, как геодезические связаны с динамикой, а в § 14.5 обрисовываем в общих чертах подход к решению этого вопроса.

1.3.4. Первые цифры степеней числа 2. В качестве иллюстрации возможностей динамики в разрешении тонких и запутанных вопросов рассмотрим вполне безобидную последовательность степеней двойки. Приведем первые пятьдесят членов этой последовательности:

2	2048	2097152	2147483648	2199023255552
4	4096	4194304	4294967296	4398046511104
8	8192	8388608	8589934592	8796093022208
16	16384	16777216	17179869184	17592186044416
32	32768	33554432	34359738368	35184372088832
64	65536	67108864	68719476736	70368744177664
128	131072	134217728	137438953472	140737488355328
256	262144	268435456	274877906944	281474976710656
512	524288	536870912	549755813888	562949953421312
1024	1048576	1073741824	1099511627776	1125899906842624.

Хотя последовательность и определялась элементарным образом, этот список выглядит довольно сложным. Однако имеются некоторые интересные особенности, на которые следует обратить внимание. Например, последние цифры повторяются периодически: 2, 4, 8, 6. Очевидно, что так и должно быть, поскольку последняя цифра следующей степени определяется последней цифрой предыдущей степени, поэтому, как только появляется одно повторение, оно перерастает в периодическую модель. (Кроме того, мы видим, что последняя цифра всегда четная и ненулевая.)

Подобное рассуждение показывает, что последние две цифры вместе должны также в конечном счете начать периодически повторяться: согласно приведенным выше замечаниям для последних двух цифр имеется не больше 40 возможностей, и, поскольку последние две цифры следующей степени определяются последними двумя цифрами предыдущей степени, чтобы получить периодическую модель, достаточно найти только одно повторение. И действительно, если мы посмотрим на нашу последовательность, то легко заметим, что последние две цифры образуют следующую периодическую последовательность с периодом двадцать, начиная со второго члена: 04, 08, 16, 32, 64, 28, 56, 12, 24, 48, 96, 92, 84, 68, 36, 72, 44, 88, 76, 52.

Заметим, что эта последовательность имеет несколько интересных особенностей. Сумма первого и одиннадцатого членов составляет 100, так же как и сумма второго и двенадцатого, третьего и тринадцатого членов и т. д. Один из способов получения этой последовательности таков. Начнем с 04 и будем применять следующее правило: если текущее число меньше 50, удвоим его, в противном случае удвоим разность между 100 и нашим числом. Аналогичные свойства наблюдаются для более простой последовательности 2, 4, 8, 6, рассмотренной выше. Теперь посмотрим на последовательность *первых* цифр. Из того же самого списка

2	2048	2097152	2147483648	2199023255552
4	4096	4194304	4294967296	4398046511104
8	8192	8388608	8589934592	8796093022208
16	16384	16777216	17179869184	17592186044416
32	32768	33554432	34359738368	35184372088832
64	65536	67108864	68719476736	70368744177664
128	131072	134217728	137438953472	140737488355328
256	262144	268435456	274877906944	281474976710656
512	524288	536870912	549755813888	562949953421312
1024	1048576	1073741824	1099511627776	1125899906842624

мы видим, что первые цифры в 50 членах равны

2481361251
 2481361251
 2481361251
 2481361251
 2481371251.

Эта последовательность близка к периодической, но в последней строке появляется небольшое отклонение, так что чисто периодической модели нет — и нет никаких причин ожидать таковую. (Если вычислить дальнейшие члены последовательности, подобное поведение сохранится: на каких-то этапах будут возникать небольшие поправки.)

Поскольку эта последовательность не является такой правильной, как предыдущая, здесь может быть полезен статистический подход. Посмотрим на *частоту* каждой цифры: как часто данная цифра появляется в этом списке? Можно составить следующую таблицу:

цифра:	1	2	3	4	5	6	7	8	9
число появлений:	15	10	5	5	5	4	1	5	0.

Эти частоты выглядят несколько неравномерными. В частности, цифры «7» и «9» — самые «непопулярные». Цифра «7» появляется в нашей последовательности только один раз и лишь на 46-м месте, а «9» вообще появится впервые только как первая цифра числа 2^{53} . Для первых 100 членов последовательности частоты немного меньше отличаются друг от друга, но все равно они оказываются меньше для больших цифр.

Таким образом, все девять цифр могут быть первыми цифрами некоторой степени числа 2. Однако мы хотели бы знать больше. Появляется ли каждая цифра бесконечно много раз? Если да, то можно ли указать какую-либо закономерность в появлении цифр? Какие из цифр появляются наиболее часто?

Прежде чем перейти к обсуждению этих вопросов, мы должны сформулировать их более точно. Для этого для каждой цифры d и каждого натурального числа n обозначим через $F_d(n)$ число тех степеней 2^m , $m = 1, \dots, n$, которые начинаются с d . Таким образом, мы только что перечислили десять значений $F_d(50)$. Частота, с которой d появляется как первая цифра среди первых n степеней числа 2, равна $F_d(n)/n$. Тогда один из наших вопросов можно сформулировать так: имеет ли каждая из величин $F_d(n)/n$ предел при $n \rightarrow \infty$ и, если эти пределы существуют, как они зависят от d . Если мы получим ответы на эти вопросы, можно задать аналогичные вопросы про степени числа 3 и сравнить их предельные частоты.

В предложении 4.2.7 мы докажем существование этих пределов и приведем формулу для них, из которой, в частности, будет следовать, что все предельные частоты положительны и что они уменьшаются с ростом d . Таким образом, вопреки тому, что наблюдается для первых 50 степеней (но в соответствии с тем, что можно увидеть для первых 100 степеней), цифра «7» в конечном счете будет появляться чаще, чем «8». Соотношения между соответствующими пределами для степеней чисел 3 и 2 также весьма интересны.

1.3.5. Последние цифры многочленов. В предыдущем примере мы успешно изучили вопрос о последних цифрах и заметили, что динамика может помочь понять поведение первых цифр. Рассмотрим теперь другую задачу, связанную с последовательностями целых чисел, для которой можно поставить подобные вопросы относительно последних цифр.

Вместо экспоненциальной последовательности рассмотрим последовательность $x_n = n^2$ для $n \in \mathbb{N}_0$. Последние цифры будут иметь вид 01496569410, а затем начнут периодически повторяться.

Упражнение 1.3.9. Докажите, что эти цифры повторяются периодически.

Упражнение 1.3.10. Объясните, почему эта последовательность является *палиндромической*, т.е. она не изменится, если записать ее в обратном порядке.

Эта последовательность почти столь же проста, как и предыдущая, поэтому попробуем заменить ее на последовательность $x_n = n^2 p/q$ для некоторых $p, q \in \mathbb{N}$. Если q отлично от 1, не все члены последовательности будут целыми числами, так что мы будем рассматривать цифру, стоящую перед десятичной запятой. Можно немного поэкспериментировать, однако и здесь легко видеть, что мы опять получаем периодическую модель, период которой не превосходит $10q$. Причина кроется в следующем: разность

$$a_{n+10q} - a_n = (n + 10q)^2 p/q - n^2 p/q = 10(2np + 10pq)$$

является целым числом, кратным 10, так что цифра перед запятой (так же, как и все цифры после запятой) одна и та же для a_{n+10q} и a_n .

Упражнение 1.3.11. Докажите, что исходные $10q$ результатов образуют палиндромический ряд.

Это был интересный, но не слишком сложный результат. Естественно попытаться заменить p/q иррациональным числом, поскольку в этом случае можно ожидать наличия «бесконечного периода», т.е. отсутствия какой-либо периодичности.

Рассмотрим последовательность $x_n = n^2 \sqrt{2}$. Последовательность последних цифр (перед запятой) начинается со следующих 100 членов: 4776493564160220725775169007481218481107379985035540580084923206134316133205911072577527011950343171.

Нет никаких очевидных причин ожидать периодичности или существования какой-либо очевидной закономерности. Как видно, здесь появляются все возможные цифры. Однако вопросы, которые мы поставили относительно первых цифр степеней числа 2, уместны и в этом случае: все ли цифры встречаются бесконечно часто? Обладают ли они определенными относительными частотами? Относительные частоты определяются, как и прежде. Обозначим через $P_n(d)$ число появлений последней цифры d в множестве $\{n^2 \sqrt{2}\}_{i=0}^{n-1}$ и рассмотрим $P_n(d)/n$ для больших n . Для первых 100 значений мы получаем следующие частоты:

$i:$	0	1	2	3	4	5	6	7	8	9
$P_{100}(i)/100:$	0,14	0,15	0,09	0,10	0,09	0,11	0,06	0,13	0,06	0,07.

Этот список никак не помогает нам ответить на поставленные вопросы, как, впрочем, и подобный список для больших значений n .

Динамика может полностью и строго разрешить эти вопросы, так же как и множество других подобных. В этом частном примере оказывается, что все

относительные частоты сходятся к $1/10$. Таким образом, мы имеем пример *равномерного распределения*, являющегося одной из центральных парадигм динамики, так же как и самой природы. Мы приводим решение задачи распределения последних цифр в § 15.1.

1.3.6. Клеточные автоматы. В 1980-х гг. была популярна игра, известная под названием «жизнь». Она была предназначена для моделирования простой популяции каких-либо организмов, живущих в определенной местности. Каждый из организмов находится в некоторой точке фиксированной решетки, например в точке на плоскости с целочисленными координатами, и может иметь несколько состояний здоровья. В простейшем варианте такие организмы могут иметь только два состояния «присутствует» и «отсутствует» (1 или 0). Но можно также рассматривать модель с большим числом возможных состояний, включая, например, состояния типа «болезненный» или «счастливый». Правила игры заключаются в том, что популяция определенным образом изменяется через дискретные промежутки времени. Каждый организм проверяет состояния некоторых из своих соседей (до некоторого расстояния) и в зависимости от них изменяет соответственно свое собственное состояние. Например, правило могло бы быть таким: если все непосредственные соседи находятся в состоянии «присутствует», организм умирает (перенаселение). Возможно, то же самое случается, если нет соседей вообще (слишком одинокий организм). Эта игра была так популярна, потому что из относительно простых правил можно было получить довольно любопытное поведение и с помощью компьютеров, даже довольно старых, можно было легко проследить путь многих поколений за короткое время.

Если число ячеек конечно, тогда с нашей точки зрения в системе не происходит ничего интересного. Она имеет лишь конечное число состояний, так что через некоторое время какое-то состояние должно повториться. Поскольку правила остаются неизменными, модель проходит через ту же последовательность состояний снова и снова. Независимо от того, какими сложными могут быть модели и какими длинными могут быть циклы, это дает полное качественное описание долгосрочного поведения.

Если имеется бесконечно много ячеек, нет никаких оснований для такого циклического поведения, и здесь возможны самые различные виды долгосрочного поведения.

Системы подобного вида называются клеточными автоматами. Поскольку правила поведения таких систем точно определены, с ними легко проводить математические рассуждения. Для простоты обозначений рассмотрим целочисленные точки не на плоскости, а на прямой. Соответственно состояние системы — это последовательность, каждый элемент которой имеет одно из конечного числа значений (состояний). Если состояния задаются числами $0, \dots, N-1$, тогда мы можем обозначить пространство этих последовательностей через Ω_N . Развитие всех организмов подчиняется одному и тому же правилу. Это правило задается функцией $f: \{0, \dots, N-1\}^{2n+1} \rightarrow \{0, \dots, N-1\}$, отображающей последовательности состояний $(0, \dots, N-1)$ длиной $2n+1$ в одно из состоя-

ний. Входными данными служат состояния всех соседей некоторого организма, находящихся от него на расстоянии не больше чем n в любом направлении, а значением функции является будущее состояние этого индивидуального организма. Итак, каждый шаг эволюции системы задается таким отображением $\Phi: \Omega_N \rightarrow \Omega_N$, что $(\Phi(\omega))_i = f(\omega_{i-n}, \dots, \omega_{i+n})$. В качестве примера возьмем $N = n = 1$ и $f(x_{-1}, x_0, x_1) = x_1$. Это означает, что каждый организм следует поведению своего правого соседа (сегодняшнее значение x_1 есть завтрашнее значение x_0). Можно было бы назвать этот пример «волной», поскольку с какого бы начального состояния мы ни начали, оно неуклонно будет перемещаться влево.

Мы здесь привели общее описание клеточных автоматов, применение которых простирается гораздо дальше, чем игра «жизнь». То же самое математическое понятие допускает совершенно другую интерпретацию. Если каждую из последовательностей понимать как поток данных, то отображение Φ попросту преобразовывает эти данные, иными словами — кодирует. Этот класс кодов известен как *скользящие блочные коды*. Такие коды подходят для кодирования и декодирования потоков данных в реальном времени. Для нас важно, что это преобразование действует на хорошем пространстве, причем оно может быть проведено повторно, т. е. мы имеем дело с динамической системой. Общий класс динамических систем, состояния которых задаются последовательностями (или массивами), называется *символической динамикой*, и некоторые из наиболее полезных для нас моделей имеют именно такой вид. Наиболее часто мы будем использовать систему типа «волна», которую будем называть (левым) сдвигом. Класс скользящих блочных кодов играет важную роль в динамике, хотя они и известны под иным названием (сопряжения).

Символическая динамика вводится в п. 7.3.4 и изучается в п. 7.3.7. С ее помощью можно получить богатый набор примеров, которые являются простыми по описанию, но вместе с тем несут в себе удивительное многообразие сложных динамических явлений.

Упражнения

1.3.12. Докажите, что при бинарном поиске корня последовательности левых и правых конечных точек сходятся и имеют один и тот же предел.

1.3.13. Рассмотрим процесс бинарного поиска корня при $a = 0$, $b = 1$ и предположим, что этот процесс продолжается бесконечно долго. Закодируйте последовательность шагов бесконечными наборами из нулей и единиц, выбирая 0, когда реализуется случай 1, и 1 — когда реализуется случай 2. Докажите, что полученная таким образом строка из нулей и единиц представляет собой двоичное представление решения, найденного при помощи бинарного поиска.

1.3.14. В предшествующем упражнении предположим, что алгоритм останавливается на каком-то шаге. Как конечная строка из нулей и единиц связана с двоичным представлением корня?

1.3.15. Решите уравнение $\cos x = x$ методом Ньютона с начальным приближением $x_0 = 3/4$.

1.3.16. Найдите приближенное значение $\sqrt{5}$ методом Ньютона, выбирая 2 в качестве начального приближения, с наибольшей возможной для вашего калькулятора точностью.

1.3.17. Используйте метод Ньютона для решения уравнения $\sin x = 0$ с начальным приближением 1 и отметьте закономерность, которой подчиняется размер абсолютной погрешности.

1.3.18. Попробуйте решить уравнение $\sqrt[3]{x} = 0$ методом Ньютона, *не* выбирая 0 в качестве начального приближения.

1.3.19. Для греческого метода арифметического/гармонического средних выразите последовательные средние арифметические как итерации некоторой функции, т. е. укажите рекуррентную формулу для первых координат.

1.3.20. Вычисление квадратного корня из числа z можно осуществлять разными способами. Сравните греческий метод арифметического/гармонического средних с методом Ньютона, взяв 1 в качестве начального приближения.

1.3.21. Найдите наименьшую такую степень числа 2, что среди первых цифр всех предшествующих ей степеней (включая ее саму) «7» встречается чаще, чем «8».

1.3.22. Рассмотрим последовательность $(a_n)_{n \in \mathbb{N}}$, определенную последними двумя цифрами степеней числа 2. Докажите, что $a_n + a_{n+10} = 100$ для всех $n \geq 2$.

1.3.23. Докажите, что последовательность, определенная последними тремя цифрами степеней числа 2 (начиная с 008), является периодической с периодом 100.

1.3.24. Рассмотрим последовательность $(a_n)_{n \in \mathbb{N}}$, определенную последними тремя цифрами степеней числа 2. Докажите, что $a_n + a_{n+50} = 1000$ для всех $n \geq 3$.

Часть I

Динамические системы: от простого к сложному

Динамика предоставляет понятия и методы, используемые для понимания и описания сложного долгосрочного поведения систем, которые развиваются во времени. Мы сможем получить лучшее представление об этих понятиях и методах и оценить их роль, если будем последовательно переходить от рассмотрения систем с простым поведением к изучению более сложных систем. При этом как примеры, так и используемый аппарат будут приобретать все более сложный характер.

Соответственно наш курс начинается с изучения простейшего динамического поведения (универсальное устойчивое состояние), которое также оказывается фундаментальным инструментом для исследования более сложных примеров. Об этом подробнее говорится в гл. 9. В гл. 2 также рассматриваются более сложные системы, которые могут иметь несколько разных устойчивых состояний. Глава 3 посвящена линейным отображениям, в которых возможно еще более сложное поведение, но которые, тем не менее, поддаются изучению. Позднее мы будем их использовать как важный инструмент в изучении нелинейных систем.

Системы со сложным поведением впервые появляются в гл. 4 (и, в случае еще более высоких размерностей, в гл. 5). Мы рассматриваем орбиты, долгосрочное поведение которых таково, что они подходят близко к первоначальному состоянию, но не достигают его (явление рекуррентности), а также изучаем распределение в пространстве состояний. Однако потенциальная сложность поведения таких систем уменьшается из-за регулярности асимптотики отдельных орбит и схожести поведения разных орбит. Это наблюдение подтверждается с помощью статистического анализа, который вводится в той же главе. Механические системы (см. гл. 6) представляют собой естественный пример ситуации, когда достаточно сложное поведение появляется неустранимым образом, хотя их внутренняя структура часто ограничивает эту сложность до уровня двух предыдущих глав.

Системы самого высокого уровня сложности рассматриваются в гл. 7. Здесь мы сталкиваемся с ситуацией, когда сложное рекуррентное поведение существенно нерегулярно для отдельных орбит, асимптотическое поведение разных орбит принципиально различно и они сложным образом переплетены друг с другом. В этой и следующей главах вводятся понятия, необходимые для изучения систем такого уровня сложности, включая основы теории статистического поведения.

Главу 9, хотя она и относится ко второй обзорной части книги, можно, тем не менее, считать органичным продолжением курса. Доказательства в ней написаны на то же уровне строгости, что и в первой части книги.

СИСТЕМЫ С УСТОЙЧИВЫМ АСИМПТОТИЧЕСКИМ ПОВЕДЕНИЕМ

В этой главе приводится материал, необходимый для дальнейшего изложения большинства последующих глав. С одной стороны, здесь приводятся простейшие примеры динамического поведения с намеками на то, каким образом может возникать более сложное поведение. С другой стороны, мы вводим некоторые основные понятия и средства анализа, которыми будем часто пользоваться в дальнейшем. Имеются два типа динамических систем, которые мы рассматриваем здесь как «простые». Во-первых, это линейные отображения, простота которых заключена в возможности их разложения на отдельные компоненты и изучения каждой из них отдельно. Второй тип — это сжимающие отображения, которые просты, поскольку итерации всех точек сходятся к единственной точке. Здесь мы обсудим линейные отображения довольно кратко и уделим наибольшее внимание начальным аспектам их применения при изучении нелинейных динамических систем. Систематически линейные отображения изучаются в гл. 3. В настоящей главе рассматривается также ряд свойств сжимающих отображений, которые будут использоваться далее. С приложениями сжимающих отображений мы будем встречаться на всем протяжении книги, а глава 9 посвящена таким приложениям почти целиком.

§ 2.1. Линейные отображения и линеаризация

2.1.1. Линейные отображения на прямой. Простейшая модель популяции с дискретным временем $x_{i+1} = f(x_i) = kx_i$ (при $k > 0$), введенная в п. 1.2.9.1, имеет простую динамику: при любом начальном значении $x_0 \neq 0$ последовательность $(x_i)_{i \in \mathbb{N}}$ расходится, если $k > 1$, и стремится к 0, если $k < 1$. Простота этой модели еще и в том, что асимптотическое поведение не зависит от начальных условий: умножая x_0 на число a , мы умножаем тем самым и все значения x_i на то же самое число. При этом допустимые типы асимптотического поведения весьма просты.

В случае $k \neq 1$ картина почти не меняется при замене $f(x) = kx$ на $g(x) := kx + b$. Действительно, замена переменных $y = x - \frac{b}{1-k}$ приводит к соотношению $y_{i+1} = ky_i$. Таким образом, мы полностью описали возможности динамики для линейных систем (за исключением случая $f(x) = kx$ при $k = \pm 1$).

Линейные отображения в более высоких размерностях приводят к более сложному поведению индивидуальных орбит. Однако одна существенная черта простого поведения сохраняется. Хотя не все орбиты имеют одно и то же долгосрочное поведение, знание динамики лишь небольшого числа орбит оказывается достаточным для получения информации о динамике всех орбит. Данный вопрос изучается систематически в гл. 3.

2.1.2. Линеаризация. Некоторые важные динамические системы, изучаемые в этой книге, можно непосредственно получить из линейных, но наиболее интересные динамические системы не линейны. Тем не менее, знание именно линейных динамических систем может оказаться полезным для изучения нелинейных. Ключевым элементом в этом случае является дифференцируемость, или, иначе говоря, существование хорошего линейного приближения вблизи произвольной заданной точки.

Дифференцируемость и линейная аппроксимация обсуждаются в п. 2.2.4.1 и приложении А.2. Основное свойство дифференцируемости состоит в том, что она гарантирует хорошее линейное приближение отображения вблизи произвольной заданной точки. В качестве простого примера можно привести аппроксимацию функции $f(x) := \sqrt{x}$ вблизи точки $x = 16$ линейной функцией $L(x) = f(16) + f'(16)(x - 16) = 4 + \frac{1}{8}(x - 16)$, из которой, в частности, следует, что $\sqrt{17} \approx 4\frac{1}{8}$. Это приближение соответствует первому шагу метода Ньютона при аппроксимации числа $\sqrt{17}$ (см. п. 1.3.2.2).

Линейное приближение такого рода может быть также полезным для изучения долгосрочного поведения систем, если орбиты нелинейного отображения остаются достаточно близко к точке, в которой берется приближение. Далее мы встретим немало примеров поведения подобного рода. Одним из таких примеров является следующий частный случай предложения 2.2.17.

Предложение 2.1.1. Пусть F — такое дифференцируемое отображение действительной прямой, что $F(b) = b$. Если все орбиты линеаризации отображения F в точке b стремятся к b , то все орбиты отображения F , которые начинаются достаточно близко к b , также стремятся к b .

Один из основных результатов анализа, связанных с линеаризацией, — это теорема о конечных приращениях А.2.3 (см. также лемму 2.2.12). Она используется много раз уже в следующем параграфе (например, в предложении 2.2.3 и п. 2.2.4.4). Линеаризация также играет важную роль при анализе сложного поведения в динамических системах (см., например, гл. 7 и 10).

Упражнения

2.1.1. Покажите, что для $k \neq 1$ замена переменной $y = x - \frac{b}{1-k}$ преобразует соотношение $x_{i+1} = f(x_i) = kx_i + b$ в соотношение $y_{i+1} = ky_i$.

2.1.2. Опишите возможные типы асимптотического поведения для отображений $f(x) = kx$ при $k = \pm 1$.

2.1.3. Опишите возможные типы асимптотического поведения для отображений $f(x) = kx + b$ при $k = \pm 1$ и $b \neq 0$.

§ 2.2. Сжимающие отображения в евклидовом пространстве

Ученые и инженеры по традиции предпочитали иметь дело с динамическими системами с устойчивым асимптотическим поведением. Идеалом является система, приходящая в стационарное состояние, возможно, после короткого «периода

релаксации». Можно привести многочисленные простые примеры такого явления из реальной жизни. Настольная лампа при включении приходит в состояние «включено» с постоянной интенсивностью света почти мгновенно (после очень короткого периода нагревания нити накала). Если нить не повреждена, она не мигает и не мерцает. Аналогичным образом, эскалаторы двигаются с установившейся постоянной скоростью. Радиоприемники при первом включении проходят через фазу довольно сложного переходного поведения в течение доли секунды, однако затем приходят в устойчивое состояние приема. Наш обзор динамических систем начинается с систем, которые демонстрируют такое простое поведение.

По аналогии с рассмотренными выше примерами с непрерывным временем из реальной жизни можно считать, что асимптотическое поведение динамической системы с дискретным временем имеет самый простой вид, если итерации любой данной точки сходятся к одной точке — стационарному состоянию. Имеется один важный общий класс примеров, где встречается этот вид поведения, а именно сжимающие отображения. Они рассматриваются здесь не только потому, что их простая динамика является естественной отправной точкой для обсуждения асимптотического поведения, но также и потому, что мы будем использовать их при решении многочисленных задач математического анализа и дифференциальных уравнений, а также при изучении динамических систем с более сложным поведением. Мы будем встречаться с подобными приложениями на протяжении всей книги, а гл. 9 посвящена в основном им.

Определим теперь сжимающие отображения и попутно разъясним использование понятий «отображение» и «функция».

2.2.1. Определения. Когда мы используем слово «отображение», мы обычно (но не всегда; см. например п. 2.2.4.1) подразумеваем, что область определения и область значений находятся в одном и том же пространстве, а часто даже более того — что область значений лежит в области определения, поскольку мы итерируем отображения, порождая таким образом динамическую систему. Кролики Фибоначчи, омары штата Мэн, явление филлотаксиса, популяции бабочек и процедуры извлечения корней представляют собой примеры таких динамических систем. Время дискретно, и законы природы (или применяемого алгоритма) превращаются в правило, которое определяет текущие данные через предыдущие, и последующее состояние системы через текущее состояние. Все это достигается применением одного отображения, в котором «закодированы» эти законы. Так что динамические системы с дискретным временем — это отображения пространства в себя. Отображения, которые мы будем рассматривать, почти всегда непрерывны.

С другой стороны, «функции» имеют числовые значения, даже если они определены на каком-нибудь экзотическом пространстве и не итерируемы. Однако иногда мы будем использовать слово «функция» для обозначения отображения вещественной прямой или ее части в себя. Имеется третья группа преобразований, которые используются для замены переменных. Они называются *заменами координат* или *сопряжениями* (а иногда также и просто отображениями). Одно отображение, которое мы всегда имеем в нашем распоряжении, — тождественное, мы обозначаем его Id . Оно определяется соотношением $\text{Id}(x) = x$.

Теперь мы определим сжимающие отображения в евклидовой метрике

$$d(x; y) := \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

Определение 2.2.1. Отображение f подмножества X евклидова пространства удовлетворяет условию Липшица с константой λ , если

$$d(f(x), f(y)) \leq \lambda d(x, y) \quad (2.2.1)$$

для всех $x, y \in X$. Если $\lambda < 1$, то f называется *сжимающим отображением*, *сжатием* или λ -*сжатием*. Если отображение f удовлетворяет условию Липшица, то мы определяем его константу Липшица как $\text{Lip}(f) := \sup_{x \neq y} d(f(x), f(y))/d(x, y)$.

Пример 2.2.2. Функция $f(x) = \sqrt{x}$ определяет сжатие на $[1, \infty)$. Чтобы доказать это, мы убедимся, что для $x \geq 1$ и $t \geq 0$ выполняется неравенство $\sqrt{x+t} \leq \sqrt{x} + (1/2)t$ (объясните, почему этого достаточно). Это легко следует из того, что

$$\left(\sqrt{x} + \frac{t}{2}\right)^2 = x + xt + \frac{t^2}{4} \geq x + xt \geq x + t.$$

2.2.2. Случай одной переменной. Приведем теперь легкий способ проверки (в терминах производной) того, является ли отображение сжатием.

Предложение 2.2.3. Предположим, что I — это отрезок или интервал (возможно, неограниченный) на вещественной прямой и $f: I \rightarrow \mathbb{R}$ — такая дифференцируемая функция, что $|f'(x)| \leq \lambda$ для всех $x \in I$. Тогда f удовлетворяет условию Липшица с постоянной λ .

Доказательство. По теореме о конечных приращениях (А.2.3) для любых двух точек $x, y \in I$ существует такая точка c между x и y , что

$$d(f(x), f(y)) = |f(x) - f(y)| = |f'(c)(x - y)| = |f'(c)|d(x, y) \leq \lambda d(x, y). \quad \square$$

Заметим, что мы не нуждаемся в какой либо информации относительно значений производной f' в концевых точках интервала I .

Пример 2.2.4. Этот критерий позволяет легко установить, что функция $f(x) = \sqrt{x}$ задает сжатие на $I = [1, \infty)$, поскольку $f'(x) = 1/2\sqrt{x} \leq 1/2$ для $x \geq 1$.

Следует отметить, что согласно упражнению 2.2.14 более слабое условие $|f'(x)| < 1$, вообще говоря, не позволяет получить неравенство (2.2.1). Иногда, однако, это возможно.

Предложение 2.2.5. Допустим, что I — отрезок и $f: I \rightarrow I$ такая непрерывно дифференцируемая функция, что $|f'(x)| < 1$ для всех $x \in I$. Тогда f — сжимающее отображение.

Доказательство. Максимум λ функции $|f'(x)|$ достигается в некоторой точке x_0 , поскольку функция f' непрерывна. Этот максимум меньше 1, потому что $|f'(x_0)| < 1$. $\square$

Для вещественной прямой такое доказательство провести нельзя, так как она не является замкнутой и ограниченной (см. также упражнение 2.2.13).

В качестве примера рекуррентной последовательности в анализе часто приводят последовательность вида $a_{n+1} = f(a_n)$, где константа a_0 и функция f заданы, причем $|f'| \leq \lambda < 1$. Это простой пример динамической системы. Для каждого начального значения a_0 последовательность однозначно определена соотношением $a_{n+1} = f(a_n)$. Если функция f к тому же обратима, тогда эта последовательность определена для всех $n \in \mathbb{Z}$.

Определение 2.2.6. Для отображения f и точки x последовательность $x, f(x), f(f(x)), \dots, f^n(x), \dots$ (если отображение f необратимо) или же последовательность $\dots, f^{-1}(x), x, f(x), \dots$ (если отображение f обратимо) называется *орбитой* точки x под действием f . *Неподвижная точка* — это такая точка x , что $f(x) = x$. Множество неподвижных точек обозначается $\text{Fix}(f)$. *Периодическая точка* — это такая точка x , что $f^n(x) = x$ для некоторого $n \in \mathbb{N}$, т.е. $x \in \text{Fix } f^n$. Такое число n называется *периодом* точки x . Наименьшее число n с таким свойством называется *наименьшим периодом* точки x .

Пример 2.2.7. Для отображения $f(x) = -x^3$ на $\mathbb{R}$ единственной неподвижной точкой является 0, а ± 1 — периодическая орбита, т.е. 1 и -1 — периодические точки с наименьшим периодом 2.

Причина сходимости многих последовательностей такого вида, которые возникают в курсах математического анализа, кроется в следующем важном факте.

Предложение 2.2.8 (принцип сжимающих отображений). Пусть $I \subset \mathbb{R}$ — отрезок, замкнутый луч или вся числовая прямая и $f: I \rightarrow I$ — λ -сжатие. Тогда f имеет единственную неподвижную точку x_0 и $|f^n(x) - x_0| \leq \lambda^n |x - x_0|$ для любого $x \in \mathbb{R}$, т.е. каждая орбита отображения f сходится к x_0 экспоненциально.

Доказательство. Итерируя неравенство $|f(x) - f(y)| \leq \lambda |x - y|$, получаем

$$|f^n(x) - f^n(y)| \leq \lambda^n |x - y| \quad (2.2.2)$$

для $x, y \in \mathbb{R}$ и $n \in \mathbb{N}$, поэтому с учетом неравенства треугольника для любой точки $x \in I$ и для любых $m \geq n$ можем записать

$$\begin{aligned} |f^m(x) - f^n(x)| &\leq \sum_{k=0}^{m-n-1} |f^{n+k+1}(x) - f^{n+k}(x)| \leq \\ &\leq \sum_{k=0}^{m-n-1} \lambda^{n+k} |f(x) - x| \leq \frac{\lambda^n}{1-\lambda} |f(x) - x|. \end{aligned} \quad (2.2.3)$$

Здесь мы использовали известную формулу для частичной суммы геометрической прогрессии

$$(1 - \lambda) \sum_{k=l}^{n-1} \lambda^k = \lambda^l + \lambda^{l+1} + \dots + \lambda^{n-1} - \lambda^{l+1} - \lambda^{l+2} - \dots - \lambda^n = (\lambda^l - \lambda^n).$$

Так как правая часть неравенства (2.2.3) становится сколь угодно малой с ростом n , из этого неравенства следует, что последовательность $(f^n(x))_{n \in \mathbb{N}}$ является последовательностью Коши. Таким образом, для любого $x \in I$ предел последовательности $f^n(x)$ при $n \rightarrow \infty$ существует, потому что последовательности Коши сходятся, и лежит в I , поскольку множество I замкнуто. Согласно (2.2.2) этот предел один и тот же для всех значений x . Обозначим его через x_0 и покажем, что x_0 — неподвижная точка отображения f . Если $x \in I$ и $n \in \mathbb{N}$, тогда

$$\begin{aligned} |x_0 - f(x_0)| &\leq |x_0 - f^n(x)| + |f^n(x) - f^{n+1}(x)| + |f^{n+1}(x) - f(x_0)| \leq \\ &\leq (1 + \lambda)|x_0 - f^n(x)| + \lambda^n |x - f(x)|. \end{aligned} \quad (2.2.4)$$

Так как $|x_0 - f^n(x)| \rightarrow 0$ и $\lambda^n \rightarrow 0$ при $n \rightarrow \infty$, мы получаем $f(x_0) = x_0$.

Если в неравенстве (2.2.2) положить $y = x_0$, то мы получим, что $|f^n(x) - x_0| \leq \lambda^n |x - x_0|$ для каждого $x \in \mathbb{R}$. $\square$

Пример 2.2.9. Рассматривая процесс размножения кроликов, Леонардо Пизанский, также известный под именем Фибоначчи, придумал модель, согласно которой число b_n пар кроликов в n -м месяце определяется рекуррентными соотношениями $b_0 = 1$, $b_1 = 2$, $b_n = b_{n-1} + b_{n-2}$ для $n \geq 2$ (см. п. 1.2.2). Предполагая, что рост этих чисел экспоненциален, попытаемся определить, насколько быстро они растут, т. е. найти предел отношения $a_n := b_{n+1}/b_n$ при $n \rightarrow \infty$. Для этого мы воспользуемся принципом сжимающих отображений. Заметим, что

$$a_{n+1} = \frac{b_{n+2}}{b_{n+1}} = \frac{b_{n+1} + b_n}{b_{n+1}} = \frac{1}{b_{n+1}/b_n} + 1 = \frac{1}{a_n} + 1,$$

т. е. $(a_n)_{n=1}^\infty$ — орбита точки 1 при отображении $g(x) := (1/x) + 1$. Так как $g(1) = 2$, можно рассматривать орбиту точки 2 под действием отображения g . Так как $g'(x) = -x^{-2}$, отображение g не является сжатием на $(0, \infty)$. Поэтому мы должны найти подходящий отрезок, который отображался бы в себя и на котором g было бы сжатием.

Поскольку $g' < 0$, функция g убывает на $(0, \infty)$. Из этого следует, что $g([3/2, 2]) \subset [3/2, 2]$, поскольку $3/2 < g(3/2) = 5/3 < 2$ и $g(2) = 3/2$. Кроме того, $|g'(x)| = 1/x^2 \leq 4/9 < 1$ на $[3/2, 2]$, так что отображение g является сжатием на отрезке $[3/2, 2]$. Согласно принципу сжимающих отображений орбита точки 2, а следовательно и точки 1, асимптотически сходится к единственной неподвижной точке x отображения g , лежащей на $[3/2, 2]$. Таким образом, $\lim_{n \rightarrow \infty} b_{n+1}/b_n = \lim_{n \rightarrow \infty} a_n$ существует. Чтобы найти этот предел, достаточно решить уравнение $x = g(x) = 1 + 1/x = (x + 1)/x$, которое эквивалентно уравнению $x^2 - x - 1 = 0$. Имеется только одно положительное решение: $x = (1 + \sqrt{5})/2$ (см. упражнение 1.2.3). Другой способ нахождения этого решения и явной формулы для чисел Фибоначчи будет приведен в п. 3.1.9.

2.2.3. Случай нескольких переменных. Сейчас мы покажем, что принцип сжимающих отображений остается также верным и в высших размерностях. При этом используется по сути то же самое доказательство, только абсолютные величины заменяются евклидовым расстоянием.

Предложение 2.2.10 (принцип сжимающих отображений). Допустим, что множество $X \subset \mathbb{R}^n$ замкнуто, $f: X \rightarrow X$ — λ -сжатие. Тогда отображение f имеет единственную неподвижную точку x_0 и $d(f^n(x), x_0) \leq \lambda^n d(x, x_0)$ для любого $x \in X$.

Доказательство. Итерируя неравенство $d(f(x), f(y)) \leq \lambda d(x, y)$, мы получаем

$$d(f^n(x), f^n(y)) \leq \lambda^n d(x, y) \quad (2.2.5)$$

для $x, y \in X$ и $n \in \mathbb{N}$. Таким образом, последовательность $(f^n(x))_{n \in \mathbb{N}}$ является последовательностью Коши, поскольку

$$\begin{aligned} d(f^m(x), f^n(x)) &\leq \sum_{k=0}^{m-n-1} d(f^{n+k+1}(x), f^{n+k}(x)) \leq \\ &\leq \sum_{k=0}^{m-n-1} \lambda^{n+k} d(f(x), x) \leq \frac{\lambda^n}{1-\lambda} d(f(x), x) \end{aligned} \quad (2.2.6)$$

для $m \geq n$ и $\lambda^n \rightarrow 0$ при $n \rightarrow \infty$. Следовательно, $\lim_{n \rightarrow \infty} f^n(x)$ существует (поскольку последовательности Коши в $\mathbb{R}^n$ сходятся) и лежит в X , так как множество X замкнуто (см. рис. 2.2.1). Согласно (2.2.5) этот предел один и тот же для всех точек x . Обозначим его через x_0 . Тогда

$$\begin{aligned} d(x_0, f(x_0)) &\leq d(x_0, f^n(x)) + d(f^n(x), f^{n+1}(x)) + d(f^{n+1}(x), f(x_0)) \leq \\ &\leq (1 + \lambda) d(x_0, f^n(x)) + \lambda^n d(x, f(x)) \end{aligned} \quad (2.2.7)$$

для $x \in X$ и $n \in \mathbb{N}$. Поскольку $d(x_0, f^n(x)) \xrightarrow{n \rightarrow \infty} 0$, мы имеем $f(x_0) = x_0$. $\square$

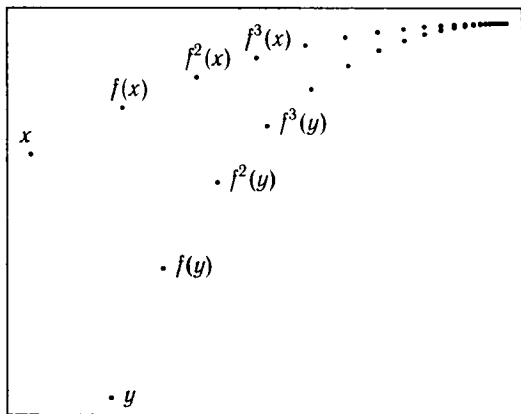


Рис. 2.2.1. Сходимость итераций

Переходя в (2.2.6) к пределу при $m \rightarrow \infty$, мы получаем, что $d(f^n(x), x_0) \leq \frac{\lambda^n}{1-\lambda} d(f(x), x)$. Это означает, что после n итераций мы можем с уверенностью говорить, что неподвижная точка находится в $\frac{\lambda^n}{1-\lambda} d(f(x), x)$ -шаре с центром в $f^n(x)$. Другими словами, при численном решении возможно делать строгие

Доказательство. Положим $v = g(b) - g(a)$, $\varphi(t) = \langle v, g(t) \rangle$. По теореме А.2.3 о конечных приращениях для функций одной переменной существует такое $t \in (a, b)$, что $\varphi(b) - \varphi(a) = \varphi'(t)(b - a)$, поэтому

$$(b - a) \|v\| \left\| \frac{d}{dt} g(t) \right\| \geq (b - a) \left\langle v, \frac{d}{dt} g(t) \right\rangle = \frac{d}{dt} \varphi(t)(b - a) = \varphi(b) - \varphi(a) = \\ = \langle v, g(b) \rangle - \langle v, g(a) \rangle = \langle v, v \rangle = \|v\|^2.$$

Чтобы завершить доказательство, достаточно разделить обе части этого неравенства на $\|v\|$. $\square$

3. **Выпуклость.** Далее нам понадобится понятие выпуклого множества.

Определение 2.2.13. *Выпуклое множество* в $\mathbb{R}^n$ — это такое множество S , что для любых $a, b \in S$ отрезок с концами в точках a и b полностью лежит в S . Множество S *строго* выпукло, если для любых точек a, b , принадлежащих замыканию множества S , весь отрезок с концами в точках a и b , кроме, возможно, одной или обеих концевых точек, содержится в S (см. рис. 2.2.2).

Рис. 2.2.2. Выпуклое, строго выпуклое и невыпуклое множества

Например, диск $\{(x, y) \in \mathbb{R}^2: x^2 + y^2 < 1\}$ — строго выпуклое множество. Открытая верхняя полуплоскость $\{(x, y) \in \mathbb{R}^2: y > 0\}$ — выпуклое, но не строго выпуклое множество. Множество $\{(r, \theta): 0 \leq r \leq 1 + (1/2) \sin \theta\}$ (в полярных координатах) невыпукло. Не является выпуклым и *круговое кольцо* $\{(x, y) \in \mathbb{R}^2: 1 < x^2 + y^2 < 2\}$.

4. **Теорема о первой производной.** Мы можем теперь сформулировать два варианта теоремы о первой производной, которая позволяет проверить, является ли отображение сжимающим, для случая нескольких переменных.

Теорема 2.2.14. *Если множество $S \subset \mathbb{R}^n$ выпукло и открыто, а отображение $f: S \rightarrow \mathbb{R}^m$ дифференцируемо, причем $\|Df(x)\| \leq M$ для всех $x \in S$, то $\|f(x) - f(y)\| \leq M\|x - y\|$ для $x, y \in S$.*

Доказательство. Отрезок, соединяющий x и y , задается параметрически формулой $c(t) = x + t(y - x)$, где $t \in [0, 1]$, и вследствие выпуклости множе-

ства C целиком в нем содержится. Положим $g(t) := F(c(t))$. Тогда в соответствии с правилом дифференцирования сложной функции имеем

$$\left\| \frac{d}{dt} g(t) \right\| = \left\| Df(c(t)) \frac{d}{dt} c(t) \right\| = \| Df(c(t))(y - x) \| \leq M \| y - x \|.$$

С учетом леммы 2.2.12 отсюда получаем, что $\|f(y) - f(x)\| = \|g(1) - g(0)\| \leq M \|y - x\|$. $\square$

Следствие 2.2.15. Если $C \subset \mathbb{R}^n$ — выпуклое открытое множество, $f: C \rightarrow C$ — отображение с непрерывными частными производными и $\|Df\| \leq \lambda < 1$ в каждой точке $x \in \mathbb{R}^n$, то отображение f является λ -сжатием.

Роль выпуклости в теореме 2.2.14, а стало быть, и в следствии 2.2.15, объяснена в упражнении 2.2.12. В частности, предположения, что любые две точки множества C могут быть связаны какой-либо хорошей кривой, недостаточно. Необходимо действительно использовать прямолинейный отрезок.

Мы не можем непосредственно применить предыдущее следствие для доказательства принципа сжимающих отображений, поскольку открытое множество может и не содержать пределы лежащих в нем последовательностей Коши. Поэтому мы приводим результат, который справедлив для замыкания такого множества. Он доказывается точно так же, как и теорема 2.2.14.

Теорема 2.2.16. Если $C \subset \mathbb{R}^n$ — открытое строго выпуклое множество, $\bar{C}$ — его замыкание, отображение $f: \bar{C} \rightarrow \mathbb{R}^n$ дифференцируемо на C и непрерывно на $\bar{C}$ и $\|Df\| \leq \lambda < 1$ в каждой точке $x \in C$, то f имеет единственную неподвижную точку $x_0 \in \bar{C}$ и $d(f^n(x), x_0) \leq \lambda^n d(x, x_0)$ для любого $x \in \bar{C}$.

Доказательство. Для $x, y \in \bar{C}$ параметризуем отрезок, соединяющий x и y , формулой $c(t) = x + t(y - x)$, где $t \in [0, 1]$, и положим $g(t) := f(c(t))$. Тогда $c((0, 1))$ содержится в C в силу строгой выпуклости и

$$\left\| \frac{d}{dt} g(t) \right\| = \left\| Df(c(t)) \frac{d}{dt} c(t) \right\| = \| Df(c(t))(y - x) \| \leq \lambda \| y - x \|.$$

В соответствии с леммой 2.2.12 из этого следует, что $\|f(y) - f(x)\| \leq \lambda \|y - x\|$. Таким образом, отображение f является λ -сжатием и имеет единственную неподвижную точку x_0 . Более того, мы видим, что $d(f^n(x), x_0) \leq \lambda^n d(x, x_0)$ для любого $x \in \bar{C}$. $\square$

2.2.5. Локальные сжатия. Теперь мы рассмотрим отображения, которые являются сжимающими не на всей их области определения, а только на некоторой ее части. Основным примером отображения, которое сжимает только локально, приводится в следующем предложении.

Предложение 2.2.17. Пусть f — непрерывно дифференцируемое отображение с неподвижной точкой x_0 и $\|Df_{x_0}\| < 1$. Тогда имеется такая замкнутая окрестность U точки x_0 , что $f(U) \subset U$ и f является сжатием на U .

Определение 2.2.18. Замкнутой окрестностью точки x мы называем замыкание открытого множества, содержащего x .

Доказательство. Так как отображение Df_x непрерывно зависит от x , найдется замкнутый шар $U = \bar{B}(x_0, \eta)$ с центром в x_0 , на котором $\|Df_x\| \leq \lambda < 1$ (см. упражнение 2.2.11). Если $x, y \in U$, то $d(f(x), f(y)) \leq \lambda d(x, y)$ в силу следствия 2.2.15, так что f — сжатие на U . Кроме того, если взять $y = x_0$, то для $x \in U$ мы имеем $d(f(x), x_0) = d(f(x), f(x_0)) \leq \lambda d(x, x_0) \leq \lambda \eta < \eta$ и, следовательно, $f(x) \in U$. $\square$

К сожалению, определение нормы $\|Df\|$ неудобно для практических вычислений. Однако, если мы готовы изменить норму или перейти к итерациям, мы можем довольно легко избежать его использования. Это будет сделано в следующей главе.

Предложение 2.2.19. Пусть f — непрерывно дифференцируемое отображение с такой неподвижной точкой x_0 , что все собственные значения отображения Df_{x_0} по абсолютной величине меньше 1. Тогда имеется такая замкнутая окрестность U точки x_0 , что $f(U) \subset U$ и f является сжатием на U в подходящей норме.

Доказательство. Как будет показано в следующей главе (предложение 3.3.3), из наших предположений относительно собственных значений следует, что можно выбрать норму (обозначим ее $\|\cdot\|'$), для которой $\|Df\|' < 1$. Теперь применим предложение 2.2.17. Другими словами, в качестве множества U можно выбрать достаточно малый замкнутый «шар» с центром в x_0 относительно нормы $\|\cdot\|'$. Этот шар является эллипсоидом в $\mathbb{R}^n$. $\square$

Ситуация, описанная в этом предложении, представляет большой интерес, так как она сохраняется при возмущениях.

2.2.6. Возмущения. Теперь посмотрим, что происходит с неподвижной точкой при возмущениях сжимающего отображения.

Предложение 2.2.20. Пусть f — непрерывно дифференцируемое отображение с неподвижной точкой x_0 и $\|Df_{x_0}\| < 1$. Тогда существует такая окрестность U точки x_0 , что любое отображение g , достаточно близкое к f , является сжатием на U . Точнее говоря, для любого $\epsilon > 0$ найдется такое $\delta > 0$, что любое отображение g , удовлетворяющее условиям $\|g(x) - f(x)\| \leq \delta$ и $\|Dg(x) - Df(x)\| \leq \delta$ на U , отображает U в U и является сжатием на U с единственной неподвижной точкой y_0 в $B(x_0, \epsilon)$.

Доказательство. В силу непрерывности линейного отображения Df имеется малый замкнутый шар $U = B(x_0, \eta)$ с центром в x_0 , на котором $\|Df_x\| \leq \lambda < 1$ (см. упражнение 2.2.11). Предположим, что $\eta\epsilon < 1$, и положим $\delta = \epsilon\eta(1 - \lambda)/2$. Тогда

$$\|Dg\| \leq \|Dg - Df\| + \|Df\| \leq \delta + \lambda \leq \lambda + (1 - \lambda)/2 = (1 + \lambda)/2 =: \mu < 1$$

на U , так что g является сжатием на U согласно следствию 2.2.15. Если $x \in U$, то $d(x, x_0) \leq \eta$ и

$$\begin{aligned} d(g(x), x_0) &\leq d(g(x), g(x_0)) + d(g(x_0), f(x_0)) + d(f(x_0), x_0) \leq \\ &\leq \mu d(x, x_0) + \delta + 0 \leq \mu\eta + \delta \leq \eta(1 + \lambda)/2 + \eta(1 - \lambda)/2 = \eta, \end{aligned} \quad (2.2.9)$$

поэтому $g(x) \in U$, откуда и следует, что $g(U) \subset U$. Наконец, поскольку $g^n(x_0) \rightarrow y_0$, мы имеем

$$d(x_0, y_0) \leq \sum_{n=0}^{\infty} d(g^n(x_0), g^{n+1}(x_0)) \leq d(g(x_0), x_0) \sum_{n=0}^{\infty} \mu^n \leq \frac{\delta}{1 - \mu} = \frac{\epsilon\eta(1 - \lambda)}{1 - \lambda},$$

а это выражение меньше чем ϵ . $\square$

Предыдущий результат, в частности, говорит нам, что неподвижная точка сжимающего отображения непрерывно зависит от сжатия. Это можно доказать и без условия дифференцируемости.

Предложение 2.2.21. Если отображение $f: \mathbb{R} \times (a, b) \rightarrow \mathbb{R}$ непрерывно и $f_y := f(\cdot, y)$ удовлетворяет условию $|f_y(x_1) - f_y(x_2)| \leq \lambda|x_1 - x_2|$ для всех $x_1, x_2 \in \mathbb{R}$ и всех $y \in (a, b)$, то неподвижная точка $g(y)$ для f_y непрерывно зависит от y (см. рис. 2.2.3).

Доказательство. Поскольку

$$|x - g(y)| \leq \sum_{i=0}^{\infty} |f_y^i(x) - f_y^{i+1}(x)| \leq \frac{1}{1 - \lambda} |x - f_y(x)|,$$

взяв $x = g(y') = f_{y'}(g(y'))$, мы получаем

$$|g(y') - g(y)| \leq \frac{1}{1 - \lambda} |f_{y'}(g(y')) - f_y(g(y'))|. \quad \square$$

Это утверждение применимо и к более общей ситуации (см. предложение 2.6.14), а теорема 9.2.4 содержит еще более сильный результат в этом направлении.

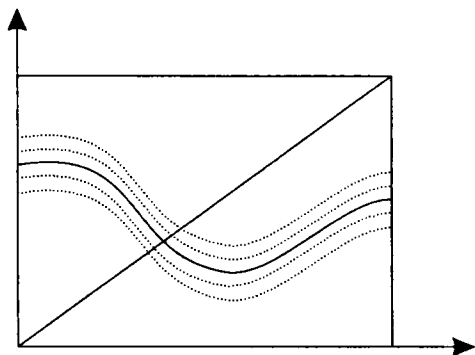


Рис. 2.2.3. Непрерывная зависимость неподвижной точки

2.2.7. Притягивающие неподвижные точки. К настоящему моменту мы столкнулись с двумя видами *устойчивости*: для заданного сжатия каждая индивидуальная орбита имеет устойчивое поведение, так что всякая близлежащая (а на самом деле вообще любая) орбита имеет в точности ту же самую асимптотику. Иначе говоря, небольшое возмущение начальной точки не оказывает никакого воздействия на асимптотическое поведение. Это устойчивость орбит по отношению к изменению начальных условий. С другой стороны, предложения 2.2.20 и 2.2.21 говорят нам, что сжатия являются устойчивыми как система, т. е. когда мы возмущаем само сжимающее отображение, качественное поведение всех орбит остается тем же самым и неподвижная точка изменяется лишь незначительно.

Сейчас настало время разъяснить, что мы подразумеваем под устойчивой неподвижной точкой. Как уже было сказано, мы хотим, чтобы каждая близлежащая орбита сходилась асимптотически к этой точке. Однако, как показывает рис. 2.2.4, на котором изображена полуустойчивая неподвижная точка, это не единственное требование. Отображение такого вида, как показано на рисунке, можно задать, например, формулой $f(x) = x + \frac{\sin^2 x}{4}$, если окружность представлена в виде $\mathbb{R}/\mathbb{Z}$ (см. п. 2.6.2). Чтобы гарантировать устойчивость, мы должны

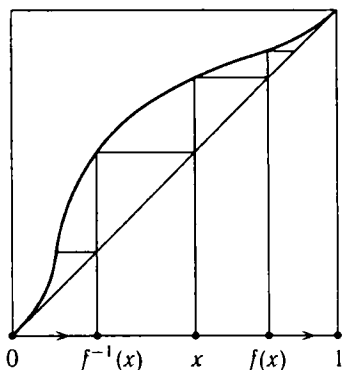
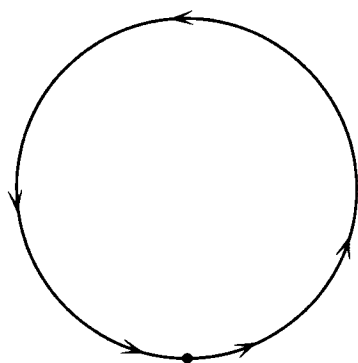


Рис. 2.2.4. Непритягивающая неподвижная точка

потребовать, чтобы близлежащие точки никогда не уходили далеко от неподвижной точки. Но, как показывает пример

$$f(x) = \begin{cases} -2x, & x \leq 0, \\ -x/4, & x > 0 \end{cases}$$

(или рис. 3.1.3), на некоторое время расстояние до неподвижной точки может увеличиться.

Определение 2.2.22. Неподвижная точка p называется *устойчивой по Ляпунову*, если для любого $\varepsilon > 0$ существует такое $\delta > 0$, что если некоторая точка находится на расстоянии не больше чем δ от p , то ее положительная полуорбита находится на расстоянии не больше чем ε от p . Точка p называется *асимптотически устойчивой* или *притягивающей неподвижной точкой*, если она устойчива по Ляпунову и имеется такое $a > 0$, что каждая точка, находящаяся на расстоянии не больше a от p , под действием итераций стремится к p .

2.2.8. Метод Ньютона. Весьма содержательное приложение метода линейных приближений к еще одной важной задаче — метод Ньютона нахождения корней уравнений (см. рис. 2.2.5), о котором говорится в п. 1.3.2.2. Нахождение точных корней — трудная, а часто и нерешаемая задача, и корни редко могут быть

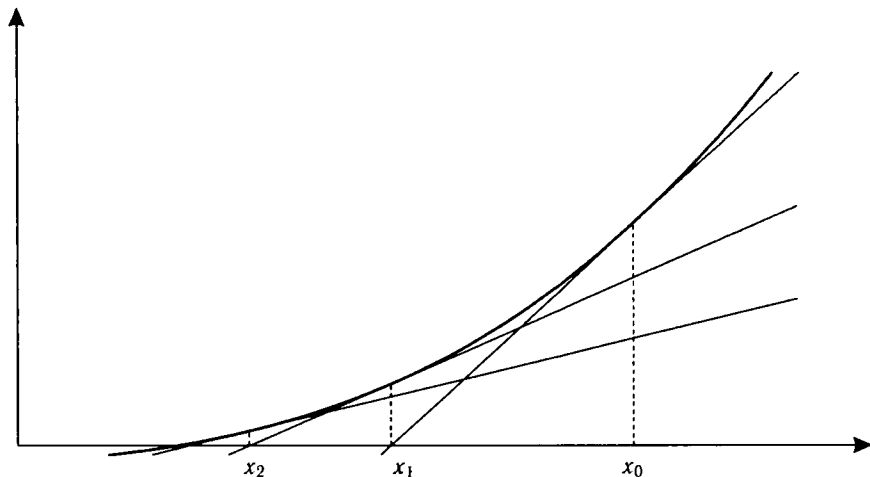


Рис. 2.2.5. Метод Ньютона

выражены в явном виде. Метод Ньютона нахождения корня работает достаточно хорошо и позволяет локализовать корень уравнения за небольшое вычислительное время при разумном начальном приближении. Чтобы понять, как этот метод работает, рассмотрим функцию f , заданную на действительной оси, и предположим, что имеется разумное начальное приближение x_0 для корня. Если график функции не пересекает ось x в точке x_0 , т. е. если $f(x_0) \neq 0$, мы должны улучшить наше приближение. Для этого проведем касательную к графику и найдем точку x_1 ,

в которой она пересекает ось x , из уравнения $f(x_0) + f'(x_0)(x_1 - x_0) = 0$. Таким образом, улучшенное приближение имеет вид

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Пример 2.2.23. Если мы начинаем с приближения $x_0 = 4$ для функции $x^2 - 17$, то улучшенная оценка корня имеет вид

$$x_1 = x_0 - \frac{x_0^2 - 17}{2x_0} = \frac{x_0}{2} + \frac{17}{2x_0} = \frac{33}{8}.$$

Еще один шаг приводит к приближению

$$x_2 = \frac{33}{16} + \frac{17 \cdot 8}{2 \cdot 33} = \frac{33^2 + 17 \cdot 64}{16 \cdot 33}.$$

Можно повторять итерации и получать более точные оценки x_3, x_4 и т. д., используя ту же самую формулу. С хорошим начальным приближением несколько шагов обычно позволяют получить довольно точное решение. (Действительно, x_2 уже отличается от корня меньше чем на 10^{-6} .)

Легко видеть, почему это так: мы несколько раз применяем отображение $F(x) := x - \frac{f(x)}{f'(x)}$, и искомая точка обладает следующим свойством.

Определение 2.2.24. Неподвижная точка x дифференцируемого отображения F называется *сверхпритягивающей*, если $F'(x) = 0$ (см. рис. 2.2.6).

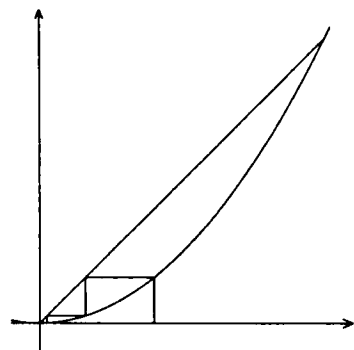


Рис. 2.2.6. Сверхпритягивающая неподвижная точка

Предложение 2.2.25. Если $|f'(x)| > \delta$ и $|f''(x)| < M$ в окрестности корня r , то r является *сверхпритягивающей неподвижной точкой* для $F(x) := x - \frac{f(x)}{f'(x)}$.

Доказательство. Наше утверждение следует из соотношений $F(r) = r$ и $F'(x) = f(x)f''(x)/(f'(x))^2$. $\square$

Замечание 2.2.26. Если первая производная мала, то точка пересечения касательной с осью x может находиться довольно далеко от x_0 . Предположение $|f''(x)| < M$ обязательно выполняется, если f'' — непрерывная функция.

В случае сверхпритягивающей неподвижной точки мы имеем квадратичную (т. е. суперэкспоненциальную) сходимость итераций, как и в случае неподвижной точки квадратичного отображения f_2 из § 2.5. Другими словами, при каждой итерации погрешность приблизительно возводится в квадрат.

Это рассуждение применимо, только если начальное приближение довольно хорошее. С неудачным начальным выбором итерации функции F могут иметь довольно неустойчивое поведение. Другими словами, функция F имеет притягивающую неподвижную точку, но тем не менее динамика этой функции может быть весьма сложной.

Аналог одного частного случая метода Ньютона нахождения корней был известен в древности.

Предложение 2.2.27. *Использование метода Ньютона с начальным значением $x_0 = 1$ для приближенного вычисления $\sqrt{2}$ приводит к тем же результатам, что и использование первых приближений греческого метода извлечения корня, приведенного в (1.3.1).*

Доказательство. С начальным предположением $x_0 = 1$ метод Ньютона приводит к рекуррентной формуле

$$x_0 = 1, \quad x_{n+1} = x_n - \frac{x_n^2 - 2}{2x_n} = \frac{x_n}{2} + \frac{2}{2x_n}.$$

Применение греческого метода начинается с приближения $(x_0, y_0) = (1, 2)$, и рекуррентная последовательность $(x_{n+1}, y_{n+1}) = f(x_n, y_n)$, определенная в (1.3.1), имеет свойство $y_n = 2/x_n$. Поэтому мы получаем

$$x_{n+1} = \frac{x_n + y_n}{2} = \frac{x_n}{2} + \frac{2}{2x_n}. \quad \square$$

2.2.9. Приложения принципа сжимающих отображений. Этот параграф в основном был посвящен принципу сжимающих отображений. Это один из наиболее важных фактов в анализе и теории динамических систем. Его приложения обширны и фундаментальны. Мы приводим многочисленные примеры применения этого принципа в нашем курсе динамики. Более того, некоторые результаты, которые лежат в основе теории, являются прямыми следствиями принципа сжимающих отображений. Глава 9 полностью посвящена этим приложениям. Хотя она и относится к обзору результатов, она тесно связана с курсом и написана на том же уровне строгости. В этой главе доказываются фундаментальные результаты анализа — теоремы об обратной функции и о неявной функции (теоремы 9.2.2 и 9.2.3). Как уже отмечалось, неподвижная точка сжимающего отображения остается устойчивой при возмущениях, и в гл. 9 приводится наиболее общее условие, которому должна удовлетворять точка, для того чтобы она была устойчивой (предложение 9.3.1). Кроме того, теория существования и единственности для дифференциальных уравнений, на которой основана, можно сказать, половина динамики, выводится из принципа сжимающих отображений (теорема 9.4.3). Один из центральных результатов теории динамических систем (обсуждаемый в гл. 7) — теорема об устойчивом многообразии (теорема 9.5.2, см. также комментарии в конце § 10.1). Этот результат также основан на принципе сжимающих отображений.

Упражнения

2.2.1. Покажите, что, вводя произвольное число на калькуляторе и многократно нажимая кнопку «sin», мы получим последовательность, которая сходится к нулю. Докажите, что сходимость не экспоненциальная, если использовать радианы, и экспоненциальная, если использовать градусы. В последнем случае выясните, сколько итераций необходимо, чтобы получить число, более чем в 10^{-10} меньше, чем исходное значение.

2.2.2. Докажите, что если набрать на калькуляторе число, большее чем 1, и многократно нажимать кнопку квадратного корня, то мы получим убывающую последовательность, и найдите ее предел. Если калькулятор хранит в памяти k двоичных разрядов, найдите приближенное число итераций, необходимое, чтобы достигнуть этого предела точности.

2.2.3. Выполните предыдущее упражнение для начальных значений из полуинтервала $(0, 1]$.

2.2.4. Покажите, что функция x^2 задает λ -сжатие на $[-\lambda/2, \lambda/2]$.

2.2.5. Это упражнение представляет собой вариацию задачи Фибоначчи о популяции кроликов, учитывающую фактор смертности.

Популяция полярных леммингов развивается согласно следующим правилам. Имеется равное количество мужских и женских особей. Каждый лемминг живет в течение двух лет и умирает на третью зиму жизни. Каждое лето каждая женская особь производит на свет четырех детенышей. В первое лето имеется одна пара однолетних леммингов. Пусть x_n обозначает общее количество леммингов в n -м году. Используя принцип сжимающих отображений, покажите, что последовательность x_{n+1}/x_n сходится к пределу $\omega > 1$. Вычислите ω .

2.2.6. Допустим, что x — такая неподвижная точка отображения f , заданного на действительной оси, что $|f'(x)| = 1$ и $f''(x) \neq 0$. Покажите, что в сколь угодно малой окрестности x найдется такая точка y , что последовательность ее итераций не сходится к x .

2.2.7. Какие из следующих множеств выпуклы: $\{(x, y) \in \mathbb{R}^2: xy > 1\}$, $\{(x, y) \in \mathbb{R}^2: xy < 1\}$, $\{(x, y) \in \mathbb{R}^2: x + y > 1\}$, $\{(x, y) \in \mathbb{R}^2: x > y^2\}$?

2.2.8. Докажите, что норма матрицы, определенная в (2.2.8), является нормой в смысле определения A.1.29.

2.2.9. Покажите, что $\|A\| \leq \sqrt{\sum_{i,j} a_{ij}^2}$ для любой $(n \times n)$ -матрицы $A = (a_{ij})_{1 \leq i,j \leq n}$.

2.2.10. Покажите, что $\|A\| \geq |\det A|^{1/n}$ для любой $(n \times n)$ -матрицы $A = (a_{ij})_{1 \leq i,j \leq n}$.

2.2.11. Докажите, что норма матрицы является непрерывной функцией ее элементов.

Задачи для углубленного изучения

2.2.12. Приведите пример такого открытого связного подмножества U плоскости $\mathbb{R}^2$ и такого непрерывно дифференцируемого отображения $f: U \rightarrow U$, что $\|Df_x\| < \lambda < 1$ для всех $x \in U$, однако при этом f не является сжатием. (Заметьте, что такое множество не может быть выпукло.)

2.2.13. Предположим, что I — отрезок и $f: I \rightarrow I$ — такое отображение, что $d(f(x), f(y)) < d(x, y)$ для всех $x \neq y$ (это более слабое предположение, чем условие принципа сжимающих отображений). Докажите, что отображение f имеет единственную неподвижную точку $x_0 \in I$ и что $\lim_{n \rightarrow \infty} f^n(x) = x_0$ для любого $x \in I$.

2.2.14. Покажите, что утверждение предыдущего упражнения неверно для $I = \mathbb{R}$, построив такое отображение $f: \mathbb{R} \rightarrow \mathbb{R}$, что $d(f(x), f(y)) < d(x, y)$ для $x \neq y$, f не имеет ни одной неподвижной точки и $d(f^n(x), f^n(y))$ не сходится к нулю для некоторых x, y .

§ 2.3. Неубывающие отображения интервала и бифуркации

Теперь мы рассмотрим отображения, которые могут иметь несколько неподвижных точек. Хотя поведение таких отображений не намного сложнее, чем для сжимающих отображений, оно может радикально изменяться при возмущениях.

2.3.1. Неубывающие отображения интервала или отрезка. Перейдем теперь к изучению ситуаций, в которых динамика подобна динамике сжатий, но скорость сходимости к неподвижной точке не обязательно экспоненциальна. Анализ таких случаев весьма полезен, поскольку мы здесь сталкиваемся с важным методом динамики малых размерностей, а именно с систематическим использованием теоремы о промежуточном значении.

Определение 2.3.1. Пусть $I \subset \mathbb{R}$ — некоторый интервал или отрезок. Мы говорим, что функция $f: I \rightarrow \mathbb{R}$ *возрастает*, если $x > y \implies f(x) > f(y)$, и *убывает*, если $x > y \implies f(x) < f(y)$. Мы говорим, что функция f *неубывающая*, если $x \geq y \implies f(x) \geq f(y)$, и *невозрастающая*, если $x \geq y \implies f(x) \leq f(y)$.

Следующий простой пример весьма полезен в теории неубывающих отображений.

Лемма 2.3.2. Если $I = [\alpha, \beta] \subset \mathbb{R}$ — отрезок и $f: I \rightarrow I$ неубывающее отображение без неподвижных точек в (α, β) , то одна концевая точка отрезка I неподвижна и все орбиты сходятся к ней, кроме другой концевой точки, если последняя также неподвижна. Если отображение f обратимо, то обе концевые точки неподвижны и все орбиты точек из интервала (α, β) при положительных итерациях асимптотически сходятся к одной из концевых точек и при отрицательных — к другой.

Замечание 2.3.3 (лестница Лемерея). Для любого конкретного примера такой функции утверждение леммы вполне очевидно, если начертить график. Для этого полезно использовать «лестницу Лемерея», или «графические вычисления». Алгоритм наших действий таков: чтобы определить орбиту точки x , отметим эту точку на горизонтальной оси и проведем через нее вертикальную прямую до пересечения с графиком в точке $(x, f(x))$. Абсцисса следующей точки в орбите — это $f(x)$, так как горизонтальная прямая, проведенная через точку $(x, f(x))$, пересекает диагональ $x = y$ в точке $(f(x), f(x))$. Теперь проведем через точку $f(x)$ на горизонтальной оси вертикальную прямую до пересечения с графиком, а затем через полученную точку — горизонтальную прямую до пересечения с диагональю и т. д. На рис. 2.3.1 этот процесс проиллюстрирован для несколько более сложной функции, чем те, что рассматриваются в лемме. Но даже для такой функции картина весьма проста.

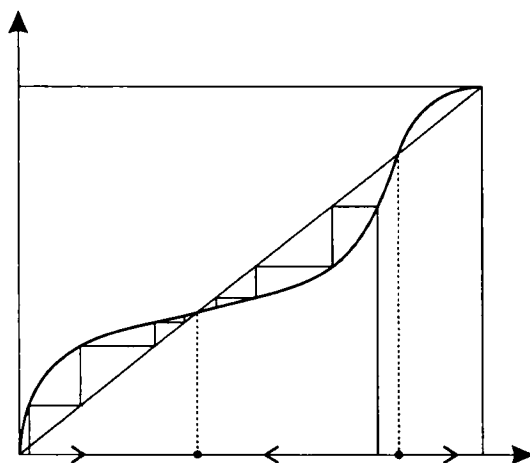


Рис. 2.3.1. Динамика возрастающего отображения отрезка

Идея, лежащая в основе леммы 2.3.2, состоит в том, что, так как график не пересекает диагональ над интервалом (α, β) (нет неподвижных точек), он располагается либо полностью выше, либо полностью ниже диагонали. Если график расположен выше диагонали, то $f(x) > x$, так что каждая орбита представляет собой возрастающую последовательность и точки постоянно перемещаются вправо. Так как любая орбита ограничена сверху, она должна сходиться, и затем нам, таким образом, необходимо показать, что предел всегда равен β . Если отображение f обратимо, то при итерациях обратного отображения точки должны двигаться в противоположном направлении (к α). Схематические рисунки помогают наметить идею доказательства, однако они не заменяют строгого доказательства, которое приводится ниже.

Доказательство. Так как $f(I) \subset I$, мы имеем $f(\alpha) \geq \alpha$ и $f(\beta) \leq \beta$, поэтому $(f - \text{Id})(\alpha) \geq 0$ и $(f - \text{Id})(\beta) \leq 0$, где Id — тождественное отображение. С другой стороны, по теореме о промежуточном значении непрерывная функция $f - \text{Id}$ не может менять знак на I , потому что она не обращается в нуль на (α, β) по предположению. Таким образом, либо $f(\alpha) = \alpha$ (если $f(x) < x$ на (α, β)), либо $f(\beta) = \beta$ (если $f(x) > x$ на (α, β)), либо реализуются оба случая. Для определенности предположим, что $f(x) > x$ на (α, β) и, следовательно, β является неподвижной точкой. Тогда для любого $x \in (\alpha, \beta)$ последовательность $x_n := f^n(x)$ является возрастающей и ограничена сверху числом β , следовательно, она сходится к некоторой точке $x_0 \in (\alpha, \beta]$. Но тогда в силу непрерывности

$$f(x_0) = f(\lim_{n \rightarrow \infty} x_n) = \lim_{n \rightarrow \infty} f(x_n) = \lim_{n \rightarrow \infty} x_{n+1} = x_0, \quad (2.3.1)$$

так что $x_0 \in \text{Fix } f \cap (\alpha, \beta] = \{\beta\}$. В случае если $f(y) < y$ на (α, β) , мы аналогичным образом заключаем, что $f^n(x) \rightarrow \alpha$ при $n \rightarrow \infty$ для любого $x \in (\alpha, \beta)$.

Если отображение f обратимо, то из неравенства $z := f(y) > y$ следует, что $f^{-1}(z) = y < f(y) = z$, так что если $f(x) > x$ на (α, β) и, следовательно, $f^n(x) \rightarrow \beta$

при $n \rightarrow \infty$, то, как видно из приведенных выше рассуждений, $f^{-1}(x) < x$ на (α, β) и $f^{-n}(x) \rightarrow \alpha$ при $n \rightarrow \infty$, поэтому положительные и отрицательные итерации любой точки $x \in (\alpha, \beta)$ сходятся к противоположным концам отрезка $[\alpha, \beta]$. $\square$

Это свойство приводит к следующему полезному определению.

Определение 2.3.4. Если $f: X \rightarrow X$ — обратимое отображение и $x \in X$ — такая точка, что $\lim_{n \rightarrow \infty} f^{-n}(x) = a$ и $\lim_{n \rightarrow \infty} f^n(x) = b$, то точка x называется *гетероклинической* для a и b . Если же $a = b$, то точка x называется *гомоклинической* для a .

Мы пока не встречались с примерами гомоклинических точек. Для неубывающих отображений интервала или отрезка таких точек быть не может, потому что все орбиты монотонны. Рис. 2.2.4 показывает, как гомоклинические точки могут возникать на окружности.

После ситуации, рассмотренной в лемме 2.3.2, перейдем к изучению еще одного простого типа асимптотического поведения, который наблюдается, если каждая орбита сходится к неподвижной точке, но различные орбиты могут сходиться к различным неподвижным точкам. Это происходит в случае возрастающих функций вещественной переменной, рассматриваемых как отображения.

Предложение 2.3.5. Если $I \subset \mathbb{R}$ — отрезок и $f: I \rightarrow I$ — неубывающее непрерывное отображение, то любая точка $x \in I$ или неподвижна, или же ее итерации асимптотически стремятся к неподвижной точке. Если отображение f возрастает (следовательно, обратимо), то любая точка $x \in I$ или неподвижна, или является гетероклинической к соседним неподвижным точкам.

Доказательство. Рис. 2.3.1 помогает понять идею доказательства. Заметим, что направление движения определяется знаком отображения $f - \text{Id}$: если $(f - \text{Id})(x) < 0$, то $f(x) - x < 0$, так что $f(x) < x$ и точка x перемещается влево, а если $(f - \text{Id})(x) > 0$, то точка перемещается вправо.

Сначала мы покажем, что неподвижные точки существуют. Положим $I = [a, b]$ и рассмотрим непрерывное отображение $f - \text{Id}: I \rightarrow \mathbb{R}$, заданное функцией $f(x) - x$. Так как $f(I) \subset I$, мы имеем $f(a) \geq a$ и $f(b) \leq b$, следовательно, $(f - \text{Id})(a) \geq 0$ и $(f - \text{Id})(b) \leq 0$, так что $(f - \text{Id})(x) = 0$ для некоторого $x \in I$ по теореме о промежуточном значении. Но тогда $f(x) = x$ и x — неподвижная точка.

Множество неподвижных точек $\text{Fix } f$ замкнуто, поскольку является множеством нулей непрерывной функции $f - \text{Id}$. Если $\text{Fix } f = I$, то каждая точка неподвижна.

В противном случае $I \setminus \text{Fix } f$ является непустым открытым множеством и, следовательно, представляет собой объединение непересекающихся интервалов. Рассмотрим интервал (α, β) из этого объединения. Либо $\alpha, \beta \in \text{Fix } f$, либо одна из точек α, β является концом отрезка I , так что в любом случае $f(\alpha) \geq \alpha$ и $f(\beta) \leq \beta$. Если $y \in (\alpha, \beta)$, то $\alpha \leq f(\alpha) \leq f(y) \leq f(\beta) \leq \beta$, поскольку отображение f неубывающее. Это показывает, что $f([\alpha, \beta]) \subset [\alpha, \beta]$. Таким образом, все орбиты в $[\alpha, \beta]$ асимптотически стремятся к неподвижной точке в силу леммы 2.3.2.

Если отображение f возрастает, а следовательно, обратимо, то, как следует из леммы 2.3.2, все точки интервала (α, β) являются гетероклиническими к концам отрезка $[\alpha, \beta]$. $\square$

На рис. 2.3.1 левая неподвижная точка является притягивающей (определение 2.2.22). Другая неподвижная точка относится к прямо противоположному типу, который столь же распространен, как и притягивающие неподвижные точки, и также заслуживает отдельного названия.

Определение 2.3.6. Неподвижная точка x называется *отталкивающей* (или *репеллером*), если для любого $\epsilon > 0$ и любой точки y , расположенной на расстоянии меньше чем ϵ от x , существует такое $n \in \mathbb{N}$, что положительная полуорбита точки $f^n(y)$ не имеет никаких точек на расстоянии меньше чем ϵ от x .

2.3.2. Бифуркации. Для сжимающих отображений устойчивость индивидуальных орбит сосуществует с устойчивостью поведения всей системы при возмущениях. Это не обязательно так в тех случаях, где динамика различных точек может иметь различные качественные особенности. Здесь имеется несколько неподвижных точек и орбиты могут притягиваться к одной из них. Число неподвижных точек не определено заранее. Это делает интересным изучение изменения качественной картины при изменении отображения. Можно пытаться искать переходы между различными типами поведения в семействах неубывающих отображений, зависящих от параметра, а также искать значения параметра, при которых такие изменения происходят.

Подобные переходы называются *бифуркациями*. Примером является семейство отображений, изображенных на рис. 2.3.2. Сначала имеются две неподвижные точки (левый рисунок), одна притягивающая и одна отталкивающая. При



Рис. 2.3.2. Простейшая бифуркация (седлоузел)

определенном значении параметра, которое называется *бифуркационным параметром*, эти точки сливаются в одну (средний рисунок), которая немедленно исчезает (правый рисунок) для больших значений параметра. Конечно, вне изображенной части графика должна быть другая неподвижная точка, поскольку, как мы видели в последнем доказательстве, по теореме о промежуточном значении всегда должна существовать по крайней мере одна неподвижная точка. Такая ситуация иногда называется *бифуркацией седлоузла*.

Эта терминология происходит из теории дифференциальных уравнений, где соответствующая бифуркация в более наглядном двумерном случае представляет собой слияние двух положений равновесия, *седла* (рис. 3.1.6) и *узла* (рис. 3.1.2), которые поглощают друг друга и исчезают.

Первоначально мы имеем одну устойчивую и одну неустойчивую неподвижные точки, и единственная неподвижная точка, имеющаяся для бифуркационного значения параметра, является полуустойчивой (как на рис. 2.2.4), т. е. другие точки приближаются к ней с одной стороны и отдаляются с другой. Для чуть больших значений параметра эта неподвижная точка исчезает, но это не оказывает влияния на поведение орбиты на не очень больших интервалах времени, поскольку для того, чтобы орбита прошла узкое «горлышко» между диагональю и графиком (большое количество ступеней на лестнице Лемерея), требуется большое время.

Упражнения

2.3.1. Докажите, что множество нулей непрерывной функции замкнуто.

2.3.2. Пусть $f: [0, 1] \rightarrow [0, 1]$ — *невозрастающее* непрерывное отображение. Каковы возможные периоды для периодических точек такого отображения?

2.3.3. Покажите, что используемое в доказательстве предложения 2.2.8 утверждение о том, что предел является неподвижной точкой, можно доказать, применяя рассуждения, подобные приведенным в (2.3.1) (без экспоненциальных оценок), вместо (2.2.4).

2.3.4. Докажите, что для любого замкнутого множества $E \subset \mathbb{R}$ существует такое непрерывное строго возрастающее отображение $f: \mathbb{R} \rightarrow \mathbb{R}$, что $\text{Fix}(f) = E$.

Задачи для углубленного изучения

2.3.5. Докажите, что для замкнутого множества $E \subset \mathbb{R}$ имеется такое непрерывное дифференцируемое строго возрастающее отображение $f: \mathbb{R} \rightarrow \mathbb{R}$, что $\text{Fix}(f) = E$.

2.3.6. Исследуя рис. 2.2.4, мы отметили, что для неубывающих отображений интервала или отрезка неподвижная точка, к которой все остальные точки асимптотически стремятся при положительных итерациях, должна быть аттрактором. Докажите, что это верно, даже если отображение не монотонно.

§ 2.4. Дифференциальные уравнения

В этом параграфе мы исследуем простую динамику отображений, с которыми мы сталкивались до настоящего момента, в контексте дифференциальных уравнений. Сначала мы рассмотрим дифференциальные уравнения на прямой, являющиеся аналогами неубывающих отображений интервала или отрезка, а затем детально обсудим логистическую модель роста с непрерывным временем. И в том, и в другом случаях нам встретятся примеры притягивающих неподвижных точек, определяемых точно так же, как для отображений. Однако имеется и другой аналог притягивающих неподвижных точек для отображений, который будет введен в п. 2.4.3.

2.4.1. Дифференциальные уравнения на прямой. Теперь мы докажем результат, аналогичный предложению 2.3.5, для непрерывного времени. Пусть имеется дифференциальное уравнение первого порядка $\dot{x} = f(x)$, где функция f удовлетворяет условию Липшица (см. определение 2.2.1). Рассмотрим множество нулей функции f , которые являются постоянными решениями (положениями равновесия).

Замечание 2.4.1. Множество нулей замкнуто, потому что функция f непрерывна. Поэтому дополнение этого множества открыто и может быть представлено как объединение непересекающихся интервалов.

Если рассматривать только один из этих интервалов, то следующий результат представляет собой полное описание нашей ситуации в точной аналогии с леммой 2.3.2.

Лемма 2.4.2. *Рассмотрим функцию f , удовлетворяющую условию Липшица, и предположим, что $f \neq 0$ на (a, b) и $f(a) = f(b) = 0$. Тогда для любого начального условия $x_0 \in I$ соответствующее решение уравнения $\dot{x} = f(x)$ монотонно. Это решение возрастает (и асимптотически стремится к b), если $f > 0$ на I , и убывает (и асимптотически стремится к a) в случае $f < 0$.*

Доказательство. Для определенности положим $f(x_0) > 0$ (другой случай вполне аналогичен). Легко видеть, что решение возрастает, пока оно находится в (a, b) . Необходимо показать, что оно не может выйти за пределы этого интервала.

Поскольку $\dot{x}(0) = f(x(0)) = f(x_0) > 0$, решение первоначально возрастает. Если оно когда-либо становится убывающим, то должен иметься максимум $x(t_0) = c$, откуда следует, что $f(c) = 0$ и поэтому $c = b$. Мы должны проверить, что такого быть не может, т. е. что при всех значениях t выполняется неравенство $x(t) \neq b$. Для этого имеются два пути, честный и легкий. Мы начнем с честного.

Мы можем записать решение дифференциального уравнения $\dot{x} = f(x)$ в виде $x(t) = x(0) + \int_0^t f(x(s)) ds$, поскольку дифференцирование обеих частей приводит к уравнению $\dot{x} = f(x)$. Для нашей задачи запишем $\frac{dx}{dt} = f(x)$ и по теореме об обратной функции получим $\frac{dt}{dx} = \frac{1}{f(x)}$, так что в интегральной форме уравнение можно переписать в виде $t(x) = \int_{x_0}^x \frac{1}{f(s)} ds$. Поскольку функция f удовлетворяет условию Липшица, мы имеем $f(s) = f(s) - f(b) \leq C(b - s)$ для некоторой константы C . Поэтому

$$t(x) = \int_{x_0}^x \frac{1}{f(s)} ds \geq \int_{x_0}^x \frac{1}{C(b-s)} ds.$$

Если $x = b$, то этот интеграл расходится, т. е. $t(x) = \infty$. Это показывает, что $x(t) < b$ для всех (конечных) значений t и, кроме того, что $x(t) \rightarrow b$ при $t \rightarrow \infty$.

Легкий путь состоит в использовании теоремы о существовании и единственности решений дифференциальных уравнений, которую мы докажем позже

(теорема 9.4.3). Поскольку функция $\bar{x}(t) = b$ для всех $t \in \mathbb{R}$ является решением, любое решение, которое в какой-либо момент времени принимает значение b , имеет вид $\bar{x}(t - t_0) = b$. Но поскольку начальное значение не равно b , наше решение не представимо в таком виде, следовательно, оно не может принимать значение b . $\square$

Таким образом, и в этой ситуации каждое решение имеет простую асимптоту: при $t \rightarrow \infty$ или $t \rightarrow -\infty$ решение или сходится к неподвижной точке, или расходится к бесконечности.

Используя лемму 2.4.2, теперь легко полностью описать качественную картину поведения любого дифференциального уравнения с одной переменной. Отметим на вещественной прямой точки, в которых $f = 0$. Они являются неподвижными точками. На каждом из интервалов дополнения определяем знак функции f . Если на таком интервале $f > 0$, рисуем стрелку направо, иначе — стрелку налево, как показано на рис. 2.4.1.

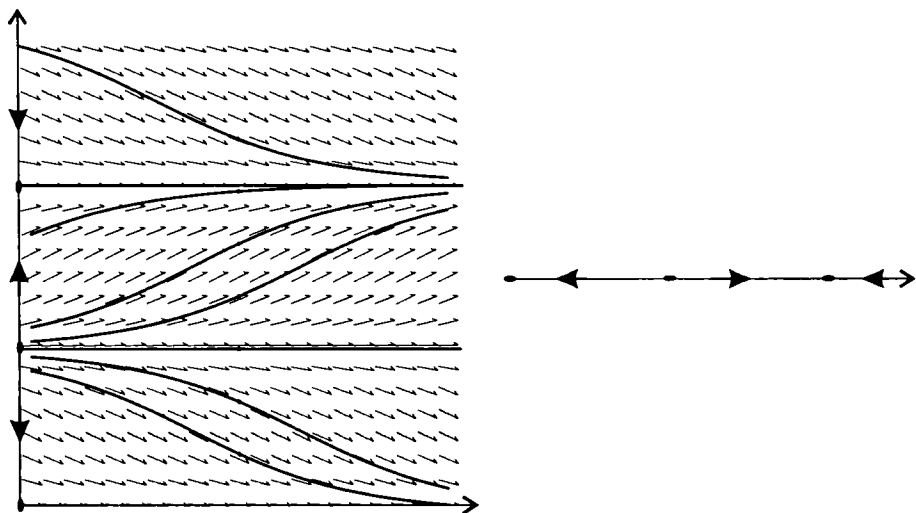


Рис. 2.4.1. Решения и фазовый портрет дифференциального уравнения

Тип неподвижных точек можно определить с помощью производной f' .

Предложение 2.4.3. Если $f(x_0) = 0$ и $f'(x_0) < 0$, то x_0 является притягивающей неподвижной точкой уравнения $\dot{x} = f(x)$, т. е. всякая близлежащая орбита асимптотически стремится к x_0 . Соответственно неподвижные точки x_0 , для которых $f'(0) > 0$, являются отталкивающими, т. е. близлежащие точки отдаляются от x_0 .

Доказательство. Если $f'(x_0) < 0$, то $f(x) < 0$ для достаточно близких точек $x > x_0$, так что такие точки двигаются к x_0 . Аналогичное рассуждение можно провести для $x < x_0$. Таким образом, каждая близлежащая орбита асимптотически стремится к x_0 . $\square$

Поэтому *фазовый портрет* уравнения $\dot{x} = f(x)$ становится понятен, если все нули функции f известны и невырождены (производная отлична от нуля). Отметим все эти точки и нарисуем стрелки к ним, если значение производной в них отрицательно, и от них, если значение производной положительно. Термин «фазовый портрет» пришел из физики (см. § 6.2).

Замечание 2.4.4. Свойство, заменяющее монотонность в модели с дискретным временем, состоит в том, что различные решения дифференциального уравнения не пересекаются. Это означает, что, как и для возрастающих отображений, сохраняется порядок. Мы сформулируем это более четко в п. 2.4.2.5.

2.4.2. Логистическое дифференциальное уравнение. В гл. 1 мы видели, что некоторые задачи из биологии популяций естественно моделируются с использованием дискретного времени. В нашем примере с колонией бабочек использовалась дискретная шкала времени с шагом в один год. Часто, однако, имеет смысл использовать модели с непрерывным временем (см. п. 1.2.1, 1.2.5—1.2.7), особенно если мы хотим исследовать рост популяции в условиях, которые изменяются медленно по сравнению с репродуктивным циклом популяции. Здесь мы также можем найти типичные примеры в биологии насекомых, поскольку довольно много разновидностей насекомых имеют короткие репродуктивные циклы (около одного дня), что делает возможным изучать прирост популяции в течение одного лета с помощью модели с непрерывным временем.

1. *Экспоненциальный рост популяции.* Самая простая модель такого рода — модель с экспоненциальным ростом (см. п. 1.2.1): предположим, что в любое данное время коэффициенты рождаемости и смертности представляют собой постоянные доли от полной численности популяции, т. е. имеется такая константа k , что если вещественная переменная x обозначает численность популяции, то $\dot{x} = kx$ или $\frac{dx}{dt} = kx$.

Лемма 2.4.5. *Решение уравнения $\dot{x} = kx$ имеет вид $x(t) = x(0)e^{kt}$.*

Доказательство. Запишем уравнение в виде $\dot{x}/x = k$, и, интегрируя по t , получим $\ln |x| = kt + C$, т. е. $|x| = e^{kt+C}$, или $x = Ae^{kt}$. Подставив в полученное соотношение $t = 0$, мы получим $A = x(0)$, т. е. $x(t) = x(0)e^{kt}$. $\square$

Приведенное выше рассуждение неприменимо, если $x(0) = 0$, однако результат все же остается справедливым.

2. *Логистическая модель.* Разумеется, любая модель с таким беспрепятственным и быстрым ростом не совсем реалистична. И действительно, не может быть, что показатели рождаемости и смертности, как доли численности популяции, не зависят от самой этой численности, т. е. что k не зависит от размера популяции. Для относительно малых популяций это хорошее приближение, но, как в примере с бабочками, для больших популяций ограниченное количество еды и, возможно, других ресурсов все же играет определенную роль. Таким образом, должна существовать популяция насыщения, которая больше не может расти, и если первоначальный размер популяции был больше, чем у популяции насы-

нения, то она должна прийти к уровню насыщения. Иными словами, k следует считать функцией от x , которая обращается в нуль при (положительном) значении насыщения $x = L$ (рост отсутствует) и отрицательна для больших значений x (уменьшающаяся популяция). Если взять линейную функцию $k = a(L - x)$, где $a > 0$, то получается дифференциальное уравнение

$$\frac{d}{dt}x = ax(L - x). \quad (2.4.1)$$

Качественное поведение решений этого уравнения легко исследовать с использованием предложения 2.4.3. Однако здесь ситуация настолько проста, что мы можем решить это дифференциальное уравнение в явном виде.

Лемма 2.4.6. *Решение уравнения $\dot{x} = ax(L - x)$ имеет вид*

$$x(t) = \frac{Lx(0)}{x(0) + (L - x(0))e^{-Lat}}.$$

Доказательство. Разделение переменных приводит к равенству

$$a = \frac{dx/dt}{x(L - x)}.$$

Интегрируя по t и раскладывая рациональную дробь в сумму простейших, мы получим

$$at + C = \int \frac{1}{x(L - x)} \frac{dx}{dt} dt = \int \frac{1}{x(L - x)} dx = \int \frac{1}{Lx} dx + \int \frac{1}{L(L - x)} dx.$$

Таким образом,

$$at + C = \frac{\ln|x|}{L} - \frac{\ln|L - x|}{L} = \frac{1}{L} \ln \left| \frac{x}{L - x} \right|.$$

Положив $t = 0$, мы найдем $CL = \ln \left| \frac{x(0)}{L - x(0)} \right|$, и, следовательно, $e^{-CL} = \left| \frac{L}{x(0)} - 1 \right|$.

Изменение знака и взятие экспоненты приводит к равенству

$$e^{-Lat} \left| \frac{L - x(0)}{x(0)} \right| = e^{-L(at+C)} = \left| \frac{L}{x} - 1 \right| = \left| \frac{L - x(t)}{x(t)} \right| = \left| \frac{L}{x(t)} - 1 \right|.$$

Величины под знаком модуля всегда имеют одинаковые знаки, так что мы можем отбросить модули и получим решение

$$x(t) = \frac{Lx(0)}{x(0) + (L - x(0))e^{-Lat}}. \quad \square$$

3. Асимптотическое поведение. Исследуем теперь асимптотическое поведение решений этого дифференциального уравнения. Для $x(0) = L$ мы получаем постоянное решение $x(t) = L$. При $t \rightarrow +\infty$ экспоненциальный член стремится к нулю, следовательно, $x(t) \rightarrow L$ для любого положительного начального условия. Если $x(0) < L$, то при $t \rightarrow -\infty$ экспоненциальный член стремится к бесконечности и $x(t) \rightarrow 0$. Для $x(0) > L$ или $x(0) < 0$ (последнее бессмысленно с биологической точки зрения) знаменатель обращается в нуль (решение имеет особенность) для

$$t = \frac{\ln(1 - [L/x(0)])}{La},$$

а это значение t отрицательно для $x(0) > L$ и положительно для $x(0) < 0$.

Поэтому асимптотическое поведение для положительного времени просто: если начальная популяция нулевая, тогда она остается нулевой навсегда. Если начальная численность популяции положительна, но ниже уровня насыщения (т. е. меньше чем L), то популяция растет и в конечном счете стремится к популяции насыщения. Скорость роста максимальна, когда численность популяции равна $L/2$, поскольку функция $x(L - x)$ имеет максимум в точке $L/2$. Если начальная численность популяции больше чем L , то популяция уменьшается, и ее численность приближается к L . Качественное поведение напоминает на рис. 2.4.1.

Говоря на языке динамических систем, мы установили, что неподвижная точка L устойчива: все близлежащие решения стремятся к ней в положительном времени. Для постоянного нулевого решения справедливо обратное. Любое близлежащее решение удаляется от него и сходится либо к решению равновесия L , либо (для биологически бессмысленных отрицательных решений) к отрицательной особенности.

4. Интерпретация взрыва. Тот факт, что решения, начинающиеся с $x(0) > L$, имеют особенность для отрицательного времени, не столь неприятен для нас. Заметим, что решение с начальным условием $x(0) = 2L$ имеет особенность в точке $t = \frac{\ln(1 - [L/x(0)])}{La} = -\ln \frac{2}{La}$. Это означает, что независимо от того, насколько большой была начальная популяция $x(0)$, мы всегда получаем $x(t) < 2L$ для $t \geq \ln \frac{2}{La}$, т. е. популяция приходит к разумному размеру (скажем, $2L$) в пределах времени, которое ограничено равномерно по $x(0)$. Чем больше особей, тем больше они голодают. Из этого также ясно, что в данной модели мы в принципе не можем получить популяцию размера больше чем L . Излишняя популяция должна была попасть в экосистему извне относительно недавно.

5. Одномерный поток. Посмотрим теперь на эти две модели (экспоненциальную и логистическую) несколько с другой стороны. В обоих случаях мы получили решения как функции времени, которые зависят от начального условия, так что множество решений описано как семейство функций времени с параметром из фазового пространства. В первом примере эта совокупность задается в виде $x_0 e^{kt}$, где x_0 — параметр. С другой стороны, мы можем считать x_0 переменной, а t параметром, или фиксированным значением. Тогда мы получаем совокупность функций $\phi^t(x) = x e^{kt}$, которая параметризована посредством t , а x выступает в качестве независимой переменной. Эта точка зрения весьма интересна, так как зависимость от параметра t не совсем произвольна. Мы имеем $\phi^{t+s}(x) = x e^{k(t+s)} = (x e^{kt}) e^{ks} = \phi^s(\phi^t(x))$. Это простое следствие из формулы $e^{a+b} = e^a e^b$ и может поэтому выглядеть как случайность. Однако то же самое свойство справедливо и во втором примере, хотя это сначала не столь очевидно: положив $\phi^t(x) := \frac{Lx}{x + (L-x)e^{-Lat}}$, мы получаем $\phi^{t+s}(x) = \phi^s(\phi^t(x))$ (см. упражнение 2.4.5). Это свойство означает, что если мы движемся вдоль решения в течение времени t и затем далее от этого значения в течение времени s , то траектория будет такая же, как и при движении вдоль решения в течение времени $t + s$.

с самого начала. Такая точка зрения будет подробно рассматриваться в п. 9.4.7 (с. 293).

Здесь же просто заметим, что, как следует из замечания 2.4.4, в имеющейся ситуации эти отображения являются возрастающими как функции переменной x (для фиксированного t).

2.4.3. Предельные циклы. Теперь перейдем к изучению менее очевидного аналога моделей с притягивающими неподвижными точками для случая непрерывного времени. Используем для этого свойства *потоков*.

Потоки будут рассматриваться в § 3.2и более детально изучаться в п. 9.4.7. Хотя доказательства в этом параграфе и несколько сложны для вводного курса, тем не менее результаты и рисунки должны помочь пониманию общего смысла.

Очевидным аналогом служат притягивающие неподвижные точки для потоков типа популяции насыщения L в предыдущем примере.

Второй аналог не встречается на прямой. Это притягивающая периодическая орбита (периодическое решение) для дифференциального уравнения на плоскости или в более высоких размерностях, как, например, уравнение Ван дер Пола из п. 1.2.8.

Мы приведем простой критерий, который позволяет проверить, что периодическая орбита является притягивающей. Этот критерий является следствием принципа сжимающих отображений, а именно предложения 2.2.19. Мы покажем, что если ϕ^t — поток и p — периодическая точка с периодом T , то из вида дифференциала ϕ^T в точке p можно заключить, что орбита $\mathcal{O}(p)$ точки p притягивающая.

В качестве первого шага заметим, что направление, совпадающее с направлением потока, не связано с устойчивостью.

Лемма 2.4.7. Если p — точка с периодом T , и она не является неподвижной для уравнения $\dot{x} = f(x)$, то 1 является собственным значением оператора $D\phi_p^T$ (см. п. 2.2.4.1).

Доказательство. Заметим, что $f(p) = f(\phi^T(p)) = \frac{d}{ds}\phi^s(p)\Big|_{s=T} = \frac{d}{ds}\phi^T \circ \phi^s(p)\Big|_{s=0} = D\phi_p^T f(p)$. Таким образом, $f(p)$ — собственный вектор оператора $D\phi_p^T$ с собственным значением 1. $\square$

Далее мы не будем учитывать это собственное значение.

Определение 2.4.8. Если p — периодическая точка с периодом T и $\lambda_1, \dots, \lambda_{n-1}$, 1 — собственные значения оператора $D\phi_p^T$ (не обязательно различные), то $\lambda_1, \dots, \lambda_{n-1}$ называются *собственными значениями в точке p* .

Замечание 2.4.9. Эти собственные значения зависят только от орбиты: если $q = \phi^s(p)$, то из равенства $\phi^T \circ \phi^s = \phi^s \circ \phi^T$ следует, что $D\phi_q^T D\phi_p^s = D\phi_p^s D\phi_q^T$, т. е. линейные отображения $D\phi_q^T$ и $D\phi_p^T$ сопряжены посредством $D\phi_p^s$, и, таким образом, собственные значения в точках p и q совпадают.

Теперь сформулируем обещанный критерий.

Предложение 2.4.10. Если p — периодическая точка, для которой все собственные значения по модулю меньше единицы, то орбита $\mathcal{O}(p)$ точки p является предельным циклом, т. е. имеется окрестность, в которой каждая точка асимптотически стремится к $\mathcal{O}(p)$.

Доказательство. Чтобы применить предложение 2.2.19, мы построим отображение, которое отражает динамику и информацию о собственных значениях. Для этого рассмотрим направление потока в точке p и возьмем небольшую часть от его ортогонального подпространства, а именно маленький диск S , содержащий p и такой, что орбита точки p пересекает его, как показано на рис. 2.4.2.

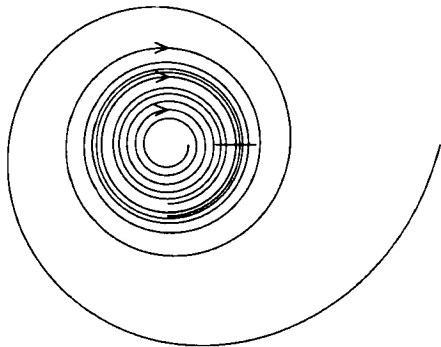


Рис. 2.4.2. Предельный цикл

При доказательстве нам несколько раз придется использовать непрерывность до момента времени $1,1 \cdot T$ (предложение 9.4.5), что можно сформулировать следующим образом.

Лемма 2.4.11. *Для данного $\epsilon > 0$ найдется такое $\delta > 0$, что любая точка, находящаяся на расстоянии меньше чем δ от $\mathcal{O}(p)$, останется на расстоянии меньше чем ϵ от $\mathcal{O}(p)$ в течение времени $1,1 \cdot T$.*

Выбрав такое ϵ , что S содержит ϵ -диск с центром в p , мы видим, что если точка $q \in S$ достаточно близка к p , то орбита снова пересекает S за время, меньшее чем $1,1 \cdot T$. Это означает, что имеется корректно определенное отображение *возвращения* (или *последования*) F_p^S на окрестность точки p в S . Гладкость (предложение 9.4.6) и теорема о неявной функции (теорема 9.2.3) обеспечивают гладкость отображения F_p^S .

Предложение 2.4.12. *Собственные значения в точке p совпадают с собственными значениями отображения DF_p^S .*

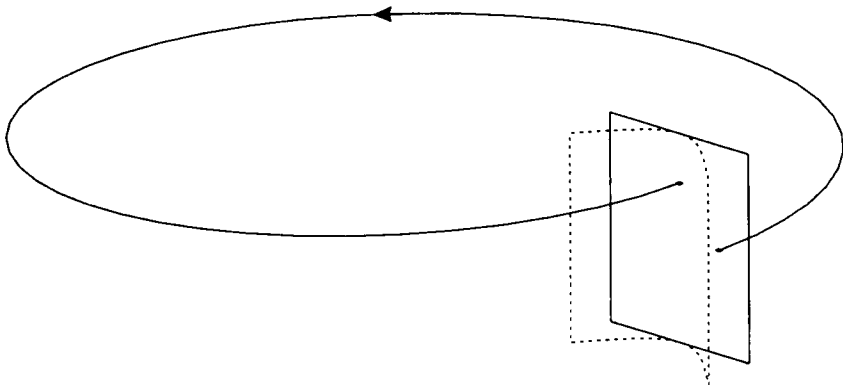


Рис. 2.4.3. Проекция на сечение

Доказательство. Если обозначить проекцию на S вдоль $f(p)$ через $\pi: \mathbb{R}^n \rightarrow S$ (см. рис. 2.4.3), то дифференциал отображения $F^S(x) = \phi^x(x)|_S$ как отображение в S имеет вид

$$DF_p^S = \pi(D\phi_p^{t_p}|_S + \dot{\phi}^{t_p}(p)Dt_p|_S).$$

Применяя отображение π к $\dot{\phi}^{t_p}(p)Dt_p|_S = F(\dot{\phi}^{t_p}(p))Dt_p|_S = F(p)Dt_p|_S$, мы получим нуль, так что $DF_p^S = \pi D\phi_p^{t_p}|_S =: A$. Но с другой стороны мы можем расширить базис в S до базиса в $\mathbb{R}^n$, добавив к нему $f(p)$, и получим координатное представление $D\phi_p^{t_p} = \begin{pmatrix} A & 0 \\ * & 1 \end{pmatrix}$. $\square$

Это означает, что согласно предложению 2.2.19 точка p является притягивающей неподвижной точкой отображения F^S с окрестностью притяжения $U \subset S$. Согласно лемме 2.4.11 каждая точка, достаточно близкая к орбите точки p , в какой-то момент времени попадет в U и начиная с этого момента будет возвращаться в U через промежутки времени, меньшие чем $1,1 \cdot T$. Полученные точки возвращения сходятся к p . Вновь используя лемму 2.4.11, мы заключаем, что вся положительная полуорбита точки q сходится к p . $\square$

Замечание 2.4.13. Предложения 2.2.20 и 2.2.21 показывают, что неподвижная точка сжатия сохраняется при возмущениях. Поскольку предельные циклы получены из принципа сжимающих отображений, они имеют то же самое свойство. Если в условиях предложения 2.4.10 динамическая система подвергается небольшим возмущениям, то отображение F_p^S на S , определенное выше, будет также немного изменяться. Так как оно является сжатием на $V \subset S$, то же самое рассуждение можно провести и для возмущений. Соответственно имеется единственная неподвижная точка в S , которая порождает периодическую точку для потока, и последний абзац доказательства предложения 2.4.10 указывает на возникновение при этом предельного цикла.

Упражнения

2.4.1. Объясните исчезновение знаков модуля в доказательстве леммы 2.4.5.

2.4.2. Рассмотрим дифференциальное уравнение $\dot{x} = -x^k$ для $k > 1$. Обозначим решение с начальным условием $x_0 > 0$ через $x(t)$. Докажите, что существует такое число $s > 0$, что предел $\lim_{t \rightarrow \infty} t^s x(t)$ конечен и отличен от нуля.

2.4.3. Приведите пример дифференциального уравнения, для которого непостоянное решение по прошествии конечного времени будет совпадать с постоянным, т. е. точка, не являющаяся неподвижной, придет к неподвижной точке за конечное время.

2.4.4. Приведите пример дифференциального уравнения, для которого решение уходит на бесконечность за конечное время.

2.4.5. Покажите, что если

$$\phi'(x) := \frac{Lx}{x + (L - x)e^{-Lat}},$$

то $\phi^{t+s}(x) = \phi^s(\phi^t(x))$.

2.4.6. Рассмотрим дифференциальное уравнение $\dot{x} = f(x)$ и предположим, что $f(x_0) = 0$ и $f'(x_0) < 0$. Покажите, что орбита любой точки $x \neq x_0$, расположенной вблизи x_0 , сходится экспоненциально к x_0 .

Задачи для углубленного изучения

2.4.7. Приведите пример дифференциального уравнения на плоскости с дифференцируемой правой частью, которое имеет такой предельный цикл, что близкие орбиты стремятся к нему не экспоненциально.

§ 2.5. Квадратичные отображения

После изучения логистического дифференциального уравнения интересно вернуться к логистическому уравнению с дискретным временем, которое рассматривалось во введении. Хотя это отображение имеет сложную динамику для больших значений параметра «плодовитости», для меньших значений параметра возможно все же сделать некоторые выводы относительно поведения, используя идеи, рассмотренные в этой главе. При доказательстве результатов предыдущего параграфа об отображениях интервала или отрезка использовалась монотонность, и поэтому они не могут быть применены непосредственно к отображению $f(x) = \lambda x(1 - x)$. Однако иногда все же удается увидеть простую динамику без использования монотонности, комбинируя локальную монотонность с понятием сжатия.

2.5.1. Притягивающие неподвижные точки для малых значений параметра. Семейство отображений $f_\lambda: [0, 1] \rightarrow [0, 1]$, $f_\lambda(x) := \lambda x(1 - x)$ для $0 \leq \lambda \leq 4$, называется *квадратичным семейством*. Это наиболее популярная модель в одномерной динамике, как в вещественной, так и в комплексной (в последнем

случае отображения продолжаются на множество $\mathbb{C}$ и часто рассматриваются комплексные значения параметра λ). Эти отображения, очевидно, не монотонны, так как они возрастают на $[0, 1/2]$ и убывают на $[1/2, 0]$. С другой стороны, как можно видеть из двух утверждений, приводимых ниже, для $\lambda < 3$ динамика проста. (Эксперименты во введении с разными значениями параметра указывают на такое поведение.)

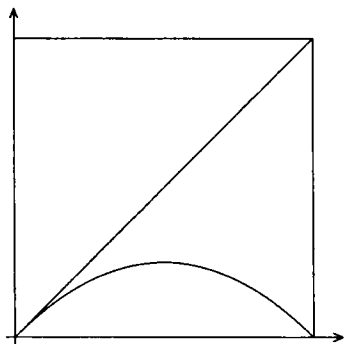


Рис. 2.5.1. Отображение f_λ

Предложение 2.5.1. Для $0 \leq \lambda \leq 1$ все орбиты отображения $f_\lambda(x) = \lambda x(1 - x)$ на $[0, 1]$ асимптотически стремятся к 0 (см. рис. 2.5.1).

Доказательство. Заметим, что $f_\lambda(x) = \lambda x(1 - x) \leq x(1 - x) < x$ для $x \neq 0$, так что последовательность $(f^n(x))_{n \in \mathbb{N}}$ убывает и ограничена снизу нулем, а следовательно, сходится в силу полноты (см. п. А.1.2). В соответствии

с соотношениями (2.3.1) предел является неподвижной точкой и, следовательно, должен быть равен нулю. (Другой подход заключается в использовании того обстоятельства, что $f_\lambda([0, 1]) \subset [0, \lambda/4] \subset [0, 1/2]$ и что отображение f_λ монотонно на $[0, 1/2]$, так что уже после первого применения f_λ можно воспользоваться рассуждениями о монотонных отображениях.) Для $\lambda < 1$ мы можем использовать принцип сжимающих отображений, поскольку $|f'_\lambda(x)| = \lambda|1 - 2x| \leq \lambda < 1$, так что f_λ является сжатием. Из этого также следует, что все орбиты для $0 \leq \lambda < 1$ (но не для $\lambda = 1$) асимптотически сходятся к неподвижной точке 0 с экспоненциальной скоростью. $\square$

Для $\lambda > 1$ ситуация несколько иная.

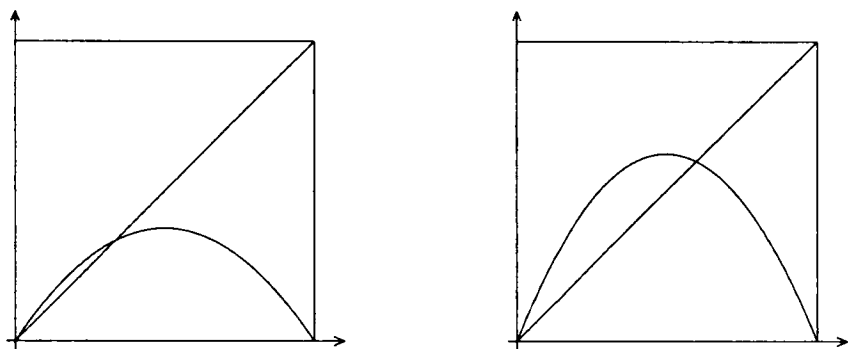
Предложение 2.5.2. Для $1 < \lambda \leq 3$ все орбиты отображения $f_\lambda(x) = \lambda x(1 - x)$ на отрезке $[0, 1]$, за исключением орбит 0 и 1, асимптотически стремятся к неподвижной точке $x_\lambda = 1 - (1/\lambda)$.

Доказательство. Уравнение $f_\lambda(x) = x$ эквивалентно квадратному уравнению $0 = \lambda x(1 - x) - x = \lambda x(1 - x - (1/\lambda))$ с ненулевым решением $x_\lambda = 1 - (1/\lambda)$, которое является новой неподвижной точкой в отрезке $[0, 1]$, если $\lambda > 1$. (Заметим, что для $\lambda < 1$ это решение также существует, однако оно отрицательное.)

Случай 1: $1 < \lambda \leq 2$. В этом случае $x_\lambda < 1/2$ (см. рис. 2.5.2) и f_λ является возрастающим отображением отрезка $[0, x_\lambda]$ в себя, причем $f_\lambda(x) = \lambda x(1 - x) > x$, и, таким образом, каждая точка x из отрезка $[0, x_\lambda]$ асимптотически стремится к x_λ .

Рассмотрим теперь точки, лежащие справа от x_λ . Заметим, что $f_\lambda(1 - x) = f_\lambda(x)$, т.е. функция f_λ симметрична относительно $x = 1/2$. Поэтому $f_\lambda([1 - x_\lambda, 1]) \subset [0, x_\lambda]$ и каждая точка $x \in (1 - x_\lambda, 1]$ также асимптотически стремится к x_λ . Остается только исследовать точки из интервала $(x_\lambda, 1 - x_\lambda)$. Вновь используя симметрию функции f_λ , мы получаем

$$f_\lambda([x_\lambda, 1 - x_\lambda]) \subset [x_\lambda, f(1/2)] = [x_\lambda, \lambda/4] \subset [x_\lambda, 1 - x_\lambda],$$

Рис. 2.5.2. Отображения $f_{1,5}$ и $f_{2,5}$

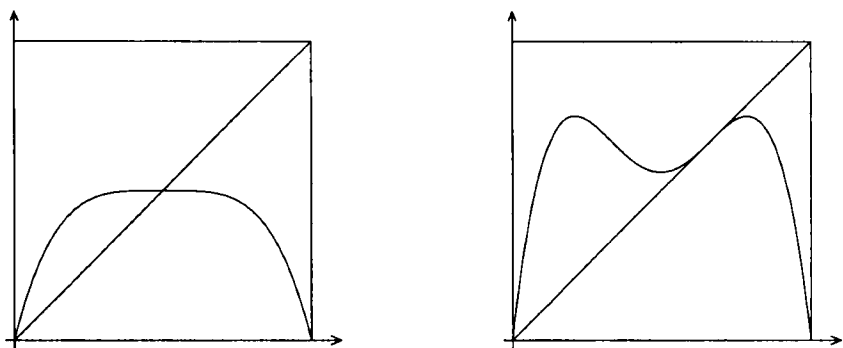
так что f отображает отрезок $[x_\lambda, 1 - x_\lambda]$ в себя. Кроме того,

$$|f'_\lambda(x)| = \lambda|1 - 2x| \leq \lambda|1 - 2x_\lambda| = \lambda\left|1 - 2\left(1 - \frac{1}{\lambda}\right)\right| = |2 - \lambda| < 1$$

для $1 < \lambda \leq 3$ и $x \in [x_\lambda, 1 - x_\lambda]$, поэтому f_λ является сжатием отрезка $[x_\lambda, 1 - x_\lambda]$.

Следовательно, все точки этого отрезка должны асимптотически стремиться к единственной неподвижной точке x_λ в нем. Таким образом, мы показали, что каждая орбита отображения f_λ для $1 < \lambda \leq 2$ (кроме 0 и 1) асимптотически стремится к ненулевой неподвижной точке этого отображения.

Случай 2: $2 < \lambda < 3$. В этом случае (см. рис. 2.5.3) мы можем применить подобное рассуждение, но оно будет более сложным.

Рис. 2.5.3. Отображения f_2^2 и f_3^2

Приведенные выше вычисления показывают, что неподвижная точка притягивающая, но трудность состоит в доказательстве того факта, что она притягивает все точки (к тому же мы не используем принцип сжимающих отображений). В этом случае ненулевая неподвижная точка $x_\lambda = 1 - 1/\lambda$ отображения f_λ находится справа от $1/2$, так что отображение f_λ уже не является возрастающим на $[0, x_\lambda]$. Сначала мы покажем, что для отрезка $I := [1 - x_\lambda, f_\lambda(1/2)] = [1/\lambda, \lambda/4]$ выполняется включение $f_\lambda(I) \subset I$. Для этого заметим, что $f_\lambda(x) \leq f_\lambda(1/2) = \lambda/4$ для $x \in [0, 1]$, в частности, для $x \in I$. С другой стороны, $f(\lambda/4) = (\lambda^2/4) - (\lambda^3/16) > 1/\lambda \in I$, потому что производная функции $q(\lambda) = \lambda^3/4 - \lambda^4/16$ положительна для $2 < \lambda < 3$, таким образом, $q(\lambda) > q(2) = 1$. Но $f(\lambda/4)$ является минимумом функ-

ции f на I , потому что $\lambda/4$ — точка отрезка I , наиболее удаленная от $1/2$. Действительно, $\lambda/4 > x_\lambda$, потому что $0 \leq ((\lambda/2) - 1)2/\lambda = (\lambda/4 - (1 - (1/\lambda)))$. Это доказывает, что $f(I) \subset I$.

Покажем теперь, что каждая орбита, кроме орбит точек 0 и 1 , в конечном счете попадает в отрезок I . Для $x \in I$ утверждение очевидно. Для $x \in (0, 1/\lambda)$ положим $x_n := f_\lambda^n(x)$ и заметим, что $f([0, 1/\lambda]) = [0, x_\lambda]$, так что если последовательность $(x_n)_{n \in \mathbb{N}}$ не имеет членов, содержащихся в I , то $x_n \leq 1/\lambda$ для всех $n \in \mathbb{N}$, откуда следует, что $x_{n+1} > x_n$ для всех $n \in \mathbb{N}$, так как $f(x) > x$ на $[0, 1/\lambda]$. Будучи возрастающей и ограниченной сверху, эта последовательность сходится к пределу $x_0 \in (0, 1/\lambda)$, который является неподвижной точкой согласно соотношениям (2.3.1). Но в этом полуинтервале неподвижных точек нет. Заметим, наконец, что для любой точки $x \in [\lambda/4, 1]$ мы имеем $f_\lambda(x) \in [0, 1/\lambda]$, следовательно, эти точки также попадают в отрезок I , и, таким образом, мы приходим к ситуации предшествующего случая всего лишь после одного шага.

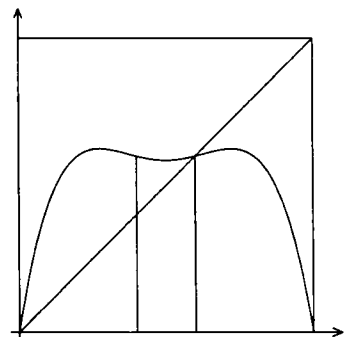


Рис. 2.5.4. Отображение $f^2_{2,5}$

Теперь мы покажем, что $f^n(x) \rightarrow x_\lambda$ при $n \rightarrow \infty$ для $x \in I$. К сожалению, f не является сжатием на I , потому что $f'_\lambda(\lambda/4) = \lambda(1 - 2(\lambda/4)) < -1$ для $\lambda > 1 + \sqrt{3} < 3$. Чтобы преодолеть эту трудность, мы рассмотрим f^2 . Заметим, что $f^2_\lambda(I) \subset I$ и, кроме того, $f_\lambda([1/\lambda, x_\lambda]) \subset [x_\lambda, \lambda/4]$ и наоборот, т.е. отрезок $[1/\lambda, x_\lambda]$ является f^2 -инвариантным. Введем обозначение $y(\lambda) := f^2_\lambda(1/2) = \lambda^2(4 - \lambda)/16$ и будем сравнивать значения y с $1/2$. Для $\lambda = 2$ мы имеем равенство, а для $\lambda = 3$ мы имеем $y = (\lambda) = 9/16 > 1/2$. Кроме того, $\frac{d^2}{d\lambda^2} y(\lambda) = \frac{8 - 6\lambda}{16} < 0$ для $\lambda \geq 2$, так что функция $y(\lambda)$ выпукла вниз, и, таким образом, $f^2_\lambda(1/2) > 1/2$ для $2 < \lambda \leq 3$. Из этого следует строгая инвариантность отрезка $J := [1/2, x_\lambda]$ относительно f^2 , т.е. $f^2(J) \subset [1/2, x_\lambda]$. На этом полуинтервале функция f^2_λ возрастает, и, следовательно, все точки асимптотически стремятся к неподвижной точке отображения f^2_λ согласно лемме 2.3.2. Единственная возможная неподвижная точка — это x_λ . Поэтому для $x \in I$ мы имеем $f^{2n}_\lambda(x) \rightarrow x_\lambda$ или $f^{2n+1}_\lambda(x) \rightarrow x_\lambda$ при $n \rightarrow \infty$. Но, например, в первом случае $f^{2n+1}_\lambda(x) \rightarrow F_\lambda(x_\lambda) = x_\lambda$, что и завершает доказательство. Это длинное рассуждение становится довольно очевидным, если посмотреть на график отображения $f^2_{2,5}$ на рис. 2.5.4. Заметим также, что согласно предложению 2.2.17 для $\lambda < 3$ сходимость к x_λ экспоненциальная, хотя мы и не используем принцип сжимающих отображений. $\square$

Поведение несколько изменяется около $\lambda = 2$. Действительно, для этого значения параметра производная неподвижной точки x_λ равна нулю, а для больших значений λ производная отрицательна. Таким образом, близлежащие точки приближаются к неподвижной точке, накапливаясь вокруг нее, в отличие от монотонности в случае $\lambda < 2$.

Случай $\lambda = 2$ интересен также и тем, что для f_2 точка $1/2$ является сверхпритягивающей (см. определение 2.2.24). Поскольку производная отображения f_2 равна нулю при $x = 1/2$, близлежащие точки приближаются к $1/2$ быстрее чем экспоненциально. Действительно, $|f_2(x) - 1/2| = |x - 1/2|^2$, т.е. погрешность возводится в квадрат на каждом шаге (см. рис. 2.2.6). Это легче заметить, если сделать замену переменных: $x' = 1/2 - x$. Мы уже встречались с подобной *квадратичной сходимостью* при изучении метода Ньютона в п. 2.2.8.

Динамика существенным образом изменяется при $\lambda \geq 3$; этот случай исследуется в гл. 11.

2.5.2. Устойчивое асимптотическое поведение. Общая черта всех примеров, приведенных в этой главе, состоит в том, что асимптотическое поведение в них является устойчивым; каждая орбита асимптотически стремится к непо-

движной точке или предельному циклу. В случае сжимающих отображений это происходит благодаря сильному предположению, из которого следует, что все орбиты стремятся к неподвижной точке. Для монотонных отображений интервала или отрезка имеется возможность более сложного поведения, потому что различные неподвижные точки получаются как предельные точки покоя. В этой ситуации мы можем наблюдать механизм, за счет которого динамика несколько усложняется: бифуркации могут увеличивать число неподвижных точек. Но все же существенно они не изменяют поведения из-за ограничений, наложенных монотонностью и теоремой о промежуточном значении. Изучение квадратичного семейства отображений пока не привело к большой дополнительной сложности. Однако в данном случае это связано лишь с тем, что мы рассматривали ограниченное множество значений параметра. Причина отсутствия какого-либо фундаментального ограничения сложности заключается в том, что квадратичные отображения необратимы и, следовательно, не сохраняют порядок. Это свойство позволяет орбитам меняться положениями и геометрически проявляется в том, что при отображении интервал или отрезок «складывается». Для больших значений параметра эта новая черта приводит к более сложной динамике.

Для дифференциальных уравнений на плоскости сложность динамики ограничена, подобно тому как она ограничена для неубывающих отображений интервала или отрезка. Любая орбита, которая не стремится к неподвижной точке, асимптотически стремится к предельному циклу. Это утверждение, известное как *теорема Пуанкаре—Бендиксона*, связано с тем, что для потоков в плоскости такие сечения, как сечение S из доказательства предложения 2.4.10, являются интервалами и отображение возвращения F^S , грубо говоря, монотонно. (Доказательство существенно опирается на *теорему Жордана*, которая также зависит от размерности и гласит, что, подобно тому как точка разделяет линию на две части, замкнутая кривая делит плоскость на две части.)

Упражнения

2.5.1. Покажите, что f_λ^2 является сжатием на $[1/2, x_\lambda]$ для $2 \leq \lambda < 3$.

2.5.2. (Альтернативная форма квадратичного семейства.) Для $g_\alpha(x) := \alpha - x^2$ и $h_\lambda = \lambda\left(x - \frac{1}{2}\right)$ покажите, что $h_\lambda(f_\lambda(x)) = g_\alpha(h_\lambda(x))$, где $\alpha = \frac{\lambda^2}{4} - \frac{\lambda}{2}$.

2.5.3. Пусть отображение $f: [0, 1] \rightarrow [0, 1]$ дифференцируемо и $|f'(x)| \leq 1$ для всех x . Покажите, что множество неподвижных точек отображения f непусто и связно (т. е. состоит из одной точки или заполняет целый интервал или отрезок).

2.5.4. Пусть отображение $f: [0, 1] \rightarrow [0, 1]$ дифференцируемо и $|f'(x)| \leq 1$ для всех x . Покажите, что всякая периодическая точка имеет период 1 или 2.

§ 2.6. Метрические пространства

Некоторые динамические системы не встречаются в естественном виде в евклидовых пространствах, и имеются ситуации, в которых полезно вводить динамические системы во вспомогательных пространствах. Мы вводим в рассмот-

рение метрические пространства в общем виде, но уделяем особое внимание случаям окружности, цилиндра и тора. Они являются фазовыми пространствами для большого числа динамических систем, в частности для многих систем, возникающих из классической механики (см. § 5.2), и билиардов. Последующие пункты посвящены канторовым множествам и канторовым лестницам¹. Подобные пространства могут показаться патологическими, однако они естественно возникают в разнообразных динамических системах (см. предложение 4.3.19 и п. 7.4.4, посвященный канторовым множествам, и рассуждения, приводящие к определению 4.4.1, а также предложению 4.4.13 для канторовых лестниц).

2.6.1. Определения. Говоря о сжимающих отображениях в евклидовых пространствах, мы не использовали специальные свойства евклидова расстояния. Действительно, мы могли бы применять то же самое рассуждение, если бы использовали другой способ измерения расстояния в $\mathbb{R}^n$, например метрику $d(x, y) = \max_{1 \leq i \leq n} |x_i - y_i|$.

Естественно, свойство сжатия зависит от того, как измеряются расстояния. Так, например, некоторые отображения, которые не являются сжимающими относительно евклидова расстояния, могут обладать этим свойством относительно расстояния максимума или же какой-либо другой метрики в $\mathbb{R}^n$. Бывает весьма полезно выбрать метрику, которая делает данное отображение сжимающим. Мы вернемся к этому в следующей главе, где сформулируем необходимое и достаточное условие применимости принципа сжимающих отображений для линейного отображения при подходящим образом выбранной метрике.

Даже предположение о том, что наше пространство является линейным или подмножеством линейного, совсем не важно для рассуждений в доказательстве предложения 2.2.10. Мы по существу используем только самые основные свойства евклидовой метрики, такие как неравенство треугольника и сходимости последовательностей Коши. Сводя наши предположения к таким основным свойствам, мы получаем весьма содержательное обобщение.

Определение 2.6.1. Пусть X — некоторое множество. Тогда отображение $d: X \times X \rightarrow \mathbb{R}$ называется *метрикой*, если

- 1) $d(x, y) = d(y, x)$ (*симметричность*),
- 2) $d(x, y) = 0 \Leftrightarrow x = y$ (*положительность*),
- 3) $d(x, y) + d(y, z) \geq d(x, z)$ (*неравенство треугольника*).

Подстановка $z = x$ в неравенство 3) и использование условий 1) и 2) приводит к неравенству $d(x, y) \geq 0$. Если d — метрика, то пара (X, d) называется *метрическим пространством*.

Подмножество метрического пространства само является метрическим пространством с той же метрикой d .

¹ В литературе на английском языке используется более выразительный термин «devil's staircase».

Нам потребуется определить несколько важных объектов, связанных с метрическими пространствами и обобщающих сходные понятия для евклидова пространства. Более подробно метрические пространства обсуждаются в приложениях.

Определение 2.6.2. Множество $B(x, r) := \{y \in X: d(x, y) < r\}$ называется (*открытым*) r -шаром с центром в x . Говорят, что последовательность $(x_n)_{n \in \mathbb{N}}$ в X *сходится* к точке $x \in X$, если для любого $\epsilon > 0$ существует такое $N \in \mathbb{N}$, что для всех натуральных чисел $n \geq N$ выполнено неравенство $d(x_n, x) < \epsilon$.

Теперь мы определим свойство метрического пространства, которое исключает «дыры», или «отсутствующие точки».

Определение 2.6.3. Говорят, что последовательность $(x_i)_{i \in \mathbb{N}}$ является *последовательностью Коши*, если для любого $\epsilon > 0$ существует такое $N \in \mathbb{N}$, что $d(x_i, x_j) < \epsilon$ для $i, j \geq N$. Метрическое пространство называется *полным*, если всякая последовательность Коши в нем сходится.

При доказательстве принципа сжимающих отображений (предложение 2.2.8 и 2.2.10) неподвижная точка получалась как предел последовательности Коши. Поэтому мы предположим, что пространство в формулировке предложения 2.6.10 является полным, и, таким образом, сможем использовать то же самое рассуждение.

Определение 2.6.4. Пусть (X, d) и (Y, d') — метрические пространства. Отображение $f: X \rightarrow Y$ называется *изометрией*, если $d'(f(x), f(y)) = d(x, y)$ для всех $x, y \in X$. Это отображение *непрерывно* в точке $x \in X$, если для любого $\epsilon > 0$ существует такое $\delta > 0$, что из неравенства $d(x, y) < \delta$ следует, что $d'(f(x), f(y)) < \epsilon$. Непрерывное взаимно однозначное (инъективное и сюръективное) отображение с непрерывным обратным отображением называется *гомеоморфизмом*. Говорят, что отображение $f: X \rightarrow Y$ *липшицево* (или удовлетворяет *условию Липшица*) с константой C , если $d'(f(x), f(y)) \leq C d(x, y)$. Отображение называется *сжимающим* (или, более точно, λ -сжимающим), если оно липшицево с константой $\lambda < 1$. Мы говорим, что две метрики *изометричны*, если тождественное отображение является изометрией между метриками. Две метрики называются *эквивалентными*, если тождественное отображение, а также обратное к нему удовлетворяют условию Липшица.

2.6.2. Окружность. Единичная окружность $S^1 = \{x \in \mathbb{R}^2: \|x\| = 1\}$ на плоскости может быть описана как совокупность комплексных чисел, по модулю равных 1. Она служит фазовым пространством для примера, приведенного в п. 2.2.7, а также и для многих других динамических систем, которые мы будем изучать в дальнейшем.

На окружности можно естественным образом ввести различные метрики. Во-первых, можно измерять расстояние между двумя точками на S^1 , используя евклидову метрику в $\mathbb{R}^2$. Это согласуется со сказанным ранее: подмножество метрического пространства само является метрическим пространством. Будем обозначать эту метрику так же, как и евклидову, через d .

С другой стороны, можно считать, что расстояние между двумя точками на S^1 — это путь, проходимый при перемещении от одной точки до другой по окружности, т.е. длина наиболее короткой дуги, соединяющей две точки. Мы называем эту метрику метрикой длины d_l , поскольку для определения расстояния здесь используются длины дуг. На первый взгляд эти две метрики совсем не похожи, однако они не так уж сильно отличаются друг от друга.

Лемма 2.6.5. *Метрики d и d_l эквивалентны.*

Доказательство. Заметим, что $d(x, y) = 2 \sin(d_l(x, y)/2)$, $d_l(x, y) \in [0, \pi/2]$ и $2t/\pi \leq 2 \sin(t/2) \leq t$ для $t \in [0, \pi/2]$. Таким образом, тождественное отображение окружности (S^1, d) с евклидовой метрикой с метрикой (S^1, d_l) с метрикой длины удовлетворяет условию Липшица с константой $\pi/2$ (см. рис. 2.6.1). Обратное отображение (тоже тождественное, но «в другом направлении») удовлетворяет условию Липшица с константой 1. Следовательно, эти две метрики эквивалентны. $\square$

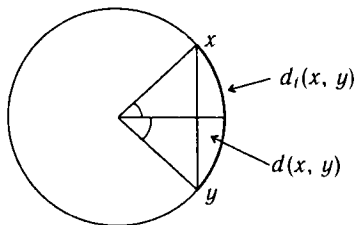


Рис. 2.6.1. Метрики d и d_l на S^1

Покажем, как иная конструкция приводит к метрическому пространству, гомеоморфному окружности S^1 и изометричному пространству (S^1, d_l) с точностью до постоянного множителя.

Мы будем использовать это пространство в дальнейшем, а сейчас это послужит интересным примером.

Рассмотрим вещественную прямую $\mathbb{R}$ и определим на ней отношение эквивалентности $\sim$, полагая $x \sim y$, если $x - y \in \mathbb{Z}$, т.е. будем считать точки эквивалентными, если они отличаются на целое число. Тем самым мы определим класс эквивалентности числа $x \in \mathbb{R}$ $[x] := \{y \in \mathbb{R} : y \sim x\}$. Класс эквивалентности нуля — само множество $\mathbb{Z}$, любой класс эквивалентности — это $\mathbb{Z}$ -сдвиги элемента класса, т.е. $[x] = x + \mathbb{Z}$. Чтобы определить новое метрическое пространство, мы рассмотрим множество $X = \mathbb{R}/\mathbb{Z} := \{[x] : x \in \mathbb{R}\}$ всех классов эквивалентности.

Замечание 2.6.6. Через $[\cdot]$ мы обозначаем классы эквивалентности. Это общепринятое обозначение, к сожалению, напоминает обозначение для целой части, которую мы, во избежание путаницы, будем обозначать $\lfloor \cdot \rfloor$.

Метрика на $\mathbb{R}/\mathbb{Z}$ индуцируется метрикой на $\mathbb{R}$.

Предложение 2.6.7. *Метрика на $X = \mathbb{R}/\mathbb{Z}$ определяется равенством $d(x, y) := \min\{|b - a| : a \in x, b \in y\}$.*

Доказательство. Метрика d , очевидным образом, симметрична. Докажем, что $d(x, y) = 0$ только в случае, когда $x = y$. Для этого сначала заметим, что метрика не изменится, если мы возьмем минимум по $a \in x$ только для фиксированного $b \in y$, поскольку наименьшее расстояние от b до элементов класса x то же самое, что наименьшее расстояние от любого целочисленного сдвига элемента b до элементов класса x . Но очевидно, что $\min\{|b - a| : a \in x\}$ достигается и, следовательно, равен нулю, если $b \in x$, а поэтому $x = y$.

Для доказательства неравенства треугольника возьмем такие элементы $x, y, z \in \mathbb{R}/\mathbb{Z}$ и $a \in x, b \in y$, что $d(x, y) = |b - a|$. Тогда для любого $c \in z$ имеем $d(x, z) \leq |c - a| \leq |c - b| + |b - a| = |c - b| + d(x, y)$. Взяв минимум по $c \in z$, приходим к неравенству $d(x, z) \leq d(y, z) + d(x, y)$. $\square$

Пример 2.6.8. Во введенной метрике имеем $d([\pi], [3/2]) = 7/2 - \pi = 0,5 - 0,14159265 \dots = 0,3584073 \dots$ и $d([0,9], [0]) = 0,1$.

Таким образом, мы получаем метрическое пространство. Чтобы понять, как оно выглядит, заметим, что каждый класс эквивалентности имеет в точности одного представителя в $[0, 1)$. Поэтому, как множество точек, мы можем отождествить $\mathbb{R}/\mathbb{Z}$ с $[0, 1)$.

Лемма 2.6.9. 1) Если $a, b \in [0, 1)$, причем $|a - b| \leq 1/2$, то $d([a], [b]) = |a - b|$.
2) Если $|a - b| \geq 1/2$, то $d([a], [b]) = 1 - |a - b|$.

Доказательство. 1) По определению $d([a], [b]) \leq |a - b|$, но неравенство не может быть строгим, поскольку каждый $\mathbb{Z}$ -сдвиг числа b расположен дальше от a , чем само число b .

2) Мы имеем $d([a], [b]) = 1 - |a - b|$, потому что это наименьшее из чисел $|a - (b - 1)|$ и $|a - (b + 1)|$. $\square$

Например, расстояние между классами $[1 - \epsilon]$ и $[0]$ равно ϵ , если $\epsilon < 1/2$. Поэтому наше построение интуитивно соответствует «приклеиванию» открытого конца полуинтервала $[0, 1)$ к 0. Или, если говорить обо всей вещественной оси $\mathbb{R}$, наша конструкция приводит к «скручиванию» всей прямой в окружность единичного радиуса, так что все целочисленные сдвиги одного и того же числа оказываются в одной точке окружности (см. рис. 2.6.2). И наоборот, переход от окружности к прямой происходит так, как если бы крутящееся велосипедное колесо оставляло за собой периодический след.

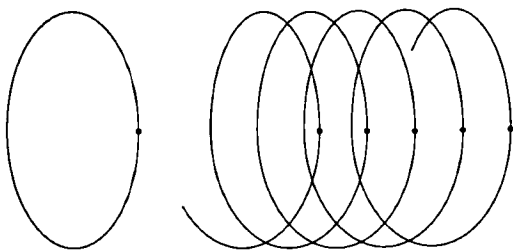


Рис. 2.6.2. Скручивание вещественной оси

Эту конструкцию можно описать аналитически: в качестве отображения прямой $\mathbb{R}$ на единичную окружность в комплексной плоскости $\mathbb{C}$ мы можем взять отображение $f(x) = e^{2\pi i x}$. Тогда $f(x + k) = e^{2\pi i(x+k)} = e^{2\pi i x} e^{2\pi i k} = e^{2\pi i x} = f(x)$, так что эта функция зависит только от класса эквивалентности $[x]$ числа x и поэтому определяет отображение из $\mathbb{R}/\mathbb{Z}$ в S^1 , задаваемое соотношением $F([x]) = f(x)$. Если использовать метрику $d_I/2\pi$ на S^1 , то F является изометрией, что оправдывает наше утверждение о том, что мы привели другой способ построения метрического пространства (S^1, d_I) .

Наша точка зрения заключается в том, что мы рассматриваем пространство классов $\mathbb{R}/\mathbb{Z}$ как *окружность*, а единичную окружность на плоскости как удобное и часто физически мотивированное *представление* окружности.

В динамике и различных приложениях математики окружность, определенная как $\mathbb{R}/\mathbb{Z}$, естественным образом возникает при изучении периодических функций. Любая система функций с общим периодом является системой функций на окружности, поскольку можно изменить масштаб независимой переменной, сделав общий период равным 1, а значения 1-периодических функций зависят только от класса эквивалентности $(\text{mod } \mathbb{Z})$ переменной. Так что такие функции корректно определены на $\mathbb{R}/\mathbb{Z}$. Особенно важный класс функций такого вида — функции $(e^{2\pi i k x})_{k \in \mathbb{Z}}$. Это помогает понять геометрический смысл задач, в которых естественным образом возникают периодические функции.

2.6.3. Цилиндр. Цилиндр — это пространство, о котором естественно думать как о трубке. Его можно определить разными способами.

Один из них связан с естественной параметризацией цилиндра

$$(\cos 2\pi t, \sin 2\pi t, z), \quad t \in \mathbb{R}, \quad -1 \leq z \leq 1.$$

Конечно же, для получения всего цилиндра нам достаточно ограничиться рассмотрением промежутка $0 \leq t \leq 1$, а в силу периодичности тригонометрических функций точки $(0, z)$ и $(1, z)$ отображаются в одну и ту же точку в $\mathbb{R}^3$. Таким образом, эту параметризацию можно представить как сворачивание единичного квадрата в трубку. Это показано на левой части рис. 2.6.3.

Рис. 2.6.3. Тор

С другой стороны, в предыдущем пункте мы обсуждали построение окружности S^1 посредством сворачивания прямой $\mathbb{R}$. При данной параметризации происходит то же самое, только появляется еще одна дополнительная «инертная» переменная z . Иными словами, мы можем описать цилиндр как произведение окружности (переменная t) и интервала (переменная z). Это пример произведения пространств, которое рассматривается в п. А.1.6.

2.6.4. Тор. Тор обычно представляют себе как поверхность бублика. Можно также рассматривать эту поверхность как результат такого вращения окружности, лежащей в плоскости xz пространства $\mathbb{R}^3$ и не пересекающей ось z , вокруг оси z , при котором центр также перемещается по окружности в плоскости xy . При таком подходе из параметризации окружности

$$(R + r \cos 2\pi\theta, 0, \sin 2\pi\theta)$$

мы получаем параметризацию тора

$$((R + r \cos 2\pi\theta) \cos 2\pi\phi, (R + r \cos 2\pi\theta) \sin 2\pi\phi, \sin 2\pi\theta).$$

Видно, что углы θ и ϕ входят только в аргументы периодических функций, поэтому тор представляется как декартово произведение двух окружностей в $\mathbb{R}^2$, вложенных в $\mathbb{R}^4$. Рассматривая тор в виде $\mathbb{T}^2 = S^1 \times S^1$, мы можем, однако, использовать описание окружности S^1 как $\mathbb{R}/\mathbb{Z}$ и тем самым задать $\mathbb{T}^2$ непосредственно как $\mathbb{R}^2/\mathbb{Z}^2$, выбирая классы эквивалентности точек $(x_1, x_2) \in \mathbb{R}^2$ при параллельных переносах на векторы с целыми координатами $(k_1, k_2) \in \mathbb{Z}^2$, т. е. $\{(x, y)\} = ([x], [y])$. Как и прежде, евклидова метрика на $\mathbb{R}^2$ индуцирует метрику на $\mathbb{T}^2$, которая совпадает с метрикой произведения $d((x_1, x_2), (y_1, y_2)) = \sqrt{(d(x_1, y_1))^2 + (d(x_2, y_2))^2}$. Продолжая процедуру «сворачивания» цилиндра, мы можем получить описание тора $\mathbb{T}^2$. Для этого возьмем единичный квадрат $[0, 1) \times [0, 1)$ и, склеивая друг с другом правое и левое ребра, получим цилиндр, а затем, склеивая друг с другом верхнее и нижнее основания цилиндра, получим тор. Можно представлять «швы» на торе как экватор вокруг отверстия бублика и меридиан вокруг его кольца. Аналогичным образом мы можем построить и описать торы $\mathbb{T}^n$ любой размерности как произведения n копий окружности или как $\mathbb{R}^n/\mathbb{Z}^n$.

Как и в случае с окружностью, мы принимаем единичный квадрат (или куб) с описанными выше отождествлениями за *определение тора*. Заметим, что поверхность «бублика» является каноническим *представлением* абстрактного тора как конкретной поверхности, в противоположность тому, что физический квадрат служит просто удобной областью для параметризации тора.

Периодические функции естественно возникают и в этом случае: такая функция $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, что $f(x + i, y + k) = f(x, y)$ при $i, k \in \mathbb{Z}$, корректно определена на торе. Так же как и в случае окружности, тор может служить полезным визуальным представлением в ситуациях, когда возникают такие дважды периодические функции, особенно в приложениях. Подобные циклические координаты часто встречаются в механике, вот почему торы играют очень важную роль во многих механических задачах.

Еще раз отметим, что как цилиндр, так и тор служат примерами произведений метрических пространств, о которых более подробно говорится в п. А.1.6.

2.6.5. Сжимающие и почти сжимающие отображения. Теперь, когда мы определили все необходимые общие понятия, мы можем показать, что принцип сжимающих отображений справедлив в любом полном метрическом пространстве.

Предложение 2.6.10 (принцип сжимающих отображений). Пусть X — полное метрическое пространство. Под действием итераций сжимающего отображения $f: X \rightarrow X$ все точки сходятся с экспоненциальной скоростью к единственной неподвижной точке отображения f .

Доказательство. Как и для случая евклидова пространства, итерирование неравенства $d(f(x), f(y)) \leq \lambda d(x, y)$ позволяет заключить, что

$$d(f^n(x), f^n(y)) \rightarrow 0 \quad \text{при } n \rightarrow \infty,$$

поэтому асимптотическое поведение всех точек одно и то же. С другой стороны, неравенство (2.2.6) показывает, что для любого $x \in X$ последовательность $(f^n(x))_{n \in \mathbb{N}}$ является последовательностью

Коши. Таким образом, для любого $x \in X$ предел этой последовательности при $n \rightarrow \infty$ существует, если пространство полное, и согласно (2.2.5) этот предел один и тот же для всех x . Как следует из неравенства (2.2.7) этот предел является неподвижной точкой x_0 отображения f .

(Заметим, что единственность неподвижной точки не зависит от полноты.) $\square$

Как и в случае евклидова пространства, мы видим, что $d(f^n(x), x_0) \leq \frac{\lambda^n}{1-\lambda} d(f(x), x)$, т.е. все орбиты сходятся к x_0 экспоненциально. Если x_0 уже известно или если оценка с помощью начального условия не требуется, то с учетом неравенств (2.2.5), мы можем получить тот же результат более непосредственно из оценки $d(f^n(x), x_0) \leq \lambda^n d(x, x_0)$.

Иногда полезно иметь в виду, что принцип сжимающих отображений может применяться и при более слабых предположениях, чем те, которые мы использовали. Действительно, посмотрев на доказательство, можно увидеть, что было бы достаточно выполнения следующего свойства.

Определение 2.6.11. Отображение f метрического пространства называется *почти сжимающим*, если найдутся такие константы $C > 0$, $\lambda \in (0, 1)$, что

$$d(f^n(x), f^n(y)) \leq C\lambda^n d(x, y) \quad (2.6.1)$$

для всех $n \in \mathbb{N}$.

Мы можем не только воспроизвести доказательство принципа сжимающих отображений при этом ослабленном предположении, но и найти метрику, в которой такое отображение становится сжимающим. В действительности эта метрика эквивалентна изначальной.

Замена метрик, которая превращает почти сжимающее отображение в просто сжимающее, имеет аналог и для отображений, не обязательно являющихся сжимающими, так что мы докажем полезный и несколько более общий результат.

Предложение 2.6.12. Если $f: X \rightarrow X$ — отображение метрического пространства и существуют такие константы $C, \lambda > 0$, что $d(f^n(x), f^n(y)) \leq C\lambda^n d(x, y)$ для всех $x, y \in X$, $n \in \mathbb{N}$, то для каждого $\mu > \lambda$ существует метрика d_μ , эквивалентная метрике d и такая, что $d_\mu(f(x), f(y)) \leq \mu d_\mu(x, y)$ для всех $x, y \in X$.

Доказательство. Возьмем такое $n \in \mathbb{N}$, что $C(\lambda/\mu)^n < 1$, и положим

$$d_\mu(x, y) := \sum_{i=0}^{n-1} d(f^i(x), f^i(y)) / \mu^i.$$

Эта метрика называется *адаптированной метрикой* или *метрикой Ляпунова* для отображения f . Метрики d и d_μ эквивалентны, так как

$$d(x, y) \leq d_\mu(x, y) \leq \sum_{i=0}^{n-1} C(\lambda/\mu)^i d(x, y) \leq \frac{C}{1 - (\lambda/\mu)} d(x, y).$$

Заметим теперь, что

$$\begin{aligned} d_\mu(f(x), f(y)) &= \sum_{i=1}^n \frac{d(f^i(x), f^i(y))}{\mu^{i-1}} = \mu(d_\mu(x, y) + \frac{d(f^n(x), f^n(y))}{\mu^n} - \\ &\quad - d(x, y)) \leq \mu d_\mu(x, y) - (1 - C(\lambda/\mu)^n) d(x, y) \leq \mu d_\mu(x, y). \quad \square \end{aligned}$$

Отсюда непосредственно следует, что мы можем получить сжимающие отображения из почти сжимающих с помощью замены метрики, поскольку для $\lambda < 1$ из определения 2.6.11 в условиях предложения 2.6.12 можно выбрать $\mu \in (\lambda, 1)$.

Следствие 2.6.13. Пусть X — полное метрическое пространство и $f: X \rightarrow X$ — почти сжимающее отображение (см. определение 2.6.11). Тогда при итерациях отображения f все точки сходятся к его неподвижной точке с экспоненциальной скоростью.

Укажем одно из главных достоинств понятия почти сжимающего отображения. Как мы только что выяснили, свойство отображения быть сжимающим может зависеть от выбранной метрики. Это не так

для почти сжимающих отображений: если отображение f удовлетворяет условию (2.6.1) и метрика d' эквивалентна метрике d , а именно если $md'(x, y) \leq d(x, y) \leq Md'(x, y)$, то

$$d'(f^n(x), f^n(y)) \leq Md(f^n(x), f^n(y)) \leq MC\lambda^n d(x, y) \leq \frac{MC}{m}\lambda^n d'(x, y).$$

Другими словами, от метрики зависит только значение постоянной C , а не существование такой постоянной.

Даже если не предполагать, что отображение гладкое, как мы это делали в предложении 2.2.20, можно утверждать, что неподвижная точка сжимающего отображения непрерывно зависит от отображения. Это свойство весьма полезно в приложениях и поэтому заслуживает особого внимания. Естественный способ выражения непрерывной зависимости состоит в том, чтобы рассматривать семейства сжимающих отображений, параметризованных элементами другого метрического пространства.

Предложение 2.6.14. Если X и Y — метрические пространства, причем пространство X полное, и $f: X \times Y \rightarrow X$ — такое непрерывное отображение, что отображение $f_y := f(\cdot, y)$ является λ -сжимающим для всех $y \in Y$, то неподвижная точка $g(y)$ отображения f_y непрерывно зависит от y .

Доказательство. Применим неравенство

$$d(x, g(y)) \leq \sum_{i=0}^{\infty} d(f_y^i(x), f_y^{i+1}(x)) \leq \frac{1}{1-\lambda} d(x, f_y(x))$$

при $x = g(y') = f(g(y'), y')$ и получим

$$d(g(y), g(y')) \leq \frac{1}{1-\lambda} d(f(g(y'), y'), f(g(y'), y)).$$

□

Упражнения

2.6.1. Покажите, что открытый r -шар является открытым множеством.

2.6.2. Покажите, что любое объединение (не обязательно конечное или счетное) открытых множеств открыто и что любое пересечение замкнутых множеств замкнуто.

2.6.3. Рассмотрим множество $\mathbb{Z}$ целых чисел как метрическое пространство с евклидовой метрикой $d(n, m) = |n - m|$. Опишите шары $\{n \in \mathbb{Z}: d(n, 0) < 1\}$ и $\{n \in \mathbb{Z}: d(n, 0) \leq 1\}$. Какой из них является открытым и какой — замкнутым?

2.6.4. Опишите все открытые множества в пространстве $\mathbb{Z}$, рассматриваемом с евклидовой метрикой $d(n, m) = |n - m|$.

2.6.5. Покажите, что внутренность любого множества открыта и что замыкание любого множества замкнуто.

2.6.6. Покажите, что граница подмножества метрического пространства является замкнутым множеством и что граница открытого множества нигде не плотна. Используя эти результаты, докажите, что граница множества нигде не плотна.

2.6.7. Определите, какие из следующих пространств являются полными с обычной метрикой): $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Z}$, $[0, 1]$.

2.6.8. Покажите, что замкнутое подмножество полного метрического пространства само полно.

Задачи для углубленного изучения

2.6.9. Пусть X — компактное метрическое пространство (см. определение А.1.18), и пусть отображение $f: X \rightarrow X$ таково, что $d(f(x), f(y)) < d(x, y)$ для любых точек $x, y, x \neq y$. Докажите, что отображение f имеет единственную неподвижную точку $x_0 \in I$ и $\lim_{n \rightarrow \infty} f^n(x) = x_0$ для любого $x \in I$.

2.6.10. Пусть X — такое полное метрическое пространство, что его диаметр не превосходит 1, и пусть отображение $f: X \rightarrow X$ таково, что $d(f(x), f(y)) \leq d(x, y) - \frac{1}{2}(d(f(x), f(y)))^2$. Докажите, что отображение f имеет единственную неподвижную точку $x_0 \in I$ и $\lim_{n \rightarrow \infty} f^n(x) = x_0$ для любого $x \in I$.

§ 2.7. Фракталы

2.7.1. Определение канторова множества. Теперь мы рассмотрим пространство, которое в курсе математического анализа часто считается весьма необычным и искусственным, а именно канторово множество. Мы увидим, однако, что множества, подобные этому, естественно возникают в динамике и составляют один из наиболее важных классов пространств, с которыми мы встретимся в нашем курсе.

1. *Геометрическое определение.* *Троичное канторово множество* или *1/3-канторово множество* описывается следующим образом. Рассмотрим единичный отрезок $C_0 = [0, 1]$ и удалим из него открытую среднюю часть $(1/3, 2/3)$, оставив два отрезка длины $1/3$, объединение которых мы обозначим как C_1 . Применим ту же самую процедуру к этим отрезкам, т. е. удалим их средние части. Оставшееся множество C_2 состоит из четырех отрезков длины $1/9$, из каждого из которых мы снова удаляем среднюю часть. Продолжая процесс индуктивно, мы получаем вложенные друг в друга множества C_n , каждое из которых состоит из 2^n отрезков длины 3^{-n} (их полная длина $(2/3)^n \rightarrow 0$). Пересечение C всех этих множеств непусто (потому что они замкнуты и ограничены и согласно предложению А.1.24). Множество C также замкнуто и ограничено, потому что таковы все множества C_n . Это множество и называется троичным канторовым множеством (см. рис. 2.7.1).

2. *Аналитическое определение.* Полезно описать это построение аналитически. Это можно сделать следующим образом.

Лемма 2.7.1. *Троичное канторово множество C есть совокупность таких чисел из отрезка $[0, 1]$, которые могут быть записаны в троичной системе счисления без использования цифры 1.*



Рис. 2.7.1. Троичное канторово множество

Доказательство. Интервал $(1/3, 2/3)$ — это в точности множество чисел, у которых в первом разряде после запятой стоит 1 (при записи в троичной системе), т. е. которые в троичной системе не могут быть записаны в виде $0,0\dots$ или $0,2\dots$ (Заметим, что $1/3$ можно записать как $0,02222\dots$, а $2/3$ как $0,20000000\dots$) Соответственно средние части оставшихся отрезков — в точности те числа, для которых второй разряд после запятой должен быть равен 1, и т. д. $\square$

3. **Свойства.** Очевидно, что троичное канторово множество в некотором смысле мало (будучи пересечением множеств, состоящих из отрезков, суммарные длины которых — постоянно уменьшающиеся числа), и, конечно, оно не может содержать какой-либо интервал. Кроме того, справедливо следующее утверждение.

Лемма 2.7.2. *Троичное канторово множество вполне несвязно (см. определение A.1.8).*

Доказательство. Любые две точки множества C находятся в различных компонентах некоторого множества C_n . Взяв достаточно малую открытую окрестность одной из них вместе с внутренностью дополнения к ней, мы получаем два таких непересекающихся открытых множества, что их объединение содержит множество C и каждое из них содержит одну из рассматриваемых точек. $\square$

Но при этом верно следующее утверждение.

Лемма 2.7.3. *Троичное канторово множество несчетно.*

Доказательство. Сопоставляя каждой точке $x = 0,\alpha_1\alpha_2\alpha_3\dots = \sum_{i=1}^{\infty} \frac{\alpha_i}{3^i} \in C$ ($\alpha_i \neq 1$) число $f(x) := \sum_{i=1}^{\infty} \frac{\alpha_i/2}{2^i} = \sum_{i=1}^{\infty} \alpha_i 2^{-i-1} \in [0, 1]$, мы определяем сюръективное отображение, потому что все двоичные разложения можно представить в таком виде. Образ множества C несчетен, следовательно, и само это множество несчетно. $\square$

В этой книге мы встретимся с многочисленными модификациями троичного канторова множества. Поэтому удобно ввести следующее понятие.

Определение 2.7.4. Любое множество, гомеоморфное троичному канторову множеству, называется просто канторовым множеством.

Предложение A.1.7 показывает, что канторовы множества характеризуются внутренним образом.

4. **Самоподобие.** Имеется интересный пример сжимающего отображения на троичном канторовом множестве, а именно отображение $f: [0, 1] \rightarrow [0, 1]$, $f(x) = x/3$. Так как отображение f сжимающее, оно является сжатием и на каждом инвариантном подмножестве, и в частности на нашем канторовом множестве. Очевидно, что 0 является единственной неподвижной точкой. Это свойство инвариантности относительно линейных сжатий часто называется *самоподобием* или *свойством перенормировки*. Его значение весьма ясно и тем не менее поразительно: микроскопическая структура канторова множества точно такая

же, как и глобальная структура; она не станет более простой в любом другом уменьшенном масштабе.

5. Канторова лестница. Троичное разложение может быть не единственным лишь для тех точек, для которых оно конечно, т. е. для троичных рациональных дробей. Они составляют в точности счетное множество концов отрезков, получаемых в процессе построения троичного канторова множества. Действительно, рассмотрим функцию $f\left(\sum_{i=1}^{\infty} \frac{\alpha_i}{3^i}\right) = \sum_{i=1}^{\infty} \frac{\alpha_i/2}{2^i}$ из доказательства леммы 2.7.3. Нетрудно видеть, что $f(1/3) = f(0,02222222\ldots) = 0,011111\ldots$ (двоичное разложение) $= 0,1$ (двоичное разложение) $= 1/2$. В то же время $f(2/3) = f(0,2000000\ldots) = 0,1$ (двоичное разложение) $= 1/2$. Аналогично мы видим, что все концы отрезков при таком отображении попарно отождествляются. Нетрудно также видеть, что f неубывает. Иногда бывает интересно продолжить отображение f на $[0, 1]$, считая его постоянным на отбрасываемых интервалах. Получающаяся непрерывная функция имеет некоторые необычные свойства. Как и в случае с канторовыми множествами, оказывается, что такие функции естественно возникают при изучении динамических систем (см. определение 4.4.1, рис. 4.4.1).

2.7.2. Другие самоподобные множества. Опишем теперь некоторые другие интересные самоподобные метрические пространства. *Ковер Серпинского* (см. рис. 2.7.2) получается из единичного квадрата удалением средней части

Рис. 2.7.2. Квадратный и треугольный ковры Серпинского

$(1/3, 2/3) \times (1/3, 2/3)$, затем удалением из каждого квадрата $(i/3, i + 1/3) \times (j/3, j + 1/3)$ среднего квадрата и т. д. Это построение можно легко описать и в терминах троичного разложения, причем способ описания немедленно обобщается и на более высокие размерности (см. упражнение 2.7.4). Можно также начать построение с равностороннего треугольника с горизонтальным основанием и разделить его на четыре конгруэнтных равносторонних треугольника, центральный из которых имеет горизонтальную верхнюю сторону. Затем мы удаляем этот центральный треугольник и продолжаем построение, оставляя каждый раз по три треугольника. *Снежинка Коха* получается из равностороннего треугольника,

если поставить на каждую его сторону равносторонний треугольник, основание которого есть средняя треть стороны, а затем продолжать этот процесс итерационно со сторонами получающегося многоугольника (см. рис. 2.7.3). Эта

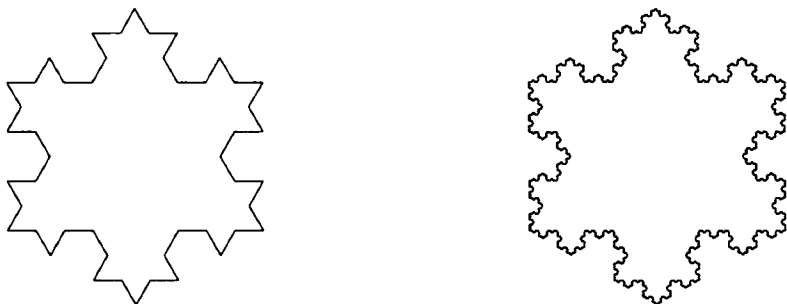


Рис. 2.7.3. Снежинка Коха

конструкция была предложена Хельге фон Кохом (1904 г.). Трехмерный вариант ковра Серпинского S — губка Серпинского или кривая Менгера, определяемая следующим образом: $\{(x, y, z) \in [0, 1]^3 : (x, y) \in S, (x, z) \in S, (y, z) \in S\}$. Это пространство получается из заполненного единичного куба, если вырезать в нем квадратную дыру со стороной $1/3$, проходящую через центр, в направлении каждой координаты, а затем восемь определенным образом расположенных квадратных дыр со стороной $1/9$ в направлении каждой координаты и т. д.

Упражнения

2.7.1. Определите, является ли троичное канторово множество полным метрическим пространством (с обычной метрикой). Ответ обоснуйте.

2.7.2. Покажите, что отображение $x \mapsto f(x) = 1 - \frac{x}{3}$ отрезка $[0, 1]$ является сжимающим и что оно отображает троичное канторово множество в себя. Где находится неподвижная точка этого отображения?

2.7.3. Прodelайте предыдущее упражнение для отображения $f(x) = \frac{x+2}{3}$.

2.7.4. Докажите, что квадратный ковер Серпинского — это множество точек в единичном квадрате, по крайней мере одна координата которых не содержит единиц в троичном разложении.

Задачи для углубленного изучения

2.7.5. Докажите, что троичное канторово множество C гомеоморфно своему декартову квадрату $C \times C$. (Мы встретимся с этим свойством в п. 7.3.3)

2.7.6. Если f — функция из доказательства леммы 2.7.3 и $(h_1, h_2): C \rightarrow C \times C$ — гомеоморфизм из предыдущего упражнения, то отображение $F(x) := (f(h_1(x)), f(h_2(x)))$ есть сюръективное отображение $F: C \rightarrow [0, 1] \times [0, 1]$. Покажите, что отображение f (а следовательно, и F) непрерывно и что F продолжается до

непрерывного отображения на $[0, 1]$. (В результате продолжения получается непрерывное сюръективное отображение $[0, 1] \rightarrow [0, 1] \times [0, 1]$, т.е. *кривая, заполняющая все пространство*, или *кривая Пеано*.)

2.7.7. Покажите, что множество C' таких точек $x \in [0, 1]$, которые могут быть записаны в системе счисления с основанием 5 без нечетных цифр, является канторовым.

2.7.8. Выясните, является ли множество таких точек $x \in [0, 1]$, которые имеют десятичную запись без нечетных цифр, канторовым.

ЛИНЕЙНЫЕ ОТОБРАЖЕНИЯ И ЛИНЕЙНЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

В этой главе рассматриваются динамические системы, имеющие несколько более сложное поведение по сравнению с уже рассмотренными примерами из гл. 2. В частности, мы изучаем системы, для которых периодические движения присутствуют и в дискретном, и в непрерывном случае. В то же самое время для большинства линейных систем, т. е. линейных отображений и линейных дифференциальных уравнений, структуру орбит легко понять (ограниченная *эллиптическая* сложность, проистекающая из комплексных собственных значений на единичной окружности, обсуждается в первых параграфах следующих двух глав).

В этой главе мы приведем детальное описание структуры орбит таких отображений и систем. Это описание опирается на линейную алгебру, но это не просто повторение, поскольку нас интересуют динамические аспекты линейных отображений и мы обращаем внимание на асимптотическое поведение при итерациях. Таким образом, в этой главе мы расширим класс известных нам видов асимптотического поведения, сначала на плоскости, а затем и в евклидовых пространствах произвольной размерности.

Кроме расширения наших знаний о возможном асимптотическом поведении, изучение линейных отображений полезно для анализа нелинейных отображений посредством линеаризации, о чем впервые говорилось в п. 2.1.2. Этот прием непосредственно применяется при изучении асимптотического поведения орбит вблизи неподвижной точки, но он также помогает и при исследовании взаимного поведения орбит и описании глобальной структуры орбит. Об этом говорится, например, в п. 6.2.2.7.

§ 3.1. Линейные отображения на плоскости

При рассмотрении линейных отображений как динамических систем, т. е. при попытках понять, как точки перемещаются под действием итераций этих отображений, нам не нужно знать всех деталей динамики рассматриваемых линейных отображений. Основное внимание следует уделять глобальным аспектам поведения, таким как приближение к началу координат, расходямость, асимптотическое приближение к прямой и скручивание в спираль.

3.1.1. Прямая. Сначала рассмотрим одномерный случай. Линейные отображения прямой описываются легко: все они имеют вид $x \mapsto \lambda x$, причем если $|\lambda| < 1$, то отображение является сжимающим и 0 — притягивающая неподвижная точка, если $|\lambda| > 1$, то все отличные от нуля орбиты стремятся к бесконечности, если же, наконец, $|\lambda| = 1$, то отображение или является тождественным, или имеет вид $x \mapsto -x$, и тогда все орбиты имеют период 2. Таким образом, линейные

отображения прямой являются (почти) сжимающими тогда и только тогда, когда $|\lambda| < 1$.

3.1.2. Собственные значения. Рассмотрим теперь отображения $x \mapsto Ax$ на плоскости также с точки зрения исследования асимптотического поведения при повторных применениях отображения. В этом случае имеется больше возможностей, и для их изучения нам понадобятся начальные сведения из линейной алгебры. Основную роль в этом анализе играют собственные значения (2×2) -матрицы A , которая представляет отображение относительно некоторого базиса. Вещественные собственные значения матрицы $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ — это такие числа λ , для которых имеется ненулевой вектор (собственный вектор) $\begin{pmatrix} x \\ y \end{pmatrix}$, удовлетворяющий условию $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}$. Геометрически это означает, что имеется прямая (подпространство, порожденное вектором $\begin{pmatrix} x \\ y \end{pmatrix}$), которая инвариантна относительно A . Динамика на прямой только что обсуждалась в п. 3.1.1, и, таким образом, она становится элементом для построения картины на плоскости. Комплексные собственные значения определены теми же самыми формулами, но с возможными комплексными векторами w . В этой ситуации вещественные векторы $w + \bar{w}$ и $i(w - \bar{w})$ порождают вещественное подпространство (собственное подпространство).

Условие $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}$ эквивалентно равенству $\begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$. Это последнее уравнение всегда имеет решение $\begin{pmatrix} x \\ y \end{pmatrix} = 0$, а ненулевое решение имеется в том и только в том случае, если матрица $\begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix}$ необратима. Таким образом, собственные векторы существуют только для тех значений λ , при которых матрица $\begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix}$ необратима, т. е. ее определитель равен нулю. Это в точности корни характеристического многочлена

$$(a - \lambda)(d - \lambda) - bc = \lambda^2 - (a + d)\lambda + ad - bc = \lambda^2 - (\operatorname{tr} A)\lambda + \det A,$$

где $\operatorname{tr} A := a + d$ и $\det A := ad - bc$. Из формулы для корней квадратного уравнения мы получаем

$$2\lambda = -\operatorname{tr} A \pm \sqrt{(\operatorname{tr} A)^2 - 4 \det A}. \quad (3.1.1)$$

Известно, что имеются три случая: это уравнение может иметь два действительных корня, один действительный корень или же два комплексно сопряженных корня, в соответствии со значением *дискриминанта* $(\operatorname{tr} A)^2 - 4 \det A$ (положительным, равным нулю или отрицательным).

3.1.3. Случай различных действительных собственных значений. Рассмотрим первый случай. Если имеются два различных действительных собственных значения λ и μ , то мы можем решить уравнения $Av = \lambda v$ и $Aw = \mu w$ для ненулевых *собственных векторов* v и w . Тогда *собственные подпространства* $\mathbb{R}v = \{tv : t \in \mathbb{R}\}$ и $\mathbb{R}w = \{tw : t \in \mathbb{R}\}$ инвариантны относительно A , и, таким образом, A сводится к одномерному линейному отображению на каждой из

этих прямых. Мы хотим выбрать эти две прямые в качестве осей новой системы координат. Это означает, что мы хотим записать матрицу отображения относительно базиса, состоящего из v и w вместо стандартных единичных векторов e_1 и e_2 . Из равенств $Au = \lambda v$ и $Aw = \mu w$ ясно, что в новых координатах матрица отображения записывается в виде $B = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$.

Предложение 3.1.1. *Линейное отображение с действительными собственными значениями $\lambda \neq \mu$ может быть приведено к диагональному виду $B = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$ линейной заменой координат.*

Полезно представлять себе, как выглядит подобная замена координат в терминах матричных операций. Нам необходима такая замена, при которой вектор v переходит в e_1 , а w — в e_2 . Поэтому матрица C обратного преобразования (относительно стандартных координат) имеет столбцы v и w (эта матрица отображает e_1 в v и e_2 в w). Таким образом, представление матрицы A в этой системе координат имеет вид $B = C^{-1}AC$. Матрица C соответствует замене новых координат на старые, матрица A осуществляет исходное преобразование, и C^{-1} отвечает за обратную замену координат. Аналогичные рассуждения приводились в п. 1.2.9.3 (с. 22).

Пример 3.1.2. Пусть $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Тогда $2\lambda = 4 \pm \sqrt{16 - 4 \cdot 3} = 4 \pm 2$ согласно (3.1.1), так что $\lambda = 1$, $\mu = 3$. Собственный вектор v для $\lambda = 1$ определяется из уравнения $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$. Так, например, мы можем взять $x = -y = 1$; $v = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Аналогично $w = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ является собственным вектором для второго собственного значения $\mu = 3$, так что матрица замены координат имеет вид $C = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ и мы получаем $B = C^{-1}AC = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$.

Частный случай $|\lambda| < 1 < |\mu|$ (или наоборот) имеет название.

Определение 3.1.3. Линейное отображение плоскости $\mathbb{R}^2$, одно собственное значение которого принадлежит интервалу $(-1, 1)$, а другое по модулю больше 1, называется *гиперболическим*.

3.1.4. Случай единственного действительного собственного значения. Во втором случае, когда имеется единственное собственное значение λ , отображение $A - \lambda \text{Id}$ необратимо и есть две возможности. В первом случае отображение $A - \lambda \text{Id}$ нулевое, и тогда $A = \lambda \text{Id}$. Такие отображения просты, но важны.

Определение 3.1.4. Отображение $A = \lambda \text{Id}$ называется *гомотетией* или *скалярным отображением*.

Во втором случае имеется одно (с точностью до скалярного множителя) ненулевое решение v уравнения $Av = \lambda v$. Пусть C — это обратимая матрица, первый столбец которой равен v . Тогда первый столбец матрицы $B = C^{-1}AC$

равен $\begin{pmatrix} \lambda & \\ 0 & \end{pmatrix}$ и другая диагональная компонента также равна λ (это число является собственным значением). Таким образом, $B = \begin{pmatrix} \lambda & s \\ 0 & \lambda \end{pmatrix}$ для некоторого $s \neq 0$ и $B^n = \begin{pmatrix} \lambda^n & ns\lambda^{n-1} \\ 0 & \lambda^n \end{pmatrix} = \lambda^n \begin{pmatrix} 1 & ns/\lambda \\ 0 & 1 \end{pmatrix}$. На самом деле можно пойти даже дальше. Для произвольного $a \neq 0$ мы можем взять $C' = \begin{pmatrix} 1 & 0 \\ 0 & 1/(sa) \end{pmatrix}$ и получим

$$B' := C'^{-1}BC = \begin{pmatrix} 1 & 0 \\ 0 & sa \end{pmatrix} \begin{pmatrix} \lambda & s \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/(sa) \end{pmatrix} = \begin{pmatrix} \lambda & a \\ 0 & \lambda \end{pmatrix}.$$

Предложение 3.1.5. *Предположим, что $a \neq 0$ и A — произвольное линейное отображение плоскости с кратным собственным значением λ , но лишь с одним линейно независимым собственным вектором. Тогда отображение A сопряжено с отображением $\begin{pmatrix} \lambda & a \\ 0 & \lambda \end{pmatrix}$. В частности, A сопряжено с отображением $\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$.*

В случае $\lambda = 1$ наблюдается характерное асимптотическое поведение, поэтому данный случай имеет специальное название.

Определение 3.1.6. Отображение, сопряженное с отображением $\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$ для некоторого $s \in \mathbb{R}$, называется *параболическим линейным отображением*.

3.1.5. Случай комплексно сопряженных собственных значений. В случае комплексно сопряженной пары корней характеристического многочлена рассмотрим матрицу A как представление отображения двумерного комплексного пространства $\mathbb{C}^2$. Пространство $\mathbb{C}^2$ — это множество пар комплексных чисел с покомпонентным сложением и умножением на (комплексные) числа. Любое линейное отображение на $\mathbb{R}^2$ определяет отображение на $\mathbb{C}^2$, если использовать матричное умножение таким же образом, как и для действительных векторов. (Можно рассматривать $\mathbb{C}^2$ как четырехмерное действительное линейное пространство почти точно так же, как мы можем считать $\mathbb{C}$ двумерным действительным линейным пространством, но нам здесь это не потребуется.)

На $\mathbb{C}^2$ отображение $A - \lambda \text{Id}$ необратимо, и, таким образом, имеются ненулевые комплексные решения уравнений $A\bar{v} = \lambda v$ и $A\bar{w} = \bar{\lambda}w$, которые можно выбрать комплексно сопряженными, $w = \bar{v}$ (потому что $\bar{\bar{A}} = A$). Взяв в качестве столбцов матрицы C векторы v и w , мы получили бы диагональную матрицу $B = C^{-1}AC$ с комплексными компонентами. Чтобы избежать этого, мы возьмем в качестве столбцов матрицы C (которая обратима) действительные векторы $v + \bar{v}$ и $-i(v - \bar{v})$ из пространства $\mathbb{R}^2$. Можно проверить, что $B := C^{-1}AC = \rho \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$, где $\lambda = \rho e^{i\theta}$. Это отображение представляет собой композицию поворота на угол θ и скалярного отображения ρId .

Предложение 3.1.7. *Линейное отображение на $\mathbb{R}^2$ с комплексным собственным значением $\rho e^{i\theta}$ сопряжено с отображением $\rho \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$.*

При $\rho = 1$ мы вновь получаем следующий важный случай.

Определение 3.1.8. Линейное отображение на $\mathbb{R}^2$, сопряженное с поворотом $\begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$, называется *эллиптическим*.

Заметьте, что если число θ/π рационально, то все орбиты такого отображения периодические.

Пример 3.1.9. Если $A = \begin{pmatrix} 7 & 8 \\ -4 & -1 \end{pmatrix}$, то $C = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$ и $B = \begin{pmatrix} 3 & 4 \\ -4 & 3 \end{pmatrix}$.

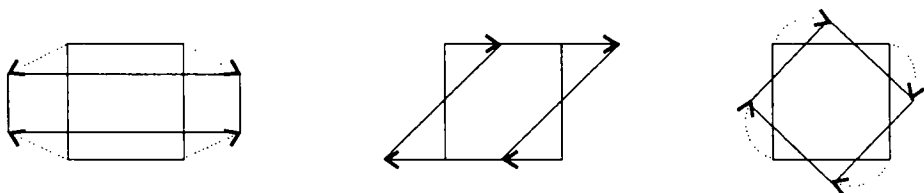


Рис. 3.1.1. Гиперболические, параболические и эллиптические отображения

Рис. 3.1.1 иллюстрирует действие матриц $\begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}$, $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ и $\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}/\sqrt{2}$ на квадрате. Это соответственно гиперболическое, параболическое и эллиптическое отображения.

3.1.6. Асимптотическое поведение. Посмотрим теперь, как может выглядеть асимптотическое поведение орбит таких отображений. Для этого полезно разъяснить, как соотносятся замена координат и замена норм, потому что во всех матричных вычислениях мы используем евклидову норму для координатных представлений векторов. Рассмотрим вектор v и его образ Cv при обратимом отображении (замене координат). Очень часто случается так, что $\|Cv\| \neq \|v\|$. Но мы можем просто определить новую норму $\|v\|' := \|Cv\|$ (это действительно норма, поскольку отображение, заданное матрицей C , линейно и обратимо; см. п. 2.6.1)¹. Таким образом, выбор евклидовой нормы в наших матричных вычислениях относительно другого базиса соответствует выбору другой нормы. Поэтому любые результаты, которые получены при таких вычислениях и не зависят от выбора нормы, позволяют нам сделать выводы об отображении относительно евклидовой нормы и не зависящие от выбора базиса.

Предложение 3.1.10. *Линейное отображение в $\mathbb{R}^2$ является почти сжимающим (см. определение 2.6.11) тогда и только тогда, когда все его собственные значения по модулю меньше единицы.*

Доказательство. Имеются три различных случая: различные действительные собственные значения, единственное двукратное действительное собственное значение и комплексные собственные значения. Во всех трех случаях отображение может быть почти сжимающим, только если модули собственных значений

¹В английском языке подобные конструкции обозначаются словом «pullback», что в дословном переводе означает «вытягивание назад».

меньше единицы. Это становится ясно, если рассмотреть канонические формы матриц в каждом случае, т. е. диагональные матрицы, верхнетреугольные матрицы и вращения с гомотетией. Поскольку свойство отображения быть почти сжимающим не зависит от нормы (потому что все нормы эквивалентны), справедливость утверждения в общем случае следует из этих рассуждений.

Теперь мы должны показать, что если собственные значения по модулю меньше 1, то отображение является почти сжимающим. В первом случае это следует из диагонализации матрицы: замена координат, которая диагонализует матрицу, определяет норму, эквивалентную стандартной. Но диагонализированная матрица определяет сжимающее отображение, а свойство отображения быть почти сжимающим сохраняется при переходе к эквивалентной метрике.

В случае одного двукратного действительного собственного значения рассмотрим такую матрицу $B = \begin{pmatrix} \lambda & a \\ 0 & \lambda \end{pmatrix}$, что $0 < 2a < 1 - |\lambda|$ (см. предложение 3.1.5). Тогда

$$\|B \begin{pmatrix} x \\ y \end{pmatrix}\| = \left\| \begin{pmatrix} \lambda x + ay \\ \lambda y \end{pmatrix} \right\| \leq |\lambda| \left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\| + \left\| \begin{pmatrix} 0 \\ y \end{pmatrix} \right\| \leq (|\lambda| + a) \left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\|,$$

поэтому отображение, заданное матрицей B , является сжимающим и, следовательно, почти сжимающим. Как мы отмечали прежде, то же самое справедливо для матрицы A , сопряженной с B , т. е. для любой матрицы с единственным собственным значением λ , потому что сопряжение приводит к замене нормы на эквивалентную. Как говорилось после следствия 2.6.13, замена нормы не затрагивает свойство отображения быть почти сжимающим.

В случае комплексных собственных значений заметим, что поворот на угол θ не изменяет евклидовой нормы любого вектора, а последующее применение отображения ρId изменяет норму на множитель $\rho < 1$, если собственные значения имеют абсолютную величину $\rho < 1$. Теперь требуемое утверждение следует из того, что всякая матрица с комплексными собственными значениями сопряжена с композицией поворота и гомотетии. $\square$

3.1.7. Структура в начале координат. Изучим теперь несколько более детально, каким образом точки приближаются к началу координат при действии этих сжатий.

1. *Различные вещественные собственные значения.* В случае двух различных действительных собственных значений мы видим, что $B^n \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \lambda^n x \\ \mu^n y \end{pmatrix}$. Предположив, что $|\mu| < |\lambda| < 1$, и записав $\begin{pmatrix} \lambda^n x \\ \mu^n y \end{pmatrix} = \lambda^n \begin{pmatrix} x \\ (\mu/\lambda)^n y \end{pmatrix}$, мы видим, что все орбиты вне оси y приближаются к нулю со скоростью $|\lambda|^n$, более «медленной» из двух скоростей. В этом случае все орбиты точек $\begin{pmatrix} x \\ y \end{pmatrix}$, $x \neq 0$, двигаются по кривым, инвариантным относительно B , которые являются касательными к оси x в нуле. Действительно, эти кривые задаются уравнением $|y| = C|x|^{|\mu|/|\lambda|}$. В результате для каждого значения C это дает четыре «ветви», по одной в каждом квадранте. В целом мы получаем рисунок с симметрией относительно обеих осей, который

называется *узлом*. Если одно или оба собственных значения отрицательны, то орбиты чередуются между двумя ветвями (см. рис. 3.1.2).

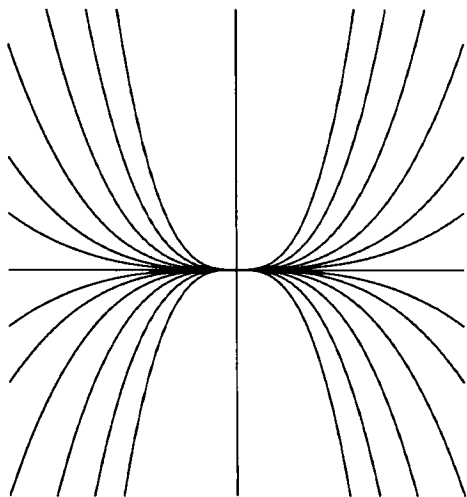


Рис. 3.1.2. Узел, $|y| = C|x|^{\mu/\ln|\lambda|}$

Чтобы проверить, что эти кривые являются инвариантными, для простоты возьмем $x, y, \lambda, \mu > 0$ и заметим, что для кривой $y = Cx^\alpha$ условие инвариантности имеет вид

$$\mu^n \cdot (Cx^\alpha) = \mu^n y = C(\lambda^n x)^\alpha = C\lambda^{\alpha n} x^\alpha,$$

и, следовательно, $\mu = \lambda^\alpha$, а значит, $\ln \mu = \alpha \ln \lambda$.

Заметим, что эти кривые гладкие (бесконечно дифференцируемые), если $|\mu| = |\lambda|^n$ для некоторого $n \in \mathbb{N}$ (в противном случае они дифференцируемы только конечное число раз). Такое соотношение между собственными значениями называется *резонансом*.

2. Одно действительное собственное значение. В случае одного действительного собственного значения λ и $A \neq \lambda \text{Id}$ мы, как было описано выше, находим матрицу $B = \begin{pmatrix} \lambda & \lambda \\ 0 & \lambda \end{pmatrix}$, сопряженную к A , и получаем $B^n = \lambda^n \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$. Отсюда нетрудно видеть, что вторая координата вектора $B^n \begin{pmatrix} x \\ y \end{pmatrix}$ равна $\lambda^n y$ и, следовательно, сходится монотонно к нулю, в то время как первая координата $\lambda^n(x + ny)$ также сходится к 0, но не обязательно монотонно, как при $y = 0$. Мы вскоре увидим, что и здесь имеются инвариантные кривые, задаваемые формулой

$$x = Cy + \frac{y \ln |y|}{\ln |\lambda|}. \quad (3.1.2)$$

Эта совокупность кривых симметрична относительно начала координат, и неподвижная точка такого типа называется *вырожденным узлом*. Здесь при отрицательном собственном значении мы тоже получаем орбиты, которые чередуются между ветвями.

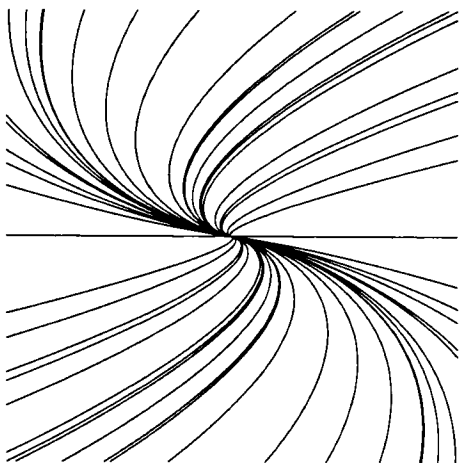


Рис. 3.1.3. Вырожденный узел, $x = Cy + \frac{y \ln |y|}{\ln |\lambda|}$

3. *Комплексные собственные значения.* В случае комплексно сопряженных собственных значений $\rho e^{\pm i\theta}$ орбиты находятся на спиралях $r = ce^{-(\theta^{-1} \ln \rho)\phi}$, $c = \text{const}$, уравнения которых записаны в полярных координатах (r, ϕ) . Подобная точка называется *фокусом* (см. рис. 3.1.4).

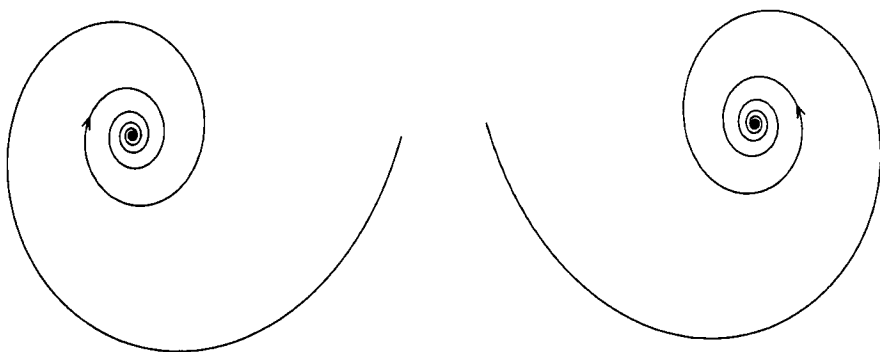
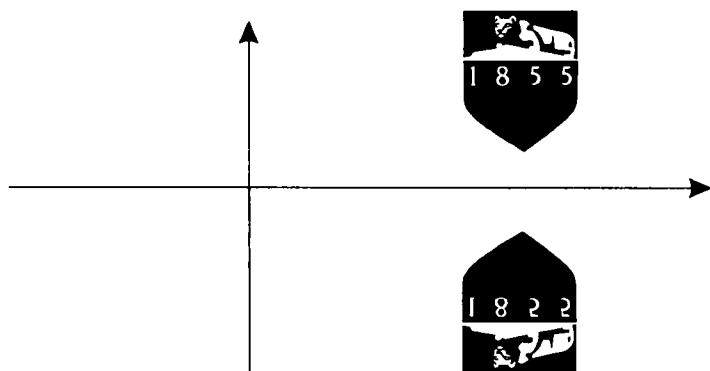


Рис. 3.1.4. Фокусы, $r = ce^{-(\theta^{-1} \ln \rho)\phi}$, $c = \text{const}$

3.1.8. Несжимающие отображения. Рассмотрим теперь ситуацию, когда отображения несжимающие, начиная со случая, когда собственные значения вещественны.

1. *Нерастягивающие отображения.* Если имеется два различных вещественных собственных значения, то они могут быть равны 1 и -1 , что соответствует отражению (см. рис. 3.1.5), или же одно из них может быть по модулю равно 1, или вообще может не быть ни одного собственного значения, по модулю равного 1. Если одно собственное значение λ по модулю равно 1, а другое собственное значение μ по модулю меньше 1, то собственное под-

Рис. 3.1.5. Собственные значения ± 1

пространство для λ состоит из неподвижных точек или точек периода 2, а все другие точки приближаются к этому собственному подпространству по прямому, параллельному собственному подпространству для μ . Иначе говоря, в этом случае $\mathbb{R}^2$ распадается на два одномерных подпространства, и на одном из них отображение A является линейным сжатием, а на другом ни A , ни A^{-1} не являются сжатиями.

2. Одно растягивающееся и одно нейтральное направление. Если $|\mu| > 1$, то все другие точки удаляются от этого собственного подпространства по тем же самым прямым, и в этом случае мы получаем разложение пространства $\mathbb{R}^2$ на растягивающееся и нейтральное подпространства.

3. Гиперболические случаи. Если оба собственных значения по модулю больше 1, то все орбиты расходятся к бесконечности по инвариантным кривым, описанным для сжимающего случая (потому что именно они возникают для обратного отображения).

Остается рассмотреть случай, когда $|\lambda| > 1$ и $|\mu| < 1$, называемый *гиперболическим (седловым)*. Переходя, как и выше, к диагональной матрице, мы получим, что точки, лежащие на оси x , расходятся вдоль оси x под действием отображения, заданного матрицей $B^n = \begin{pmatrix} \lambda^n & 0 \\ 0 & \mu^n \end{pmatrix}$, и все точки, лежащие вне оси y , расходятся вправо или влево, в то время как ординаты стремятся к 0, т. е. все орбиты приближаются к оси x асимптотически. Обратно, под действием отображения, заданного матрицей B^{-n} , все орбиты вне оси x расходятся, и ось y является асимптотой. И снова орбиты располагаются на инвариантных кривых $y = Cx^{\ln |\mu| / \ln |\lambda|}$.

Заметим, что здесь показатель степени отрицательный. Такая картина называется *седлом*. В частном случае $\mu = 1/\lambda$ эти кривые представляют собой просто стандартные гиперболы $y = C/x$, поэтому случай и назван гиперболическим. Таким образом, в этом случае $\mathbb{R}^2$ распадается на два подпространства, одно сжимающееся и одно растягивающееся, которые являются собственными подпространствами для μ и λ соответственно. Как и в случае узла, при отрицательных собственных значениях мы получаем орбиты, которые чередуются между ветвями.

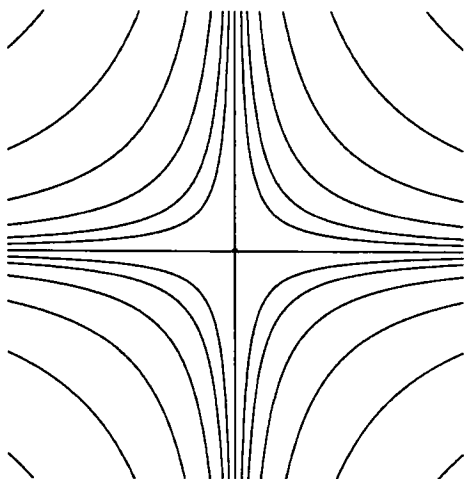


Рис. 3.1.6. Седло

4. Комплексные собственные значения. Наконец, случай комплексных собственных значений достаточно прост. Если оба собственных значения находятся на единичной окружности, тогда отображение сопряжено с поворотом. В этом случае динамика оказывается довольно интересной, даже если ограничить отображение на произвольную (инвариантную) окружность $r = \text{const}$. Такая ситуация исследуется в следующих главах. Если же собственные значения находятся вне единичной окружности, то все орбиты «раскручиваются» наружу. Все выкладки в этом случае проводятся точно так же, как и для сжимающих отображений.

3.1.9. И снова Фибоначчи.

1. Гиперболическая матрица. Полезно рассмотреть более внимательно один частный пример гиперболического отображения. Матрица $A := \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ имеет характеристический многочлен $\det \begin{pmatrix} -\lambda & 1 \\ 1 & 1-\lambda \end{pmatrix} = \lambda^2 - \lambda - 1$, так что собственные значения равны $(1 \pm \sqrt{5})/2$. Поскольку $2 < \sqrt{5} < 3$, одно из собственных значений больше 1, а другое находится в интервале $(-1, 0)$, следовательно, матрица гиперболическая. Так как два собственных вектора ортогональны (это следует из симметричности матрицы A или может быть явно проверено), картина орбит для этого отображения напоминает картину, полученную при повороте рис. 3.1.6. Собственный вектор для «растягивающего» собственного значения можно найти из уравнения

$$0 = (A - \lambda \text{Id}) \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} (-1 - \sqrt{5})x/2 + y \\ x + (1 - \sqrt{5})y/2 \end{pmatrix}.$$

Таким образом, собственным подпространством является прямая, заданная уравнением $y = (1 + \sqrt{5})x/2$.

При повторном применении гиперболического отображения все точки все более и более приближаются к растягивающемуся подпространству, как пока-

зано на рис. 3.1.6. В частности, если взять в качестве начального приближения вектор $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ и положить $\begin{pmatrix} x_n \\ y_n \end{pmatrix} := A^n \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, то мы получим $\lim_{n \rightarrow \infty} y_n/x_n = (1 + \sqrt{5})/2$.

Заметим, что последовательные значения x_n равны 1, 1, 2, 3, 5, 8, 13, ... Это последовательность Фибоначчи, которая рассматривалась в п. 1.2.2 и в примере 2.2.9 и которая представляет значительный интерес сама по себе.

2. Числа Фибоначчи. Как и в примере 2.2.9, мы обозначим числа Фибоначчи через b_n и напомним, что $b_{n+2} = b_{n+1} + b_n$. Из этого следует, что

$$\begin{pmatrix} b_{n+1} \\ b_{n+2} \end{pmatrix} = \begin{pmatrix} b_{n+1} \\ b_{n+1} + b_n \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} b_n \\ b_{n+1} \end{pmatrix} = A \begin{pmatrix} b_n \\ b_{n+1} \end{pmatrix}. \quad (3.1.3)$$

Таким образом, $x_n = b_n$ и $y_n = b_{n+1}$. Используя то обстоятельство, что $\lim_{n \rightarrow \infty} y_n/x_n = (1 + \sqrt{5})/2$, мы можем получить новое доказательство асимптотики отношения, которое было выведено в примере 2.2.9.

Рассмотрим соотношение (3.1.3) с другой точки зрения. Оно описывает преобразование первоначальной рекуррентной формулы второго порядка $b_{n+2} = b_{n+1} + b_n$ с одной переменной в рекуррентную формулу первого порядка с двумя переменными. Это пример *сведения к первому порядку*, которое бывает полезным в теории дифференциальных уравнений. В ситуации дискретного времени мы можем использовать это преобразование, чтобы получить явную (т. е. не рекуррентную) формулу для чисел Фибоначчи.

Предложение 3.1.11. Числа Фибоначчи $b_0 = 1$, $b_1 = 1$, ... могут быть вычислены по формуле

$$b_n = \frac{(1 + \sqrt{5})^{n+1} - (1 - \sqrt{5})^{n+1}}{2^{n+1}\sqrt{5}}.$$

Доказательство. Как и в п. 3.1.7.1, приведем матрицу A к диагональному виду. Для этого возьмем матрицу C , столбцы которой являются собственными векторами матрицы A , и запишем

$$C = \frac{1}{2} \begin{pmatrix} 2 & 2 \\ 1 + \sqrt{5} & 1 - \sqrt{5} \end{pmatrix}, \quad C^{-1} = \frac{1}{2\sqrt{5}} \begin{pmatrix} \sqrt{5} - 1 & 2 \\ \sqrt{5} + 1 & -2 \end{pmatrix},$$

$$C^{-1}AC = \frac{1}{2} \begin{pmatrix} 1 + \sqrt{5} & 0 \\ 0 & 1 - \sqrt{5} \end{pmatrix}.$$

Отсюда следует, что

$$A = \frac{1}{2} \begin{pmatrix} 2 & 2 \\ 1 + \sqrt{5} & 1 - \sqrt{5} \end{pmatrix} \frac{1}{2} \begin{pmatrix} 1 + \sqrt{5} & 0 \\ 0 & 1 - \sqrt{5} \end{pmatrix} \frac{1}{2\sqrt{5}} \begin{pmatrix} \sqrt{5} - 1 & 2 \\ \sqrt{5} + 1 & -2 \end{pmatrix}$$

и

$$A^n = \frac{1}{2^{n+2}\sqrt{5}} \begin{pmatrix} 2 & 2 \\ 1 + \sqrt{5} & 1 - \sqrt{5} \end{pmatrix} \begin{pmatrix} 1 + \sqrt{5} & 0 \\ 0 & 1 - \sqrt{5} \end{pmatrix}^n \begin{pmatrix} \sqrt{5} - 1 & 2 \\ \sqrt{5} + 1 & -2 \end{pmatrix} =$$

$$= \begin{pmatrix} 2(1 + \sqrt{5})^n(\sqrt{5} - 1) + 2(1 - \sqrt{5})^n(1 + \sqrt{5}) & 4(1 + \sqrt{5})^n - 4(1 - \sqrt{5})^n \\ (1 + \sqrt{5})^{n+1}(\sqrt{5} - 1) - (1 - \sqrt{5})^{n+1}(1 + \sqrt{5}) & 2(1 + \sqrt{5})^{n+1} - 2(1 - \sqrt{5})^{n+1} \end{pmatrix}.$$

Следовательно,

$$\begin{pmatrix} b_n \\ b_{n+1} \end{pmatrix} = \begin{pmatrix} x_n \\ y_n \end{pmatrix} = A^n \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1} \\ \left(\frac{1+\sqrt{5}}{2}\right)^{n+2} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+2} \end{pmatrix}. \quad \square$$

Замечание 3.1.12. Предложение 3.1.11 показывает, что числа Фибоначчи b_n являются самыми близкими целыми числами к $\frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^{n+1}$ (причем мы имеем попеременно $b_n < \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^{n+1}$ или $b_n > \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^{n+1}$). В этом смысле последовательность чисел Фибоначчи настолько близка к экспоненциальной модели роста, насколько это возможно для целочисленной модели. Пример 15.2.5 служит обобщением этого случая.

3. Разностные уравнения второго порядка. Метод, который мы только что применили к рекуррентной формуле для чисел Фибоначчи, на самом деле является общим. Если последовательность определена с помощью линейного рекуррентного соотношения второго порядка $a_{n+1} = pa_{n-1} + qa_n$, ее всегда можно преобразовать к рекуррентному соотношению первого порядка

$$\begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ p & q \end{pmatrix} \begin{pmatrix} a_{n-1} \\ a_n \end{pmatrix}.$$

В большинстве случаев эти рекуррентные соотношения так же легко исследуются, как это было для случая последовательности Фибоначчи. Чтобы убедиться в этом, заметим, что явная формула для чисел Фибоначчи представляет собой линейную комбинацию одинаковых степеней двух собственных значений. Это не удивительно, поскольку справедлив следующий результат.

Предложение 3.1.13. Если матрица $\begin{pmatrix} 0 & 1 \\ p & q \end{pmatrix}$ имеет два различных собственных значения λ и μ , то всякая последовательность, удовлетворяющая рекуррентному соотношению $a_{n+1} = pa_{n-1} + qa_n$, имеет вид $a_n = x\lambda^n + y\mu^n$.

Доказательство. Пусть v и w — собственные векторы для λ и μ соответственно. Запишем $\begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \alpha v + \beta w$. Тогда

$$\begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} = \alpha \lambda^n v + \beta \mu^n w,$$

т. е. мы получаем нужный результат для $x = \alpha v_1$ и $y = \beta w_1$. □

Замечание 3.1.14. Заметим, что x и y можно определить непосредственно из начальных условий, без нахождения собственных векторов.

Пример 3.1.15. Рекуррентное соотношение $a_{n+1} = a_{n-1}/2 + a_n/2$ из п. 1.2.3 соответствует матрице $\begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}$ с собственными значениями 1, $-1/2$, так что, как и ожидалось, улов омаров, определяемый согласно предложенной модели,

задается формулой $a_n = x + (-1/2)^n y$ для некоторых x, y . Это соответствует относительно быстрой стабилизации колебаний. С такой же ситуацией мы встречаемся при изучении логистической модели популяции для значений параметра между 2 и 3 (см. случай 2 в предложении 2.5.2).

Упражнения

3.1.1. Покажите, что функция $\|\cdot\|'$, определенная в п. 3.1.6, является нормой.

3.1.2. Пусть A — симметрическая матрица и $\lambda \neq \mu$ — ее собственные значения с собственными векторами v, w соответственно. Покажите, что $v \perp w$, т. е. что векторы v и w ортогональны.

3.1.3. Покажите, что всякая (2×2) -матрица с действительными собственными значениями и двумя ортогональными собственными векторами симметрична.

3.1.4. Сформулируйте и докажите аналог предложения 3.1.13 для случая двукратного собственного значения с единственным собственным вектором.

3.1.5. Предположим, что имеется неограниченный запас прямоугольников со сторонами 1×2 и 2×2 , которые можно сдвигать, но нельзя поворачивать. Сколькими различными способами можно заполнить полосу размера $n \times 2$?

§ 3.2. Линейные дифференциальные уравнения на плоскости

Аналоги линейных отображений для непрерывного времени — линейные дифференциальные уравнения. Здесь мы приступим к их изучению с точки зрения их асимптотического поведения.

Наличие инвариантных кривых в приведенных выше примерах не случайно. Мы можем получить описанные в этих примерах линейные отображения, решив тесно связанные с ними дифференциальные уравнения, решения которых интерполируют итерации данных отображений. Эти уравнения имеют вид

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix}$$

или, в развернутой форме,

$$\begin{aligned} \dot{x} &= a_{11}x + a_{12}y, \\ \dot{y} &= a_{21}x + a_{22}y. \end{aligned}$$

3.2.1. Узел. Случай двух различных положительных (действительных) собственных значений соответствует дифференциальному уравнению

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \ln \lambda & 0 \\ 0 & \ln \mu \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

где $\lambda, \mu > 0$, решения которого имеют вид

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} x(0)e^{t \ln \lambda} \\ y(0)e^{t \ln \mu} \end{pmatrix} = \begin{pmatrix} x(0)\lambda^t \\ y(0)\mu^t \end{pmatrix} = \begin{pmatrix} \lambda^t & 0 \\ 0 & \mu^t \end{pmatrix} \begin{pmatrix} x(0) \\ y(0) \end{pmatrix}.$$

Вернемся еще раз к п. 2.4.2.5, где мы рассматривали множество решений как семейство отображений, зависящих от параметра t . В этом примере отображения задаются матрицами $\begin{pmatrix} \lambda^t & 0 \\ 0 & \mu^t \end{pmatrix}$. Для $t = 1$ мы получаем отображение $\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$, и поэтому кривые решений параметризуют инвариантные кривые, которые мы нашли прежде. Как и в соответствующем дискретном случае, картина орбит или «фазовый портрет» называется *узлом*.

Для частного случая $A = (\ln \lambda) \text{Id}$ и $t = 1$ мы получаем отображение λId .

3.2.2. Вырожденный узел. Для линейного отображения $A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$, где $\lambda > 1$, рассмотрим дифференциальное уравнение

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \ln \lambda & 1 \\ 0 & \ln \lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

решения которого имеют вид

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} x(0)e^{t \ln \lambda} + y(0)te^{t \ln \lambda} \\ y(0)e^{t \ln \lambda} \end{pmatrix} = \begin{pmatrix} \lambda^t & t\lambda^t \\ 0 & \lambda^t \end{pmatrix} \begin{pmatrix} x(0) \\ y(0) \end{pmatrix} = \lambda^t \begin{pmatrix} x(0) + ty(0) \\ y(0) \end{pmatrix}.$$

В данном случае отображения пространства с параметром t задаются матрицами $\begin{pmatrix} \lambda^t & t\lambda^t \\ 0 & \lambda^t \end{pmatrix}$ и для $t = n$ мы получаем действие отображения $\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}^n$. Отсюда легко получить инвариантные кривые (3.1.2): логарифм модуля второй координаты и решим уравнение относительно t , после чего подставим найденное значение t в выражение для первой координаты.

3.2.3. Фокус. Линейные отображения с собственными значениями $\rho e^{\pm \theta}$ получаются из линейного дифференциального уравнения

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \ln \rho & \theta \\ -\theta & \ln \rho \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

решения которого имеют вид

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \rho^t \begin{pmatrix} x(0) \cos \theta t + y(0) \sin \theta t \\ y(0) \cos \theta t - x(0) \sin \theta t \end{pmatrix} = \rho^t \begin{pmatrix} \cos \theta t & \sin \theta t \\ -\sin \theta t & \cos \theta t \end{pmatrix} \begin{pmatrix} x(0) \\ y(0) \end{pmatrix}.$$

Таким образом, решение за время t задается отображением с матрицей

$$\rho^t \begin{pmatrix} \cos \theta t & \sin \theta t \\ -\sin \theta t & \cos \theta t \end{pmatrix},$$

т.е. поворотом с гомотетией, и эти решения параметризуют инвариантные спирали. Придерживаясь терминологии дискретного случая, назовем эту картину *фокусом*. В исключительном случае $|\rho| = 1$ решения являются просто поворотами и произвольная окружность $r = \text{const}$ является периодической орбитой. Эта картина называется *центром*. Таким образом, в отличие от дискретного случая (см. п. 3.1.8.4), динамика в данном случае полностью понятна.

3.2.4. Седло. Модель *седла* с непрерывным временем получается из уравнения

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix},$$

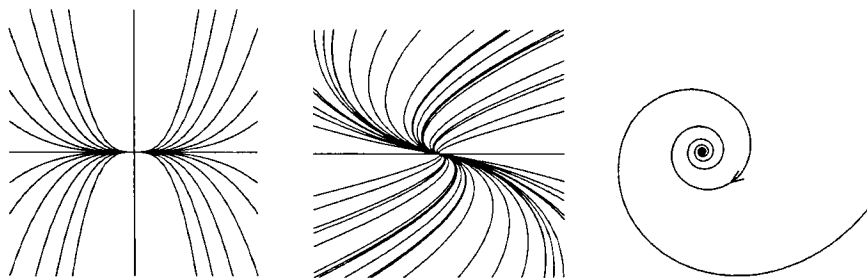


Рис. 3.2.1. Узел, вырожденный узел и фокус

где матрица A имеет одно положительное и одно отрицательное собственное значение, т. е. из уравнения

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \ln \lambda & 0 \\ 0 & \ln \mu \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

решения которого имеют вид

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} x(0)e^{t \ln \lambda} \\ y(0)e^{t \ln \mu} \end{pmatrix} = \begin{pmatrix} x(0)\lambda^t \\ y(0)\mu^t \end{pmatrix} = \begin{pmatrix} \lambda^t & 0 \\ 0 & \mu^t \end{pmatrix} \begin{pmatrix} x(0) \\ y(0) \end{pmatrix}.$$

Для $t = 1$ это в точности отображение $\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$, и поэтому кривые решения параметризуют инвариантные кривые, которые были найдены прежде. Как и в соответствующем дискретном случае, эта картина называется *седлом*.

Заметим, что мы получили почти все возможные линейные отображения как решения дифференциальных уравнений. Одно исключение — это линейные отображения с простыми отрицательными собственными значениями, которые нельзя получить из решений дифференциальных уравнений. Действительно, решение дифференциального уравнения, которое начинается на собственном подпространстве, не может его покинуть и даже не может пройти через 0 (в силу единственности решения). Другое исключение — это жордановы клетки с отрицательным собственным значением (см. задачу 3.2.5).

3.2.5. Экспонента матрицы. Связь между линейными отображениями на плоскости и двумерными линейными дифференциальными уравнениями с постоянными коэффициентами может быть выражена в более явном виде, если вспомнить, что решение дифференциального уравнения $\dot{x} = ax$ с начальным условием $x(0) = x_0$ имеет вид $x(t) = e^{at}x_0$. Аналогичным образом решение уравнения $\dot{x} = Ax$, где $x \in \mathbb{R}^n$ и A — $(n \times n)$ -матрица, определяется формулой

$$x(t) = e^{At}x(0), \quad \text{где } e^{At} := \sum_{i=0}^{\infty} \frac{A^i t^i}{i!}.$$

Каждый член ряда является $(n \times n)$ -матрицей, так что суммирование вполне корректно. Ряд сходится абсолютно, потому что каждый элемент матрицы A^i ограничен сверху числом $\|A^n\| \leq \|A\|^n$. Например, если $A = \begin{pmatrix} \ln \lambda & 0 \\ 0 & \ln \mu \end{pmatrix}$, то $e^{At} = \begin{pmatrix} \lambda^t & 0 \\ 0 & \mu^t \end{pmatrix}$,

поскольку

$$\sum_{i=0}^{\infty} \begin{pmatrix} (\ln \lambda)^i & 0 \\ 0 & (\ln \mu)^i \end{pmatrix} \frac{t^i}{i!} = \begin{pmatrix} \sum_{i=0}^{\infty} \frac{(\ln \lambda)^i t^i}{i!} & 0 \\ 0 & \sum_{i=0}^{\infty} \frac{(\ln \mu)^i t^i}{i!} \end{pmatrix} = \begin{pmatrix} e^{t \ln \lambda} & 0 \\ 0 & e^{t \ln \mu} \end{pmatrix}.$$

Лишь ненамного сложнее проверить, что

$$e \begin{pmatrix} 0 & \theta \\ -\theta & 0 \end{pmatrix} t = \begin{pmatrix} \cos \theta t & \sin \theta t \\ -\sin \theta t & \cos \theta t \end{pmatrix} \quad \text{и} \quad e \begin{pmatrix} \ln \rho & \theta \\ -\theta & \ln \rho \end{pmatrix} t = \rho^t \begin{pmatrix} \cos \theta t & \sin \theta t \\ -\sin \theta t & \cos \theta t \end{pmatrix}.$$

3.2.6. Периодические коэффициенты. В предыдущей ситуации (линейное) отображение при $t = 1$ равнялось e^A . Отсюда для начального условия $x(0)$ мы можем найти решение $x(1)$ в момент времени $t = 1$. Чтобы получить $x(n)$, достаточно проитерировать это отображение n раз. Это вполне корректно, потому что дифференциальное уравнение не содержит параметра времени в явном виде, или, иначе, $e^{Ai} = (e^A)^i$.

Это также справедливо, если время входит в дифференциальное уравнение как периодический параметр. Пусть $\dot{x} = A(t)x$, где $A(t+1) = A(t)$ и матрица M такова, что $x(1) = Mx(0)$ для любого решения $x(\cdot)$. Тогда если мы знаем $x(0)$, то можем найти $x(2)$, решив дифференциальное уравнение $\dot{x} = A(t+1)x = A(t)x$ с начальным условием $x(1)$ для $t = 1$, поэтому $x(2) = M^2x(0)$ и по индукции $x(i) = M^i x(0)$. Следовательно, дифференциальные уравнения с периодическими коэффициентами также могут быть исследованы нашими методами.

Заметим, что это наблюдение не использует линейность дифференциального уравнения. Такое же рассуждение и заключение применимы к дифференциальному уравнению $\dot{x} = f(x, t)$, где $f(x, t+1) = f(x, t)$ для $x \in \mathbb{R}^n$.

Упражнения

3.2.1. Для каждой из следующих матриц A определите, к какому (если таковой имеется) из указанных выше случаев принадлежит дифференциальное уравнение $\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix}$, а также изобразите схематически фазовый портрет (в том числе стрелки) аналогично рис. 3.2.1 (заметим, что выбранные оси будут наклонными):

$$\text{а) } \begin{pmatrix} 0 & 2 \\ -1 & 3 \end{pmatrix}, \quad \text{б) } \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix}, \quad \text{в) } \begin{pmatrix} 3 & -2 \\ 2 & 3 \end{pmatrix}, \quad \text{г) } \begin{pmatrix} 1 & -1 \\ 1 & 3 \end{pmatrix}.$$

3.2.2. Получите инвариантные кривые (3.1.2) для матрицы $\begin{pmatrix} \lambda' & t\lambda' \\ 0 & \lambda' \end{pmatrix}$.

3.2.3. Выведите уравнение инвариантных спиралей фокуса в полярных координатах.

3.2.4. Рассмотрим дифференциальные уравнения

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{3} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 10 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

а также кривые решений как графики вида $y = \varphi(x)$. В каждом из этих случаев выясните, сколько раз соответствующая функция может быть продифференцирована в нуле.

Задачи для углубленного изучения

3.2.5. Предположим, что при решении дифференциального уравнения появляется линейное отображение с двукратным отрицательным собственным значением. Докажите, что оно пропорционально тождественному отображению.

§ 3.3. Линейные отображения и дифференциальные уравнения в более высоких размерностях

Изучим линейные отображения в пространствах более высокой размерности. По сравнению со случаем отображений плоскости здесь имеется большее многообразие возможностей, поскольку существует больше различных комбинаций собственных значений — могут встречаться одновременно действительные и комплексные значения, могут быть кратные действительные или комплексные собственные значения, возможно с одновременным наличием простых или кратных собственных значений. Таким образом, логично не пытаться найти столь же подробную классификацию, как для отображений плоскости, а скорее разбить возможные типы динамически на классы в соответствии с асимптотическим поведением. На самом деле этого вполне достаточно для большинства приложений.

В качестве первого шага воспользуемся рассуждениями из предложения 3.1.10, однако в данном общем случае это потребует несколько более сложных выкладок. Для начала сформулируем более тщательно основные необходимые понятия.

3.3.1. Спектральный радиус.

Определение 3.3.1. Пусть $A: \mathbb{R}^n \rightarrow \mathbb{R}^n$ — линейное отображение. Множество его собственных значений называется *спектром* отображения A и обозначается $\text{sp } A$. Максимальный модуль собственного значения отображения A обозначается $r(A)$ и называется *спектральным радиусом* отображения A (см. рис. 3.3.1).

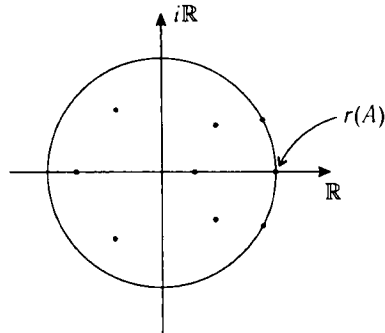


Рис. 3.3.1. Собственные значения и спектральный радиус

Для произвольного выбора нормы (см. определение A.1.29) спектральный радиус ограничен сверху нормой отображения A , определенной в (2.2.8): $r(A) \leq \|A\|$. В этом можно убедиться, если рассмотреть собственный вектор с максимальным по модулю собственным значением при условии, что это собственное значение действительное, или же если взять комплексный собственный вектор w для (комплексного) собственного значения с максимальным модулем и применить отображение A к $v := w + \bar{w}$. Для евклидовой нормы мы имеем $\|A\| = r(A)$, если матрица отображения A диагональна. Иногда бывают полезны некоторые оценки, связывающие $\|A\|$ и величину элементов матрицы отображения A .

Лемма 3.3.2. Для $(n \times n)$ -матрицы A обозначим ее элементы через a_{ij} и определим норму $|A| := \max_{ij} |a_{ij}|$. Тогда $|A| \leq \|A\| \leq \sqrt{n}|A|$.

Доказательство. Мы имеем $\|Av\| = \sqrt{\sum_{i=1}^n (\sum_{j=1}^n a_{ij} v_j)^2} \leq |A| \sqrt{\sum_{i=1}^n (\sum_{j=1}^n v_j^2)} = \sqrt{n}|A|\|v\|$, и обратное $|a_{ij}| = \langle e_i, Ae_j \rangle \leq \|A\|$. □

Следующий результат полезен для понимания динамики линейных отображений.

Предложение 3.3.3. Для всякого $\delta > 0$ существует такая норма в $\mathbb{R}^n$, что $\|A\| < r(A) + \delta$.

Доказательство этого предложения использует лемму, которая аналогична предложению 2.6.12.

Лемма 3.3.4. Рассмотрим пространство $\mathbb{R}^m$ с произвольной нормой $\|\cdot\|$ и линейное отображение $A: \mathbb{R}^m \rightarrow \mathbb{R}^m$. Если $C, \lambda > 0$ — такие константы, что $\|A^n\| \leq C\lambda^n$ для всех $n \in \mathbb{N}$, и если $\mu > \lambda$, то существует норма $\|\cdot\|'$ на $\mathbb{R}^m$, относительно которой $\|A\|' \leq \mu$.

Доказательство. Если $n \in \mathbb{N}$ таково, что $C(\lambda/\mu)^n < 1$, то можно определить норму $\|v\|' := \sum_{i=0}^{n-1} \|A^i v\|/\mu^i$, для которой

$$\|Av\|' = \sum_{i=1}^n \|A^i v\|/\mu^{i-1} = \mu \left(\|Av\| + \frac{\|A^n v\|}{\mu^n} - \|v\| \right) \leq \mu \|Av\| - \left(1 - C \frac{\lambda^n}{\mu^n}\right) \|v\| \leq \mu \|Av\|. \quad (3.3.1)$$

□

Доказательство предложения 3.3.3. Согласно предыдущей лемме достаточно показать, что существуют замена координат и норма, для которых $\|A^n\| \leq C(r(A) + (\delta/2))^n$ при всех $n \in \mathbb{N}$.

Для каждого действительного собственного значения (кратности k) рассмотрим *корневое подпространство* $E_\lambda := \{v: (A - \lambda \text{Id})^k v = 0\}$. Стандартные рассуждения линейной алгебры показывают, что $\dim(E_\lambda) = k$. (Это довольно понятно, поскольку образ единичного куба под действием отображения $A - t \text{Id}$ имеет объем $\det(A - t \text{Id}) \approx (t - \lambda)^k$, так что должно иметься k направлений «сжатия» при $t \rightarrow \lambda$.) Таким образом, эти пространства порождают все пространство. Тот же результат можно получить с помощью жордановой нормальной формы.

Используя биномиальную формулу, на E_λ мы получаем

$$A^n = (\lambda \text{Id} + \Delta)^n = \sum_{l=0}^{k-1} \binom{n}{l} \lambda^{n-l} \Delta^l = \lambda^n \sum_{l=0}^{k-1} \binom{n}{l} \lambda^{-l} \Delta^l.$$

Элементы матрицы $\Delta_n := \sum_{l=0}^{k-1} \binom{n}{l} \lambda^{-l} \Delta^l$ являются многочленами от n , так что из леммы 3.3.2 следует, что отношение $\|A^n\|/\| \lambda^n \|$ ограничено многочленом $p(n)$ от n . Если $\delta > 0$, то $\frac{p(n)\lambda^n}{(|\lambda| + \delta)^n} \rightarrow 0$; поэтому существует такое $C > 0$, что $\|A^n\|/\| \lambda^n \| \leq p(n) \leq C(|\lambda| + (\delta/2))^n / |\lambda|^n$ для всех n .

Комплексные собственные значения можно исследовать отдельно, или же можно считать A линейным отображением пространства $\mathbb{C}^n$, допуская комплексные числа в компонентах векторов. Тогда предшествующий анализ применяется также и к корневым пространствам для комплексных собственных значений.

Чтобы вывести окончательный результат из утверждения, доказанного выше для корневых пространств, мы используем норму нужного вида на каждом из этих корневых пространств. Если мы запишем векторы v в виде $(v_1, \dots, v_l)$, где каждый элемент v_l лежит в своем корневом пространстве с нормой $\|\cdot\|_l$, то требуемая норма на всем пространстве определяется формулой $\|v\| := \sum_{l=1}^l \|v_l\|_l$. □

Упражнение 3.3.2 показывает, что любая норма обладает свойством, полученным в доказательстве, хотя C и зависит от самой нормы. Однако норма, упомянутая в предложении 3.3.3, имеет специальный вид. Ее можно было получить из приведенных выше рассуждений без использования леммы, а простым применением средств линейной алгебры (жорданова нормальная форма плюс линейная замена координат, которая делает элементы вне главной диагонали произвольно малыми).

Следствие 3.3.5. Если $r(A) < 1$, то отображение A является почти сжимающим. В частности, положительные итерации каждой точки сходятся к началу координат с экспоненциальной скоростью. Если, кроме того, отображение A обратимо, т.е. если нуль не является его собственным значением, то отрицательные итерации каждой точки уходят в бесконечность с экспоненциальной скоростью.

Противоположное утверждение применимо для отображений, все собственные значения которых по модулю больше единицы.

3.3.2. Нелинейные сжатия. Наши результаты, касающиеся асимптотики линейных отображений, иногда могут быть применены и к нелинейным системам. Этот прием неоднократно применяется в гл. 9. В следующем полезном простом примере мы воспользуемся тем обстоятельством, что в подходящей норме $\|A^n\| \leq C(r(A) + (\delta/2))^n$ (см. доказательство предложения 3.3.3 или же следствие

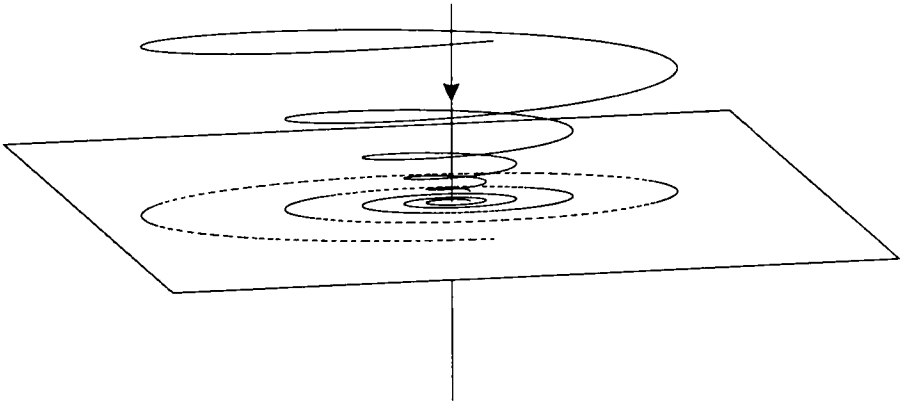


Рис. 3.3.2. Притягивающая неподвижная точка с действительным и комплексным собственными значениями

предложения 3.3.3). Применяя, как и в доказательстве предложения 2.2.17, теорему о первой производной (теорема 2.2.14), мы получаем следующий результат.

Лемма 3.3.6. Пусть f — непрерывно дифференцируемое отображение с неподвижной точкой x_0 , для которой $r(Df_{x_0}) < 1$. Тогда существует такая замкнутая окрестность U точки x_0 , что $f(U) \subset U$ и отображение f является почти сжимающим на U .

3.3.3. Случай несжимающего отображения. Остается понять, что происходит в ситуации, когда лишь некоторые собственные значения располагаются внутри или снаружи единичного круга. Подобно тому как это было сделано в двумерном случае, можно разложить $\mathbb{R}^n$ на подпространства, которые сжимаются, растягиваются или нейтральны, с той лишь разницей, что в данном случае все три возможности могут встречаться одновременно. Как и в двумерном случае, эти подпространства соответствуют множествам собственных значений внутри единичного круга, снаружи и на единичной окружности. Однако, как и в случае лишь одного собственного значения в $\mathbb{R}^2$, недостаточно рассматривать только собственные подпространства. Вместо этого следует рассмотреть *корневые подпространства* отображения A , определенные в предыдущем доказательстве для случая действительных собственных значений (хотя такая степень общности весьма желательна, читатель может заранее предположить диагонализируемость отображения, чтобы сделать рассуждения более прозрачными). Для пары комплексно сопряженных собственных значений $\lambda, \bar{\lambda}$ мы определяем $E_{\lambda, \bar{\lambda}}$ как пересечения пространства $\mathbb{R}^n$ с суммой корневых пространств, соответствующих E_λ and $E_{\bar{\lambda}}$ для комплексификации отображения A (т.е. для его расширения на комплексное пространство $\mathbb{C}^n$). Для краткости мы назовем $E_{\lambda, \bar{\lambda}}$ также корневым пространством. Пусть

$$E^- = E^-(A) = \bigoplus_{-1 < \lambda < 1} E_\lambda \oplus \bigoplus_{|\lambda| < 1} E_{\lambda, \bar{\lambda}} \quad (3.3.2)$$

обозначает пространство, порожденное всеми обобщенными корневыми пространствами для собственных значений внутри единичного круга, и подобным же образом

$$E^+ = E^+(A) = \bigoplus_{|\lambda| > 1} E_\lambda \oplus \bigoplus_{|\lambda| > 1} E_{\lambda, \bar{\lambda}}. \quad (3.3.3)$$

Если отображение A обратимо, то $E^+(A) = E^-(A^{-1})$. Наконец, положим

$$E^0 = E^0(A) = E_1 \oplus E_{-1} \oplus \bigoplus_{|\lambda|=1} E_{\lambda, \bar{\lambda}}. \quad (3.3.4)$$

Пространства E^-, E^+, E^0 очевидным образом инвариантны относительно A и $\mathbb{R}^n = E^- \oplus E^+ \oplus E^0$. Поскольку ограничение отображения A на пространство $E^-(A)$ является линейным отображением,

Рис. 3.3.3. Проекция седел и спиралей на плоскость

собственные значения которого по модулю меньше единицы, в соответствии со следствием 3.3.5 и предложением 3.3.3 мы получаем следующий результат.

Следствие 3.3.7. *Ограничение $A|_{E^-(A)}$ линейного отображения A на пространство $E^-(A)$ является почти сжимающим. Если отображение A обратимо, то и $A^{-1}|_{E^+(A)}$ является почти сжимающим. Кроме того, для любого $\delta > 0$ существует норма, относительно которой $\|A|_{E^-(A)}\| \leq r(A|_{E^-(A)}) + \delta$ и $\|A^{-1}|_{E^+(A)}\| \leq r(A^{-1}|_{E^+(A)}) + \delta$.*

Чтобы получить *норму Ляпунова*, применим предложение 3.3.3 к $E^-(A)$ и $E^+(A)$ по отдельности, получив, таким образом, нормы $\|\cdot\|_-$ и $\|\cdot\|_+$ на этих пространствах соответственно, и затем определим норму точек $x = (x_-, x_0, x_+)$ соотношением $\|(x_-, x_0, x_+)\| := \|x_-\|_- + \|x_0\| + \|x_+\|_+$.

Определение 3.3.8. Пространство $E^-(A)$ называется *сжимающимся* подпространством, а пространство $E^+(A)$ — *растягивающимся*. Отображение A называется *гиперболическим*, если $E^0 = \{0\}$, или, что эквивалентно, если $\mathbb{R}^n = E^+ \oplus E^-$.

Упражнения

3.3.1. Для каждой из следующих матриц найдите спектральный радиус соответствующего линейного отображения и опишите возможные типы долгосрочного поведения орбит $(A^n v)_{n \in \mathbb{N}}$:

$$\text{а) } \begin{pmatrix} 0 & -2 & 2 \\ 1 & 3 & -2 \\ 2 & 4 & -3 \end{pmatrix}, \quad \text{б) } \frac{1}{2} \begin{pmatrix} -1 & -1 & 0 \\ 2 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix}, \quad \text{в) } \begin{pmatrix} 2 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 1 \\ 0 & -1 & 0 & 1 \end{pmatrix}.$$

3.3.2. Пусть A — линейное отображение. Покажите, что для любой нормы существует такая константа C , что $\|A^n\| \leq C(r(A) + (\delta/2))^n$ для всех $n \in \mathbb{N}$.

3.3.3. Докажите, что $r(A) = \lim_{n \rightarrow \infty} \|A^n\|^{1/n}$.

3.3.4. Пусть линейное отображение $x \mapsto Ax$ является сжатием на $\mathbb{R}^n$. Покажите, что $\text{tr } A < n$.

3.3.5. Пусть A — такая (3×3) -матрица, что $\det A = 1/10$ и $\operatorname{tr} A = 2,7$. Покажите, что A не является сжатием.

Задачи для углубленного изучения

3.3.6. Предположим, что все собственные значения (3×3) -матрицы A по модулю равны единице. Докажите, что существует такая константа C , что для любого $v \in \mathbb{R}^3$ и любого $n \neq 0$ выполняется неравенство $\|A^n v\| \leq C n^2 \|v\|$.

3.3.7. Обобщите результат предыдущего упражнения для случая матрицы порядка $n \times n$.

3.3.8. Норма, определенная в доказательстве леммы 3.3.4, не может быть получена из скалярного произведения. Докажите, что существует модифицированное скалярное произведение, для которого индуцированная норма $\|x\| = \sqrt{\langle x, x \rangle}$ совпадает с нормой из доказательства.

РЕКУРРЕНТНОСТЬ И РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ НА ОКРУЖНОСТИ

До настоящего момента мы имели дело с динамическими системами, в которых асимптотическое поведение описывалось очень просто: каждая орбита или была неподвижной (иногда периодической), или же притягивалась к (возможно, различным) неподвижным точкам при стремлении времени к плюс или минус бесконечности. В отдельных случаях, как, например, в предложении 2.3.5, мы показали, что никакое другое поведение вообще невозможно.

В этой главе мы впервые столкнемся с существенно иным типом поведения. Аналитики используют довольно безобидный термин «квазипериодический», чтобы как-то описать это поведение и указать, что оно является всего лишь некоторым обобщением периодического поведения. Но с точки зрения динамики это отправная точка для понимания *нетривиальной рекуррентности*, центральной парадигмы теории динамических систем. Мы начинаем с тщательного изучения этого явления в самом простом возможном случае — для поворотов окружности — и уже в § 4.2 получаем замечательный набор интересных приложений. В последнем параграфе обобщаются некоторые из наших результатов на случай нелинейных отображений окружности.

§ 4.1. Повороты окружности

Описание нашего первого примера удивительно просто — оно тесно связано с некоторыми линейными динамическими системами, которые вводились в гл. 3, а именно в п. 3.1.4, при $\rho = 1$. Для линейной системы с парой комплексно сопряженных собственных значений, по модулю равных 1, сложное поведение может возникнуть на инвариантных окружностях $r = \text{const}$. Итак, мы приступим к изучению поворотов окружности.

4.1.1. Повороты окружности. В п. 2.6.2 мы рассмотрели два различных удобных способа представления окружности, которые позволяют писать разные формулы в удобном виде. Можно использовать мультипликативные обозначения, когда рассматривается единичная окружность в комплексной плоскости

$$S^1 = \{z \in \mathbb{C} : |z| = 1\} = \{e^{2\pi i \phi} : \phi \in \mathbb{R}\},$$

или же аддитивные обозначения

$$S^1 = \mathbb{R}/\mathbb{Z},$$

когда окружность представляется как множество вещественных чисел с отождествленными целочисленными сдвигами (см. рис. 2.6.2). В мультипликативной записи все алгебраические действия понимаются как операции над комплексными числами. В аддитивной записи мы можем использовать сложение и вычитание (но

не умножение или деление) так же, как и обычные действия над вещественными числами, но имея в виду, что все равенства имеют смысл с точностью до целого числа. Общепринято писать после таких равенств «(mod 1)». Таким образом, выражение $a = b \pmod{1}$, где a и b — вещественные числа, означает, что $a - b$ является целым числом.

Логарифмическое отображение

$$e^{2\pi i\phi} \mapsto \phi$$

задает изоморфизм между этими представлениями. Будем измерять длину дуг на окружности параметром ϕ , т. е. будем считать, что длина всей окружности равна единице. Пусть $\ell(\Delta)$ обозначает длину дуги Δ , измеренной таким способом. Чтобы подобным образом определить расстояние, введем метрику на множестве $X = \mathbb{R}/\mathbb{Z} := \{[x] : x \in \mathbb{R}\}$, состоящем из классов эквивалентности, полагая $d(x, y) := \min\{|b - a| : a \in x, b \in y\}$.

Поворот на угол $2\pi\alpha$ будем обозначать R_α . В мультипликативной записи

$$R_\alpha(z) = z_0 z, \quad \text{где } z_0 = e^{2\pi i\alpha}.$$

Не удивительно, что в аддитивной записи мы получаем

$$R_\alpha(x) = x + \alpha \pmod{1}. \quad (4.1.1)$$

Итерации поворотов можно выразить как

$$R_\alpha^n(z) = R_{n\alpha}(z) = z_0^n z$$

в мультипликативной записи и

$$R_\alpha^n(x) = x + n\alpha \pmod{1}$$

в аддитивной записи.

Ключевое различие в динамике поворотов окружностей появляется при рассмотрении рациональных и иррациональных значений угла вращения α .

В первом случае мы пишем $\alpha = p/q$, где p, q — взаимно простые целые числа. Тогда $R_\alpha^q(x) = x$ для всех x , так что R_α^q — тождественное отображение и после q итераций преобразование просто повторяет себя. Таким образом, полная орбита любой точки — это конечное множество, а все орбиты являются q -периодическими.

4.1.2. Плотность орбит. Случай иррационального числа α намного интереснее. Прежде всего, из вышеуказанных формул для итераций немедленно вытекает, что орбита каждой точки — бесконечное множество. Мы можем, однако, сказать нечто большее.

Предложение 4.1.1. *Если число α иррационально, то каждая положительная полуорбита поворота R_α представляет собой всюду плотное подмножество окружности.*

Доказательство. Пусть $x, z \in S^1$. Покажем, что точка z лежит в замыкании положительной полуорбиты точки x . Выберем $\epsilon > 0$. Положительная полуорбита точки x бесконечна, и ни для какого множества из $k \geq \lceil 1/\epsilon \rceil + 1$ точек

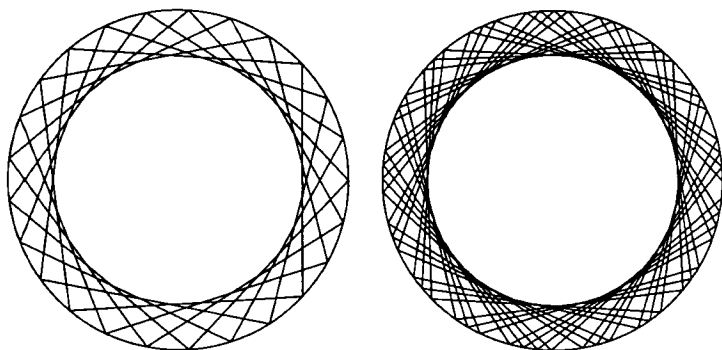


Рис. 4.1.1. Периодическая орбита и отрезок плотной орбиты

все попарные расстояния между его точками не могут превышать ε . Поэтому найдутся такие числа $k, l \in \mathbb{N}$, что $l < k \leq \lfloor 1/\varepsilon \rfloor$ и $d(R_\alpha^k(x), R_\alpha^l(x)) < \varepsilon$. Тогда $d(R_\alpha^{k-l}(x), x) < \varepsilon$, поскольку отображение R_α^{-l} сохраняет расстояние. При этом последнее расстояние не зависит от x , так как если $y \in S^1$, то $y = R_{y-x}(x)$ и

$$\begin{aligned} d(R_\alpha^{k-l}(y), y) &= d(R_\alpha^{k-l}(R_{y-x}(x)), R_{y-x}(x)) = d(R_{(k-l)\alpha+y-x}(x), R_{y-x}(x)) = \\ &= d(R_{y-x}(R_\alpha^{k-l}(x)), R_{y-x}(x)) = d(R_\alpha^{k-l}(x), x), \end{aligned}$$

поэтому числа k и l можно выбрать независимо от x .

Выберем такое $\theta \in [-1/2, 1/2]$, что $\theta = (k-l)\alpha \pmod{1}$. Тогда $\rho := |\theta| < \varepsilon$ и $R_\alpha^{k-l} = R_\theta$. Пусть $N = \lfloor 1/\rho \rfloor + 1$ (N не зависит от x). Тогда подмножество $\{R_{i\theta}(x) : i = 0, 1, \dots, N\}$ положительной полуорбиты точки x делит окружность на отрезки, длина которых меньше чем $\rho < \varepsilon$, следовательно, найдется такое $n \leq N(k-l)$, что $d(R_\alpha^n(x), z) < \varepsilon$. $\square$

Замечание 4.1.2. Так как отрицательная полуорбита вращения R_α совпадает с положительной полуорбитой вращения $R_{-\alpha}$, мы также доказали плотность отрицательных полуорбит.

Альтернативное доказательство минимальности показывает отсутствие нетривиальных замкнутых инвариантных множеств рассуждением от противного.

Другое доказательство предложения 4.1.1. Пусть $A \subset S^1$ — инвариантное замкнутое множество. Тогда дополнение к нему $S^1 \setminus A$ — непустое открытое инвариантное множество, состоящее из непересекающихся интервалов. Пусть I — самый длинный из этих интервалов (или один из самых длинных, если несколько интервалов имеют равную длину). Поскольку поворот сохраняет длину любого интервала, итерации $R_\alpha^n(I)$ не должны пересекаться. Действительно, в противном случае множество $S^1 \setminus A$ содержало бы более длинный интервал, чем I . Так как число α иррационально, никакие итерации интервала I не могут совпадать, потому что иначе конечная точка x некоторой итерации интервала I перешла бы в себя и мы бы получили $x + k\alpha = x \pmod{1}$, где $k\alpha = l$ — целое число и $\alpha = l/k$ — рациональное число. Итак, все интервалы $R_\alpha^n(I)$ имеют равную

длину и не пересекаются. Но это невозможно, поскольку окружность имеет конечную длину и сумма длин непересекающихся интервалов не может превышать эту длину. $\square$

Предложение 4.1.1 мотивирует следующие общие определения.

Определение 4.1.3. Гомеоморфизм (см. определение A.1.16) $f: X \rightarrow X$ называется *топологически транзитивным*, если существует такая точка $x \in X$, что ее орбита $\mathcal{O}_f(x) := (f^n(x))_{n \in \mathbb{Z}}$ плотна в X . Эквивалентно, можно сказать, что каждое f -инвариантное открытое инвариантное множество плотно. Необратимое отображение f называется *топологически транзитивным*, если существует такая точка $x \in X$, что ее (положительная) орбита $\mathcal{O}_f^+(x) := (f^n(x))_{n \in \mathbb{N}_0}$ плотна в X .

Определения для систем с непрерывным временем вполне аналогичны.

Определение 4.1.4. Гомеоморфизм $f: X \rightarrow X$ называется *минимальным*, если орбита каждой точки $x \in X$ плотна в X , или, что равносильно, если f не имеет собственных замкнутых инвариантных множеств. Замкнутое инвариантное множество называется *минимальным*, если оно не содержит собственных замкнутых инвариантных подмножеств, или, что эквивалентно, если оно является замыканием орбиты любой из своих точек.

Таким образом, предложение 4.1.1 показывает, что любой поворот окружности на угол, несоизмеримый с π , т. е. на иррациональное число градусов (мы будем называть такой поворот просто *иррациональным поворотом*), минимален и, следовательно, топологически транзитивен.

В то время как из минимальности всегда следует топологическая транзитивность, обратное неверно. В гл. 7 мы встретимся с различными примерами, показывающими, что топологическая транзитивность (существование плотных орбит) может иметь место одновременно с существованием многих орбит весьма различных видов, например бесконечного числа периодических (конечных) орбит, объединение которых, в свою очередь, плотно.

4.1.3. Плотные орбиты. Было бы интересно понять, каким образом орбита плотно заполняет окружность. Сделаем это для конкретного примера, исследуя орбиту точки 0 при повороте R_α , где

$$\alpha = \frac{1}{3 + \frac{1}{5 + \frac{1}{c}}}$$

для некоторого $c > 1$. Число α рационально тогда и только тогда, когда c рационально. В конце этого пункта станет понятно, почему мы представили число α в такой необычной форме.

Поскольку $1/4 < \alpha < 1/3$ и, следовательно, $3\alpha < 1 < 4\alpha$, орбита точки 0 впервые подходит к точке 0 ближе, чем при всех предшествующих итерациях, после трех шагов. Первые три точки α , 2α , и 3α находятся на одинаковом расстоянии

друг от друга, и, поскольку $4\alpha > 1$, точка 3α находится ближе к целому числу, чем предыдущие точки. Это расстояние равно

$$\delta := 1 - 3\alpha = 1 - \frac{3}{3 + \frac{1}{5 + \frac{1}{c}}} = \frac{1}{5 + \frac{1}{c}} = \frac{1}{3 + \frac{1}{5 + \frac{1}{c}}} = \frac{1}{16 + \frac{3}{c}}.$$

Чтобы определить следующий момент, когда орбита подойдет к нулю ближе, чем при всех предыдущих итерациях, начнем с четвертого шага и используем соотношение $4\alpha = \alpha - \delta \pmod{1}$. Итак, после трех α -шагов мы перешли от α к $\alpha - \delta$. Сколько таких троек α -шагов потребуется, чтобы достичь следующего момента ближайшего подхода орбиты к 0? Как и прежде, это число n должно быть порядка α/δ , точнее, для него справедлива оценка $n\delta < \alpha < (n+1)\delta$. Нам подойдет $n = 5$. Действительно,

$$5\delta = \frac{5}{15 + \left(1 + \frac{3}{c}\right)} = \frac{1}{3 + \left(\frac{1}{5} + \frac{3}{5c}\right)} < \frac{1}{3 + \frac{1}{5}} < \frac{1}{3 + \frac{1}{5 + \frac{1}{c}}} = \alpha$$

и

$$6\delta = \frac{6}{16 + \frac{3}{c}} > \frac{6}{18} = \frac{1}{3} > \alpha.$$

Эти пять троек α -шагов равномерно заполняют интервал $(0, \alpha)$ и одновременно три интервала, являющиеся его образами. При достижении следующей ближайшей к 0 точки построенный отрезок орбиты будет представлять собой δ -плотное подмножество окружности, состоящее из точек, находящихся на одинаковом расстоянии друг от друга (за исключением маленького интервала для новой ближайшей к 0 точки). Следующая ближайшая к 0 точка определяется величиной c , и мы можем утверждать, что она будет достигнута примерно через c шагов.

Если выбрать число c порядка миллиарда, то, чтобы попасть в следующую ближайшую к 0 точку, нам потребуется около миллиарда пятерок δ -шагов, т. е. около 15 миллиардов итераций поворота R_α . В частности, для первых 7 миллиардов итераций будет всегда оставаться промежуток больше чем $\delta/2 > 1/35$. Таким образом, большие члены при представлении в виде *цепной дроби* не очень хороши для равномерного заполнения окружности. Цепные дроби обсуждаются более подробно в § 15.2.

Итак, имеется естественная последовательность дальнейших более длинных интервалов времени, в течение каждого из которых орбита достигает более высокой степени плотности довольно однородным образом. Значит, поведение орбиты периодическое, за исключением небольшой погрешности δ , которая производит возмущение с намного более высоким периодом, приводя к еще меньшей погрешности, и т. д.

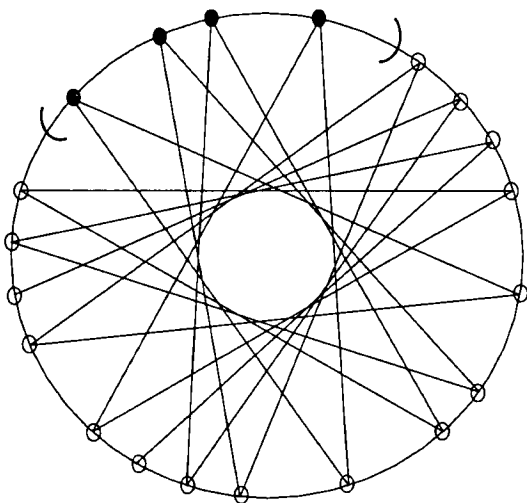


Рис. 4.1.2. Частоты

4.1.4. Равномерное распределение для интервалов. Соображения, изложенные в предыдущем пункте, предлагают взглянуть на то, как орбиты иррационального поворота распределены на окружности, с количественной точки зрения. Для этой цели мы изучаем *частоты*, с которыми итерации точки появляются на различных частях окружности. Для определенности зафиксируем дугу $\Delta \subset S^1$ и для $x \in S^1$ и $n \in \mathbb{N}$ положим

$$F_{\Delta}(x, n) := \text{card}\{k \in \mathbb{Z}: 0 \leq k < n, \mathbb{R}_{\alpha}^k(x) \in \Delta\}.$$

Заметим, что эта функция не убывает по n для фиксированных x и Δ . Так как положительная полуорбита любой точки плотна, имеется произвольно большое количество положительных итераций точки x , которые принадлежат дуге Δ . Следовательно, мы видим, что

$$F_{\Delta}(x, n) \rightarrow \infty \quad \text{при} \quad n \rightarrow \infty.$$

Мы хотели бы видеть, как часто происходят эти появления. Естественным критерием этого была бы *относительная частота* посещений

$$\frac{F_{\Delta}(x, n)}{n}. \quad (4.1.2)$$

Напомним, что $\ell(\Delta)$ обозначает длину дуги Δ , измеренную параметром ϕ , как показано в начале п. 4.1.1.

Рассуждение из доказательства предложения 4.1.1 приводит к следующему результату.

Предложение 4.1.5. Пусть R_{α} — иррациональный поворот, и пусть Δ, Δ' — две такие дуги, что $\ell(\Delta) < \ell(\Delta')$. Тогда существует такое число $N_0 \in \mathbb{N}$, зависящее от Δ, Δ' и α , что для всех $x \in S^1$, $N \geq N_0$ и $n \in \mathbb{N}$ выполняется неравенство

$$F_{\Delta'}(x, n + N) \geq F_{\Delta}(x, n).$$

Доказательство. Плотность положительной полуорбиты левого конца дуги Δ позволяет найти такое число $N_0 \in \mathbb{N}$, что $R_\alpha^{N_0}(\Delta) \subset \Delta'$. Тогда если $R_\alpha^n(x) \in \Delta$, то $R_\alpha^{n+N_0}(x) \in \Delta'$ и $F_{\Delta'}(x, n+N) \geq F_{\Delta'}(x, n+N_0) \geq F_\Delta(x, n)$ для $N \geq N_0$. $\square$

Пока мы не определили, какие дуги мы рассматриваем: открытые, замкнутые или полукоткрытые. Очевидно, что если рассматривать предельное поведение частот, то это неважно, так как разница между числом появлений на открытой дуге и на ее замыкании не превосходит двух. Так что всегда удобно брать дуги, замкнутые слева и открытые справа. Для таких дуг мы имеем следующее свойство *аддитивности*: если правый конец дуги Δ_1 совпадает с левым концом дуги Δ_2 , тогда $\Delta_1 \cap \Delta_2 = \emptyset$, $\Delta_1 \cup \Delta_2$ образует дугу и

$$F_{\Delta_1}(x, n) + F_{\Delta_2}(x, n) = F_{\Delta_1 \cup \Delta_2}(x, n).$$

Также удобно определить $F_A(x, n) := \text{card}\{k \in \mathbb{Z}: 0 \leq k < n, R_\alpha^k(x) \in A\}$ для любого множества A , которое является объединением непересекающихся дуг. Пока мы не знаем, существуют ли пределы относительных частот. Однако можно рассматривать верхние пределы

$$\bar{f}_x(A) := \overline{\lim}_{n \rightarrow \infty} \frac{F_A(x, n)}{n}.$$

Эти величины, очевидно, *полуаддитивны*, т. е.

$$\bar{f}_x(A_1 \cup A_2) \leq \bar{f}_x(A_1) + \bar{f}_x(A_2).$$

В частности, если $\bigcup_{i=1}^n A_i = S^1$, то $\sum_{i=1}^n \bar{f}_x(A_i) \geq 1$. Из предложения 4.1.5 немедленно выводится такой результат.

Следствие 4.1.6. Если $\ell(\Delta) < \ell(\Delta')$, то $\bar{f}_x(\Delta) \leq \bar{f}_x(\Delta')$.

Подобным образом мы определяем нижние асимптотические частоты

$$\underline{f}_x(A) := \underline{\lim}_{n \rightarrow \infty} \frac{F_A(x, n)}{n}.$$

Очевидно, что для любого множества A мы имеем $F_A(x, n) = n - F_{A^c}(x, n)$, где A^c обозначает дополнение $S^1 \setminus A$ к A , и, следовательно,

$$\bar{f}_x(A) = \overline{\lim}_{n \rightarrow \infty} \frac{F_A(x, n)}{n} = 1 - \underline{\lim}_{n \rightarrow \infty} \frac{F_{A^c}(x, n)}{n} = 1 - \underline{f}_x(A^c). \quad (4.1.3)$$

Теперь мы можем сформулировать наш основной результат, касающийся асимптотических частот.

Предложение 4.1.7. Для любой дуги $\Delta \subset S^1$ и любого $x \in S^1$ выполнено равенство

$$f(\Delta) := \lim_{n \rightarrow \infty} \frac{F_\Delta(x, n)}{n} = \ell(\Delta),$$

причем сходимость равномерна по x .

Замечание 4.1.8. Свойство последовательности $a_n := R_\alpha^n(x)$, $n = 0, 1, 2, \dots$, выраженное этим предложением, называется *равномерным распределением*: асимптотическая частота появлений одна и та же для дуг равной длины, независимо от того, где на окружности они находятся.

Доказательство. Сначала мы покажем, что частота появлений не может быть слишком большой.

Лемма 4.1.9. Если $\ell(\Delta) = 1/k$, то $\bar{f}_x(\Delta) \leq 1/(k-1)$.

Доказательство. Рассмотрим $k-1$ непересекающихся дуг $\Delta_1, \Delta_2, \dots, \Delta_{k-1}$ длины $1/(k-1)$ каждая. Согласно предложению 4.1.5 можно найти такие натуральные числа N_i , что если $x \in S^1$, то

$$F_{\Delta_i}(x, n + N_i) \geq F_{\Delta}(x, n),$$

а следовательно, $F_{\Delta_i}(x, n + N) \geq F_{\Delta}(x, n)$, где $N = \max_i N_i$, и

$$(k-1)F_{\Delta}(x, n) \leq \sum_{i=1}^{k-1} F_{\Delta_i}(x, n + N).$$

Так как число N фиксировано, переходя к пределу при $n \rightarrow \infty$, мы получаем

$$(k-1)\bar{f}_x(\Delta) \leq \bar{f}_x\left(\bigcup_{i=1}^{k-1} \Delta_i\right) = 1.$$

□

Возьмем теперь произвольную дугу Δ , выберем $\epsilon > 0$ и найдем достаточно большое k и дугу $\Delta' \supset \Delta$ длины $l/k < \ell(\Delta) + \epsilon$. По лемме 4.1.9 мы заключаем, что

$$\bar{f}_x(\Delta) < \bar{f}_x(\Delta') < \frac{l}{k-1} < (\ell(\Delta) + \epsilon) \frac{k}{k-1}.$$

Переходя к пределу при $\epsilon \rightarrow 0$ и $k \rightarrow \infty$, мы получаем

$$\bar{f}_x(\Delta) \leq \ell(\Delta).$$

Используя равенство (4.1.3) для $A = \Delta^c$, мы приходим к неравенству

$$\underline{f}_x(\Delta) \geq \ell(\Delta),$$

откуда следует, что предел существует и равен $\ell(\Delta)$.

□

4.1.5. Равномерное распределение для функций. Ясно, что частоты могут быть определены также и для любого множества A , являющегося конечным объединением дуг. Функция

$$\chi_A(x) := \begin{cases} 1, & \text{если } x \in A, \\ 0, & \text{если } x \notin A, \end{cases}$$

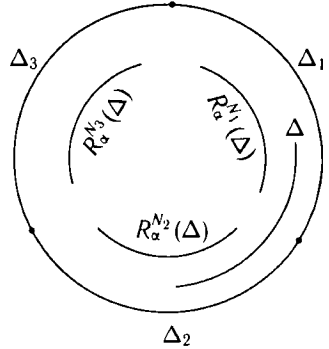


Рис. 4.1.3. Верхние асимптотические частоты

называется *характеристической функцией* множества A . Определим

$$F_A(x, n) := \sum_{k=0}^{n-1} \chi_A(R_\alpha^k(x)).$$

Тогда относительная частота равна $\sum_{k=0}^{n-1} \chi_A(R_\alpha^k(x))/n$. Так как по определению интеграла $\ell(\Delta) = \int_{S^1} \chi_\Delta(\phi) d\phi$, предложение 4.1.7 можно переформулировать следующим образом

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(R_\alpha^k(x)) = \int_{S^1} \chi_\Delta(\phi) d\phi. \quad (4.1.4)$$

1. *Усреднение Биркгофа.* Мы можем рассматривать подобные выражения также и для других функций φ , не обязательно являющихся характеристическими.

Определение 4.1.10. *Оператор усреднения Биркгофа* $\mathcal{B}_n$ — это оператор, который ставит в соответствие функции φ функцию $\mathcal{B}_n(\varphi)$, заданную формулой

$$\mathcal{B}_n(\varphi)(x) := \frac{1}{n} \sum_{k=0}^{n-1} \varphi(R_\alpha^k(x)). \quad (4.1.5)$$

Замечание 4.1.11. Перечислим некоторые полезные свойства оператора $\mathcal{B}_n$:

- 1) оператор $\mathcal{B}_n$ линеен: $\mathcal{B}_n(a\varphi + b\psi) = a\mathcal{B}_n(\varphi) + b\mathcal{B}_n(\psi)$;
- 2) оператор $\mathcal{B}_n$ неотрицателен: если $\varphi \geq 0$, то $\mathcal{B}_n(\varphi) \geq 0$; этот оператор также положителен (или монотонен): если $\varphi > 0$, то $\mathcal{B}_n(\varphi) > 0$;
- 3) оператор $\mathcal{B}_n$ нестягивающий: $\sup_{x \in S^1} \mathcal{B}_n(\varphi)(x) \leq \sup_{x \in S^1} \varphi(x)$;
- 4) оператор $\mathcal{B}_n$ сохраняет среднее: $\int_{S^1} \mathcal{B}_n(\varphi)(\phi) d\phi = \int_{S^1} \varphi(\phi) d\phi$.

Это приводит к следующим заключениям.

Предложение 4.1.12. 1) Для любой ступенчатой функции φ , т. е. для функции, являющейся линейной комбинацией характеристических функций дуг, $\lim_{n \rightarrow \infty} \mathcal{B}_n(\varphi) = \int_{S^1} \varphi(\phi) d\phi$.

2) Для любой функции φ , которая является равномерным пределом ступенчатых функций, также справедливо равенство $\lim_{n \rightarrow \infty} \mathcal{B}_n(\varphi) = \int_{S^1} \varphi(\phi) d\phi$.

Доказательство. Так как отображение, ставящее в соответствие интегрируемой функции интеграл по S^1 , имеет свойства, аналогичные перечисленным в замечании, мы можем начать с соотношения (4.1.4), переходя к линейным комбинациям и равномерным пределам, и сравнить результаты.

Чтобы доказать второе утверждение, зафиксируем $\varepsilon > 0$, возьмем такую ступенчатую функцию φ_ε , что $\sup_{\phi \in S^1} |\varphi(\phi) - \varphi_\varepsilon(\phi)| < \varepsilon$, и применим операторы $\mathcal{B}_n$ к $\varphi = \varphi_\varepsilon + (\varphi - \varphi_\varepsilon)$. В результате мы получим

$$\begin{aligned} \int_{S^1} \varphi(\phi) d\phi - 2\varepsilon &\leq \int_{S^1} (\varphi(\phi) - \varepsilon) d\phi - \varepsilon \leq \int_{S^1} \varphi_\varepsilon(\phi) d\phi - \varepsilon = \\ &= \lim_{n \rightarrow \infty} \mathcal{B}_n(\varphi_\varepsilon) - \varepsilon \leq \lim_{n \rightarrow \infty} \mathcal{B}_n(\varphi) \leq \overline{\lim}_{n \rightarrow \infty} \mathcal{B}_n(\varphi) \leq \lim_{n \rightarrow \infty} \mathcal{B}_n(\varphi_\varepsilon) + \varepsilon = \\ &= \int_{S^1} \varphi_\varepsilon(\phi) d\phi + \varepsilon \leq \int_{S^1} (\varphi(\phi) + \varepsilon) d\phi + \varepsilon \leq \int_{S^1} \varphi(\phi) d\phi + 2\varepsilon \end{aligned} \quad (4.1.6)$$

для любого $\varepsilon > 0$. □

Лемма 4.1.13. *Всякая непрерывная функция является равномерным пределом ступенчатых функций. То же самое справедливо для всякой функции с конечным числом точек разрыва, для которой существуют односторонние пределы в этих точках (т. е. для кусочно непрерывной функции).*

Доказательство. Напомним, что непрерывная на S^1 функция равномерно непрерывна, т. е. для всякого $\varepsilon > 0$ найдется такое $n \in \mathbb{N}$, что на всякой дуге длины $1/n$ изменение функции меньше чем ε . Разделив S^1 на n таких дуг, мы получим ступенчатую функцию, которая постоянна на каждой дуге и отличается от данной функции меньше чем на ε . По существу то же самое рассуждение применяется и для функций с конечным числом точек разрыва, имеющей односторонние пределы в этих точках. $\square$

Из последних двух утверждений вытекает такой результат.

Предложение 4.1.14. Для любого иррационального числа α и любой непрерывной функции φ справедливо равенство

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(R_\alpha^k(x)) = \int_{S^1} \varphi(\phi) d\phi$$

равномерно по x .

Имеется, однако, даже более общий класс функций, для которых среднее Биркгофа сходится к интегралу, а именно все функции, интегрируемые в обычном (римановом) смысле.

Теорема 4.1.15. Если число α иррационально и функция φ интегрируема по Риману, то

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(R_\alpha^k(x)) = \int_{S^1} \varphi(\phi) d\phi \quad (4.1.7)$$

равномерно по x .

Доказательство. Зададим разбиение окружности S^1 на конечное число дуг I_i . Соответствующие нижние и верхние суммы Римана $\sum_i \min \varphi|_{I_i} l(I_i)$ и $\sum_i \max \varphi|_{I_i} l(I_i)$ можно интерпретировать как интегралы ступенчатых функций φ_1 и φ_2 , определенных соотношениями $\varphi_1 = \min \varphi|_{I_i}$ на I_i и $\varphi_2 = \max \varphi|_{I_i}$ на I_i . По определению интегрируемости по Риману разбиение может быть выбрано так, что

$$\int_{S^1} \varphi(\phi) d\phi - \varepsilon \leq \int_{S^1} \varphi_1(\phi) d\phi \leq \int_{S^1} \varphi_2(\phi) d\phi \leq \int_{S^1} \varphi(\phi) d\phi + \varepsilon.$$

Из этого следует, что

$$\begin{aligned} \int_{S^1} \varphi(\phi) d\phi - \varepsilon &\leq \int_{S^1} \varphi_1(\phi) d\phi = \lim_{n \rightarrow \infty} B_n(\varphi_1) \leq \lim_{n \rightarrow \infty} B_n(\varphi) \leq \\ &\leq \overline{\lim}_{n \rightarrow \infty} B_n(\varphi) \leq \lim_{n \rightarrow \infty} B_n(\varphi_2) = \int_{S^1} \varphi_2(\phi) d\phi \leq \int_{S^1} \varphi(\phi) d\phi + \varepsilon. \end{aligned} \quad (4.1.8)$$

Переходя к пределу $\varepsilon \rightarrow 0$, мы получаем требуемое соотношение. $\square$

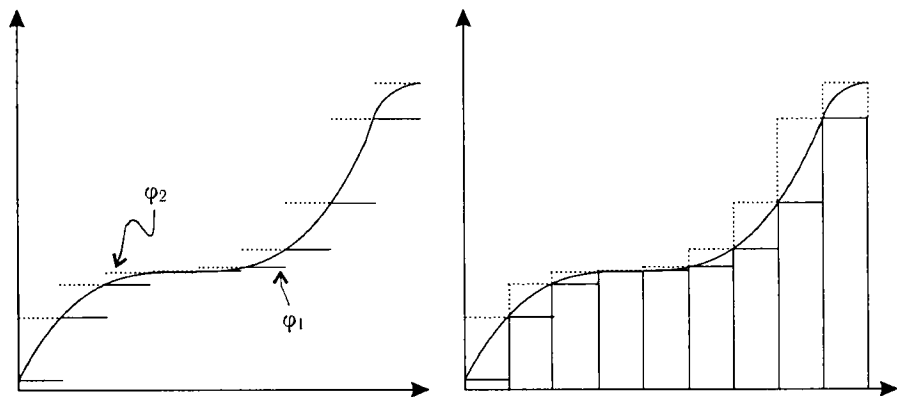


Рис. 4.1.4. Аппроксимация ступенчатыми функциями, суммы Римана

Замечание 4.1.16. Условие интегрируемости по Риману существенно. Чтобы убедиться в этом, выберем точку x_0 и определим множество A как объединение дуг длины 2^{-k+2} с центрами в точках $R_\alpha^k(x_0)$, $k \geq 0$. Некоторые из этих дуг могут пересекаться, но тем не менее множество A представляет собой объединение дуг, суммарная длина которых меньше чем $1/2$, в то время как

$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(R_\alpha^k(x)) = 1$. Конечно же, в этом случае функция χ_A не является интегрируемой по Риману.

2. *Временные средние и пространственные средние.* Величины в левой и правой частях неравенства (4.1.7) являются средними.

Определение 4.1.17. Для данной функции φ мы называем

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(R_\alpha^k(x))$$

временным средним, полученным при движении по орбите точки x при итерациях поворота R_α .

Интеграл $\int_S \varphi(\phi) d\phi$ называется *пространственным средним* функции φ .

Оба эти понятия заимствованы из физики, где важно измерять наблюдаемые величины, связанные с динамической системой. Это означает, что имеется (измеримая) величина, связанная с динамикой рассматриваемой системы, которая изменяется с изменением состояния системы, — другими словами, для системы существует функция, определенная на фазовом пространстве, значение которой в конкретном состоянии динамической системы может быть показано измеряющим устройством. Весьма естественно, особенно для систем, которые ведут себя непредсказуемым образом, брать большое количество последовательных измерений и усреднять их. Предел этих усреднений — в точности временное среднее для начального условия, при котором были начаты измерения.

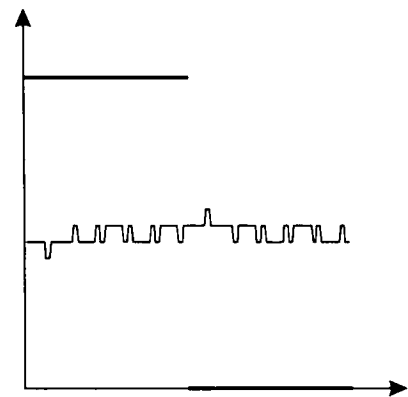


Рис. 4.1.5. Временное среднее

Пространственное среднее может быть получено в результате вычислений, связанных с математической моделью физической системы. Если известно, как мы и предполагаем в нашем простом примере, что пространственное среднее и временное среднее обязаны совпадать для модельной системы, то пространственное среднее позволяет предсказать значение временного среднего, с учетом которого предложенная модель либо принимается, либо отвергается.

Возвращаясь к нашей ситуации, мы замечаем, что, как следует из предыдущего результата, для любой интегрируемой по Риману функции временное среднее существует для орбиты любой точки x и всегда совпадает с пространственным средним. Это важное свойство иррациональных поворотов равносильно равномерному распределению орбит и обычно называется *строгой эргодичностью*. Это понятие может быть определено в абстрактной ситуации непрерывного отображения компактного метрического пространства, даже если не имеется никакого понятия интеграла.

Определение 4.1.18. Пусть X — компактное метрическое пространство. Тогда непрерывное отображение $f: X \rightarrow X$ называется *строгой эргодическим*, если сумма

$$\frac{1}{n} \sum_{k=0}^{n-1} \varphi(f^k(x))$$

сходится к постоянной величине равномерно (по x) для каждой непрерывной функции φ .

4.1.6. Метод Кронекера—Вейля. В наших рассуждениях важную роль играли ступенчатые функции. Вместо этого можно доказывать строгую эргодичность иррационального поворота немногим более коротким, но зато менее элементарным способом, с использованием тригонометрических

многочленов для аппроксимации непрерывных функций. Это возможно в силу классической теоремы Вейерштрасса, гласящей, что непрерывные функции — это равномерные пределы тригонометрических полиномов. Эта теорема является близким аналогом более знакомой теоремы Вейерштрасса, которая имеет дело с равномерным приближением непрерывной на отрезке функции многочленами. В этом рассуждении более удобно использовать комплексные функции.

Другое доказательство предложения 4.1.14. Определим *характеры* $c_m(x) := e^{2\pi i m x} = \cos 2\pi m x + i \sin 2\pi m x$. Для $m \neq 0$ мы имеем

$$c_m(R_\alpha(x)) = e^{2\pi i m(x+\alpha)} = e^{2\pi i m \alpha} e^{2\pi i m x} = e^{2\pi i m \alpha} c_m(x)$$

и

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} c_m(R_\alpha^k(x)) \right| = \left| \frac{1}{n} \sum_{k=0}^{n-1} e^{2\pi i m k \alpha} \right| = \frac{|1 - e^{2\pi i m n \alpha}|}{n|1 - e^{2\pi i m \alpha}|} \leq \frac{2}{n|1 - e^{2\pi i m \alpha}|} \rightarrow 0$$

при $n \rightarrow \infty$, поскольку $\sum_{k=0}^n x^k = \frac{1 - x^{n+1}}{1 - x}$.

Так как операторы усреднения Биркгофа линейны, это показывает, что для тригонометрического многочлена $p(x) = \sum_{i=-l}^l a_i c_i(x)$ предел $\lim_{n \rightarrow \infty} \mathcal{B}_n(p)(x)$ существует, постоянен он равен a_0 . Действительно, эта постоянная должна быть интегралом функции p по S^1 , так как операторы $\mathcal{B}_n$ не изменяют интеграл. Такие же рассуждения, как и выше, позволяют нам перейти к равномерным пределам тригонометрических многочленов, т. е. ко всем непрерывным функциям. $\square$

Это рассуждение более аналитическое и требует гораздо более прямых вычислений, чем доказательство, использующее ступенчатые функции. Заметим, однако, что с его помощью нельзя получить первоначальный результат о равномерном распределении (предложение 4.1.7), так как характеристические функции, очевидно, не являются равномерными пределами тригонометрических полиномов. Чтобы все же получить равномерное распределение для интервалов, можно использовать доказательство теоремы 4.1.15 в обратную сторону, т. е. приблизить χ_A такими непрерывными функциями $\varphi_1 \leq \chi_A \leq \varphi_2$, что $\int (\varphi_2 - \varphi_1) < \varepsilon$ (рис. 4.1.6), и повторить вычисления, сделанные в (4.1.8).

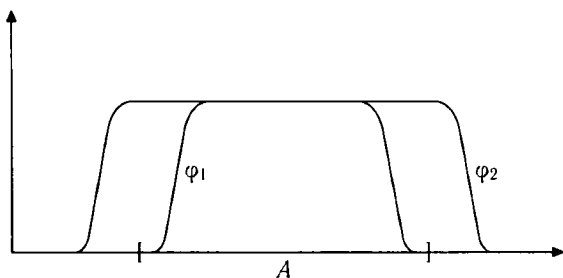


Рис. 4.1.6. Приближение непрерывными функциями

4.1.7. Сдвиги на группах. Иррациональные повороты служат отправной точкой для нескольких полезных обобщений. Рассмотрим одно из них. Окружность — это пример компактной абелевой группы, и поворот может быть представлен в терминах умножения в группе, или сдвига

$$L_{g_0}: G \rightarrow G, \quad L_{g_0} g = g_0 g.$$

Орбита единичного элемента $e \in G$ — циклическая подгруппа $\{g_0^n\}_{n \in \mathbb{Z}}$. Предложение 4.1.1 тесно связано с тем фактом, что окружность не имеет собственных бесконечных замкнутых подгрупп. Чтобы говорить, что орбита плотна, требуется понятие аппроксимации, так что мы определяем *топологическую группу* как группу с метрикой, для которой каждый сдвиг L_g является гомеоморфизмом и операция взятия обратного элемента непрерывна.

Предложение 4.1.19. Если сдвиг L_{g_0} на топологической группе G топологически транзитивен, то он минимален.

Доказательство. Для $g, g' \in G$ обозначим через $A, A' \subset G$ замыкания орбит элементов g и g' соответственно. Тогда $g_0^n g' = g_0^n g(g^{-1}g')$, так что $A' = Ag^{-1}g'$ и $A' = G$ в том и только в том случае, если $A = G$. $\square$

Упражнения

4.1.1. Докажите, что для метрики $d(x, y) := \min\{|b - a| : a \in x, b \in y\}$ на множестве $X = \mathbb{R}/\mathbb{Z} := \{[x] : x \in \mathbb{R}\}$ каждый поворот является изометрией (см. определение A.1.16).

4.1.2. В обозначениях п. 4.1.3 возьмите $c = 7,1$ и определите следующую точку ближайшего подхода к 0.

4.1.3. Докажите свойства из замечания 4.1.11.

4.1.4. Для вращения R_α найдите такое число $N \in \mathbb{N}$, зависящее от α , что $F_{(0,1/2)}(x, n)/n \geq 0,45$ для всех $n \geq N$ (см. п. 4.1.4).

4.1.5. Предположим, что движения Солнца и Луны, наблюдаемые из некоторой точки на Земле, являются строго периодическими и промежутки времени между восходом Солнца и Луны всегда различны. Докажите, что эти промежутки равномерно распределены.

4.1.6. Приведите пример гомеоморфизма полного метрического пространства, который имеет плотную орбиту, но не имеет плотной полуорбиты.

4.1.7. Приведите пример гомеоморфизма компактного метрического пространства, который имеет плотную орбиту, но не имеет плотной полуорбиты.

4.1.8. Докажите, что два минимальных множества (см. определение 4.1.4) либо не пересекаются, либо совпадают.

4.1.9. Докажите, что сжимающее отображение компактного пространства строго эргодично.

4.1.10. Приведите пример такого непрерывного отображения f компактного метрического пространства X , что последовательность

$$\frac{1}{n} \sum_{k=0}^{n-1} \varphi(f^k(x))$$

сходится равномерно (по x) для любой непрерывной функции φ , но отображение f не строго эргодично.

4.1.11. Используя достаточное число разрядов в десятичной записи числа π , найдите классические приближения $22/7$ и $355/113$ и запишите результат в таком виде, как показано в п. 4.1.3. Найдите четвертый знаменатель в разложении π в цепную дробь и объясните, как качество аппроксимации зависит от величины этого числа.

Задачи для углубленного изучения

4.1.12. Пусть G — метризуемая компактная топологическая группа. Предположим, что для некоторого $g_0 \in G$ соответствующий сдвиг L_{g_0} топологически транзитивен. Докажите, что группа G абелева.

4.1.13. Докажите, что конечная абелева группа имеет строго эргодический сдвиг тогда и только тогда, когда она циклическая.

4.1.14. Докажите, что отображение окружности $x \mapsto x + (1/4) \sin^2 x$ (см. рис. 2.2.4) строго эргодично.

4.1.15. Определим следующую метрику d_2 на группе $\mathbb{Z}$ всех целых чисел: $d_2(m, n) = \|m - n\|_2$, где

$$\|n\|_2 = 2^{-k}, \quad \text{если } n = 2^k l, \quad l \text{ нечетно.}$$

Полношение группы $\mathbb{Z}$ относительно этой метрики называется группой *2-адических целых чисел* и обычно обозначается $\mathbb{Z}_2$. Это компактная топологическая группа. Пусть $\mathbb{Z}_2^+$ — это замыкание множества четных целых чисел относительно метрики d_2 . Тогда $\mathbb{Z}_2^+$ является подгруппой группы $\mathbb{Z}_2$ индекса два.

Докажите, что для $g_0 \in \mathbb{Z}_2$ сдвиг $L_{g_0}: \mathbb{Z}_2 \rightarrow \mathbb{Z}_2$ топологически транзитивен тогда и только тогда, когда $g_0 \in \mathbb{Z}_2 \setminus \mathbb{Z}_2^+$.

Это пример класса систем, называемых *одометрами* (adding machines). Эквивалентное описание дается в определении 11.3.10, и теорема 11.3.11 показывает, что эта динамическая система является подсистемой квадратичного отображения $f_\lambda: [0, 1] \rightarrow [0, 1]$, $f_\lambda(x) := \lambda x(1 - x)$, из § 2.5 для некоторого специального значения λ .

§ 4.2. Некоторые приложения плотности и равномерного распределения

Во многих ситуациях информацию асимптотического характера можно получить с использованием (возможно, неявным) динамике поворотов окружности или сдвигов на торе, рассматриваемых в следующей главе, на основе уже изученных нами фактов. В этом параграфе мы укажем некоторые такие примеры.

4.2.1. Распределение значений для периодических функций. Пусть $(x_n)_{n \in \mathbb{N}}$ — последовательность вещественных чисел. Чтобы описать распределение значений такой последовательности, естественно рассматривать асимптотические частоты, с которыми эта последовательность «посещает» различные интервалы.

Определение 4.2.1. Для данной последовательности $(x_n)_{n \in \mathbb{N}}$ и чисел $a < b$ обозначим через $F_{a,b}(n)$ количество таких целых чисел k , что $1 \leq k \leq n$ и $a < x_k < b$. Мы говорим, что последовательность $(x_n)_{n \in \mathbb{N}}$ имеет *асимптотическое распределение*, если для любых a, b , $-\infty \leq a < b \leq \infty$, существует предел

$$\lim_{n \rightarrow \infty} \frac{F_{a,b}(n)}{n}.$$

В этом случае функция

$$\Phi_{(x_n)_{n \in \mathbb{N}}}(t) := \lim_{n \rightarrow \infty} \frac{F_{-\infty, t}(n)}{n}$$

называется *функцией распределения* последовательности.

Для последовательностей вида $y_n = \varphi(x_n)$ мы можем определить функцию распределения в терминах информации относительно функции φ .

Определение 4.2.2. Если множество $A \subset \mathbb{R}$ представляет собой конечное объединение непересекающихся интервалов, тогда мы определяем его *меру* $m(A)$ как сумму длин этих интервалов.

Функция φ , определенная на интервале, называется *кусочно монотонной*, если область ее определения может быть разделена на конечное число интервалов (или отрезков), на каждом из которых φ монотонна. В этом случае прообраз каждого интервала I — конечное объединение интервалов, так что мы можем определить $m_\varphi(I) := m(\varphi^{-1}(I))$. Для кусочно монотонной функции мы определяем ее *функцию распределения* $\Psi_\varphi: \mathbb{R} \rightarrow \mathbb{R}$, полагая $\Psi_\varphi(t) = m_\varphi((-\infty, t))$.

Заметим, что $\Psi_\varphi(t)$ — сумма длин интервалов, составляющих множество $\varphi^{-1}((-\infty, t))$, или, что равносильно, $\Psi_\varphi(t) = \int \chi_{\varphi^{-1}((-\infty, t))}$.

С использованием теоремы 4.1.15 можно показать, что последовательности, полученные вычислением значений периодических функций на арифметических прогрессиях, имеют асимптотическое распределение, и вычислить функции распределения таких последовательностей.

Теорема 4.2.3. Пусть φ — такая функция периода T , определенная на действительной оси, что ее ограничение $\varphi_T = \varphi|_{[0, T]}$ на отрезок $[0, T]$ кусочно монотонно, и пусть $\alpha \notin \mathbb{Q}$. Тогда для каждого вещественного числа t_0 последовательность $x_n := \varphi(t_0 + n\alpha T)$ имеет асимптотическое распределение с функцией распределения Ψ_{φ_T}/T .

Доказательство. Вводя новую переменную $s = t/T$, мы можем предполагать, что $T = 1$. Периодическая функция φ может рассматриваться как функция на окружности, которую мы тоже будем обозначать φ . Тогда последовательность $\{x_n\}$ совпадает с последовательностью значений этой функции на окружности вдоль орбиты иррационального поворота R_α . Применение теоремы 4.1.15 к характеристической функции множества $\varphi^{-1}((a, b))$ показывает, что последовательность $\{x_n\}$ имеет асимптотическое распределение. Действительно, $\varphi(s_0 + n\alpha) \in \varphi^{-1}((a, b))$ тогда и только тогда, когда $a < x_n < b$. Отсюда также следует справедливость утверждения относительно функции распределения. $\square$

Пример 4.2.4. Рассмотрим последовательность $(\sin n)_{n \in \mathbb{N}}$. Так как синус — 2π -периодическая функция и π — иррациональное число, мы можем применить теорему 4.2.3, взяв $\varphi(t) = \sin t$, $T = 2\pi$, $t_0 = 0$, $\alpha = (2\pi)^{-1}$, и тем самым доказать существование асимптотического распределения. В этом случае функция распре-

деления Ψ может быть вычислена в явном виде (см. рис. 4.2.1):

$$\Phi_{(\sin n)_{n \in \mathbb{N}}}(t) = \Psi_{\sin 2\pi}(t)/(2\pi) = \begin{cases} 0, & \text{если } t < -1, \\ \frac{1}{2} + \frac{1}{\pi} \arcsin t, & \text{если } t \in [-1, 1], \\ 1, & \text{если } t > 1. \end{cases} \quad (4.2.1)$$

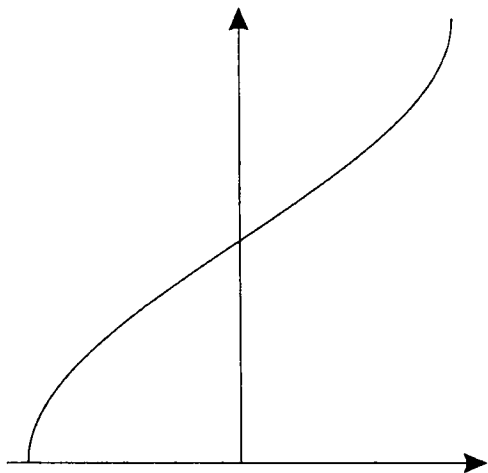


Рис. 4.2.1. Функция распределения

4.2.2. Распределение первых цифр степеней целого числа. Интересное арифметическое приложение теоремы 4.2.3, а точнее предложения 4.1.7, состоит в том, что мы можем теперь ответить на сформулированные в п. 1.3.4 вопросы относительно распределения первых цифр степеней числа 2, а также и степеней любого целого числа. Кроме того, мы сможем ответить на аналогичный вопрос для распределения не одной, а несколько начальных цифр.

Если k — степень числа 10, то первая цифра любой степени числа k равна 1, так что этот случай в дальнейшем рассматриваться не будет. Теперь мы покажем, что в других случаях для любого заданного набора цифр найдется такая степень числа k , что ее десятичная запись будет начинаться с этого набора цифр.

Предложение 4.2.5. Пусть k — натуральное число, не являющееся степенью числа 10, и пусть $r \in \mathbb{N}$. Тогда существует такое число $n \in \mathbb{N}$, что начальные цифры десятичной записи числа k^n составляют число r .

Пример 4.2.6. Для иллюстрации утверждения возьмем $k = 2$ и $r = 81$. Тогда при $n = 13$ мы получим $2^{13} = 8192$.

Доказательство предложения 4.2.5. Утверждение может быть переформулировано следующим образом: существует число $l \in \mathbb{N}$, для которого $k^n = 10^l p + q$, где q — целое число, $0 \leq q < 10^l$. Это, в свою очередь, эквивалентно неравенствам $10^l p \leq k^n < 10^l(p+1)$, или

$$L + \lg p \leq n \lg k < L + \lg(p+1),$$

где $\lg = \log_{10}$ — логарифм по основанию 10. Пусть теперь $m = \lfloor \lg p \rfloor + 1$ — количество разрядов числа p . Тогда

$$0 \leq \lg p - (m-1) \leq n \lg k - l - (m-1) < \lg(p+1) - (m-1) \leq 1,$$

а эти неравенства можно переписать в виде

$$\lg(p/10^{m-1}) \leq \{n \lg k\} \leq \lg((p+1)/10^{m-1}), \quad (4.2.2)$$

где $\{ \cdot \}$ обозначает дробную часть. Так как $\lg k$ — иррациональное число¹, последовательность $(\{n \lg k\})_{n=1}^{\infty}$ в $\mathbb{R}/\mathbb{Z}$ (см. п. 4.1.1 или п. 2.6.2) плотна на окружности (см. предложение 4.1.1), и, следовательно, множество $\{\{n \lg k\} : n \in \mathbb{N}\}$ плотно в полуинтервале $[0, 1)$. В частности оно содержит точки из отрезка $[\lg(p/10^{m-1}), \lg((p+1)/10^{m-1})]$. $\square$

С помощью наших результатов о равномерных распределениях можно не только доказать существование данного набора цифр в первых разрядах, но и найти асимптотические частоты таких наборов.

Предложение 4.2.7. Пусть k — натуральное число, не являющееся степенью числа 10, $p \in \mathbb{N}$ и $F_p^k(n)$ — количество таких целых чисел i между 0 и $n-1$, что p совпадает с начальными разрядами десятичного разложения числа k^i . Тогда

$$\lim_{n \rightarrow \infty} \frac{F_p^k(n)}{n} = \lg(p+1) - \lg p.$$

Доказательство. Применяя предложение 4.1.7 либо теорему 4.2.3 к функции $\varphi(t) := \{t\}$ и учитывая условие (4.2.2), мы видим, что $\lim_{n \rightarrow \infty} \frac{F_p^k(n)}{n} = \lg((p+1)/10^{m-1}) - \lg(p/10^{m-1}) = \lg(p+1) - \lg p$. $\square$

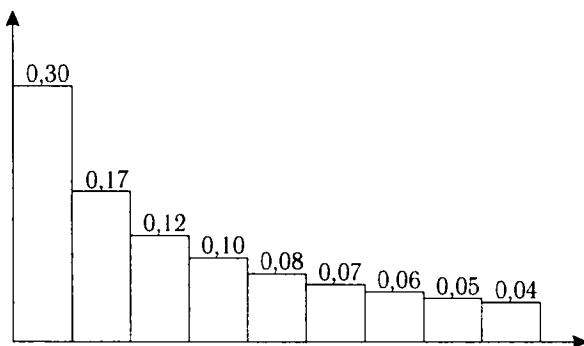


Рис. 4.2.2. Распределение первых цифр числа 2^n

Рис. 4.2.2 изображает приблизительные значения асимптотических частот цифр в первых разрядах степеней числа 2 (или 3, или 7 и т. д.), вопрос о которых был поставлен в п. 1.3.4.

¹Если $\lg k = p/q$, то, используя разложение на простые множители, можно записать $2^p 5^q = 10^p = k^q = 2^{mq} 5^{nq}$. Тогда $n = m$ и $k = 10^m$.

4.2.3. Линейный поток на двумерном торе. Некоторые дальнейшие приложения равномерного распределения появляются в связи с потоками, связанным с поворотами окружности и сдвигами на торе. В частности, наш анализ иррациональных поворотов имеет непосредственные следствия для следующей системы дифференциальных уравнений на двумерном торе (мы используем аддитивную форму записи; см. п. 2.6.4 выше):

$$\frac{dx_1}{dt} = \omega_1, \quad \frac{dx_2}{dt} = \omega_2. \quad (4.2.3)$$

Эта система дифференциальных уравнений может быть легко решена в явном виде. Получающийся в результате поток $(T_\omega^t)_{t \in \mathbb{R}}$ имеет вид

$$T_\omega^t(x_1, x_2) = (x_1 + \omega_1 t, x_2 + \omega_2 t) \pmod{1}. \quad (4.2.4)$$

Этот поток изображен на рис. 4.2.3. Как мы уже упоминали в п. 2.6.4, тор $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ может быть представлен как единичный квадрат $I^2 = \{(x_1, x_2) : 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1\}$, в котором отождествлены пары противоположных сторон: $(x, 0) \sim (x, 1)$ и $(0, x) \sim (1, x)$. В этом представлении интегральные кривые системы (4.2.3) представляют собой участки прямых линий с угловым коэффициентом $\gamma = \omega_2/\omega_1$. Движение по орбитам равномерно с мгновенными «скачками» к соответствующим точкам, когда орбита достигает границы квадрата.

Предложение 4.2.8. *Если число γ иррационально, то всякая орбита плотна в $\mathbb{T}^2$ (т. е. поток минимален). Если число γ рационально, то всякая орбита замкнута.*

Доказательство. Окружность $C_1 = \{x_1 = 0\}$ определяет так называемое глобальное сечение. Это означает, что все положительные и отрицательные полуорбиты пересекают эту окружность бесконечно часто. Таким образом, мы можем

определить *отображение возвращения* на C_1 , поставив в соответствие каждой точке окружности C_1 точку, в которой положительная полуорбита впервые возвращается на C_1 .

Это отображение — в точности поворот R_γ , потому что между последовательными моментами времени, в которые орбита пересекает C_1 , координата x_2 изменяется в точности на $\gamma \pmod{1}$. Таким образом, согласно предложению 4.1.1, если число γ иррационально, замыкание каждой орбиты содержит окружность C_1 , и, так как образы этой окружности при действии потока $\{T_\omega^t\}$ покрывают весь тор, всякая орбита плотна в $\mathbb{T}^2$. Таким образом, поток минимален (см. определение 4.1.4; для потоков определение минимальности аналогично). Если число γ рационально, то каждая орбита замкнута в соответствии с соотношением (4.2.4). $\square$

4.2.4. Линейные дифференциальные уравнения и фигуры Лиссажу. Потоки на двумерном торе появляются естественным образом во многих задачах теории обыкновенных дифференциальных уравнений. Часто надлежащий выбор координат приводит систему дифференциальных уравнений к стандартной форме (4.2.3). Наиболее тесно связаны с потоками некоторые системы линейных дифференциальных уравнений с постоянными коэффициентами. Пусть A — вещественная матрица размера 4×4 с двумя парами различных чисто мнимых собственных значений $\pm i\alpha_1$, $\pm i\alpha_2$. Рассмотрим систему обыкновенных дифференциальных уравнений

$$\frac{dx}{dt} = Ax. \quad (4.2.5)$$

Заменой координат матрица A приводится к виду

$$\begin{pmatrix} 0 & \alpha_1 & 0 & 0 \\ -\alpha_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_2 \\ 0 & 0 & -\alpha_2 & 0 \end{pmatrix}.$$

Эта система является *линеаризацией* (см. п. 6.2.2.7) сферического маятника, который будет рассмотрен в п. 6.2.7, т. е. она отражает поведение сферического маятника при малых колебаниях.

Теперь положим $x = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4$ и рассмотрим две квадратичные функции $x_1^2 + x_2^2$ и $x_3^2 + x_4^2$. Прямое вычисление производных по времени этих функций показывает, что обе они инвариантны относительно системы (4.2.5). Таким образом, для любых двух положительных чисел r_1, r_2 тор T_{r_1, r_2} , определенный уравнениями $x_1^2 + x_2^2 = r_1^2$, $x_3^2 + x_4^2 = r_2^2$, инвариантен. Теперь введем естественные нормализованные угловые координаты φ_1, φ_2 на торе, определяя $x_1 = r_1 \cos 2\pi\varphi_1$, $x_2 = r_1 \sin 2\pi\varphi_1$, $x_3 = r_2 \cos 2\pi\varphi_2$, $x_4 = r_2 \sin 2\pi\varphi_2$. Положим $\alpha_1/2\pi = \omega_1$, $\alpha_2/2\pi = \omega_2$. В этих координатах система (4.2.5) записывается следующим образом:

$$\frac{d\varphi_1}{dt} = -\omega_1, \quad \frac{d\varphi_2}{dt} = -\omega_2,$$

Рис. 4.2.4. Вложенные торы

а ее решения имеют вид

$$\begin{aligned}x_1(t) &= r_1 \cos \omega_1(t - t_0), & x_2(t) &= r_1 \sin \omega_1(t - t_0), \\x_3(t) &= r_2 \cos \omega_2(t - t_0), & x_4(t) &= r_2 \sin \omega_2(t - t_0).\end{aligned}\quad (4.2.6)$$

Таким образом, если отношение $\frac{\alpha_2}{\alpha_1} = \frac{\omega_2}{\omega_1}$ иррационально, то на каждом торе T_{r_1, r_2} поток, определенный системой (4.2.5), минимален.

Теперь рассмотрим проекцию решений системы (4.2.6) на плоскость x_1, x_3 . Получающиеся кривые называются *фигурами Лиссажу*. Из системы (4.2.6) мы получаем следующее параметрическое представление фигуры Лиссажу (мы используем координаты x и y вместо x_1 и x_3):

$$x(t) = r_1 \cos \omega_1(t - t_0), \quad y(t) = r_2 \cos \omega_2(t - t_0).$$

Простая физическая интерпретация фигуры Лиссажу такова: это конфигурационная траектория $x(t), y(t)$ пары *независимых линейных* или *гармонических осцилляторов* (см. пример 6.2.2 из п. 6.2.2), заданных дифференциальными уравнениями второго порядка

$$\ddot{x} = \omega_1 x, \quad \ddot{y} = \omega_2 y.$$

Для соизмеримых частот ω_1 и ω_2 фигуры Лиссажу — семейства замкнутых кривых, обычно самопересекающихся и имеющих довольно сложную форму (рис. 4.2.5). Но если частоты несоизмеримы, то любая фигура Лиссажу плотно заполняет прямоугольник $|x| \leq r_1, |y| \leq r_2$. Кроме того, используя соответствующую версию теоремы 4.2.3, можно доказать существование *предельной плотности* фигуры Лиссажу и явно ее вычислить.

Мы можем легко получить физическую модель, если подсоединим простые осциллирующие цепи ко входам x и y осциллографа. Осуществляя незначительную

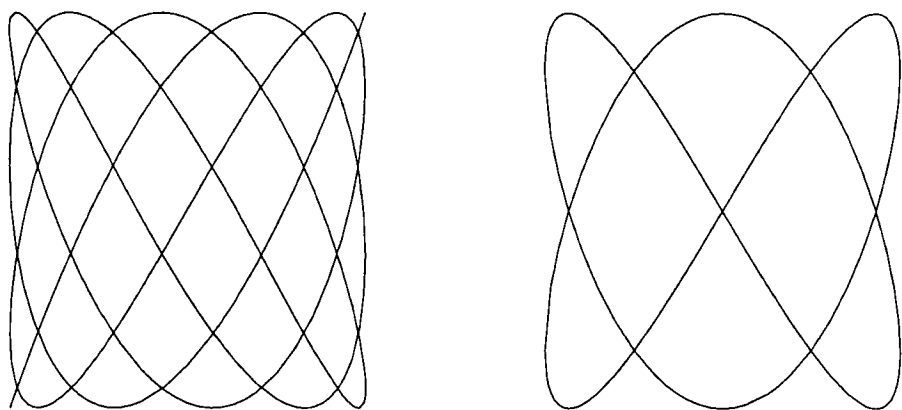


Рис. 4.2.5. Фигуры Лиссажу

настройку, можно легко воспроизводить различные фигуры Лиссажу. Если при настройке мы удаляемся от соразмерных частот, картинка медленно растягивается. При этом возникает кривая, изображенная на рис. 4.2.5 справа, подобная кривым, получающимся при проектировании с медленно вращающегося цилиндра².

4.2.5. Материальные точки на отрезке и бильярды. Имеются различные задачи механики, анализ которых может быть сведен к анализу поворотов окружности и линейных потоков на $\mathbb{T}^2$. В этом пункте мы подробно рассмотрим один простой пример.

1. Материальные точки на отрезке. Рассмотрим две точечные массы, расположенные на отрезке, которые сталкиваются упруго с концами отрезка и друг с другом. Мы будем рассматривать стандартный единичный отрезок $[0, 1]$ и полагать, что обе массы равны единице.

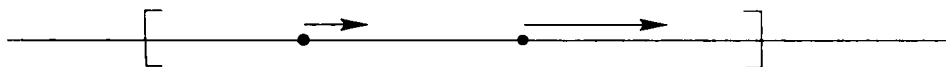


Рис. 4.2.6. Две материальные точки на отрезке

Конфигурационное пространство этой системы — двумерное множество $T = \{(x_1, x_2): 0 \leq x_1 \leq x_2 \leq 1\}$, которое является прямоугольным равнобедренным треугольником (см. рис. 4.2.7). Движение обеих масс может быть описано как движение точки внутри множества T .

Следующие несколько страниц посвящены изучению этой системы. Мы докажем следующий результат.

Предложение 4.2.9. *Для любого начального условия в системе двух материальных точек на отрезке отношение скоростей может принимать*

²Детальное описание можно найти в книге: В. И. Арнольд. Математические методы в классической механике. М.: УРСС, 2000, § 2.5.

с течением времени только восемь различных значений. Если начальное отношение рационально, тогда движение периодическое; если же начальное отношение иррационально, то траектория в конечном счете подойдет произвольно близко к любой конфигурации и движение будет обладать свойством равномерного распределения.

Вспомогательные результаты, используемые для доказательства этого предложения, являются богатым источником идей и для решения других вопросов, касающихся этой системы, а также для изучения других моделей. Например, впоследствии из геометрических соображений станет очевидно, что случай неравных масс очень существенным образом отличается от рассматриваемого сейчас случая, когда массы равны.

2. Биллиардный поток. Между столкновениями скорости частиц постоянны, т. е. «конфигурационная» точка в множестве T , которая представляет обе эти материальные точки, имеет постоянные вертикальную и горизонтальную скорости. Таким образом, конфигурационная точка перемещается по прямой с постоянной скоростью, и, когда она достигает горизонтальной или вертикальной границ, которые соответствуют столкновению с одной из концевых точек отрезка, нормальная компонента вектора скорости меняет знак. При столкновении с гипотенузой компоненты скорости меняются местами (при упругом столкновении две материальных точки обмениваются импульсами, т. е., поскольку их массы равны, скоростями). Таким образом, во всех трех случаях столкновения конфигурационная точка (x_1, x_2) отражается от границы согласно правилу «угол падения равен углу отражения». Другими словами, движение конфигурационной точки подобно движению маленького биллиардного шара внутри треугольного стола T . Поэтому динамическая система с непрерывным временем, описывающая это движение, обычно называется *биллиардным потоком*.

Рассмотрим биллиардный поток более подробно. Сначала мы должны описать фазовое пространство. Очевидно, что оно состоит из касательных векторов с основаниями в T . Нам необходимо принять некоторые соглашения относительно векторов на границе. В силу нашего предположения о мгновенном изменении скорости при столкновении мы можем рассматривать только исходящие скорости. Другими словами, мы предполагаем, что только те векторы с началом на границе, которые лежат внутри треугольника, принадлежат фазовому пространству. Углы соответствуют одновременным столкновениям, и нужно посмотреть, можно ли определить дальнейшее движение для конфигурационной точки, которая попадает в угол. Точка $(0, 1)$, т. е. вершина прямого угла, соответствует одновременному столкновению двух материальных точек с разными концами отрезка, и ее рассмотрение не вызывает никаких сложностей; каждая материальная точка изменяет знак скорости, так что вектор скорости просто меняет знак. Таким

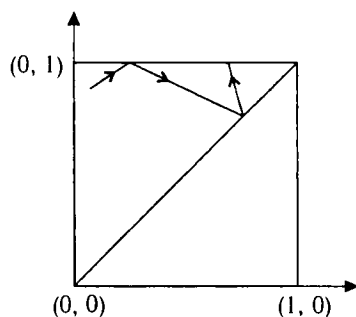


Рис. 4.2.7. Конфигурационное пространство

образом, в этом случае исходящий вектор скорости также направлен внутрь множества T . Случаи двух других вершин менее тривиальны. Они соответствуют одновременному столкновению материальных точек с одним и тем же концом отрезка. Чтобы понять, как после этого будет продолжаться движение, можно аппроксимировать это событие рядом простых столкновений и рассмотреть их предел, если таковой существует. Оказывается, такой ряд всегда содержит четыре столкновения, и предел существует. Однако вместо того, чтобы проводить столь трудоемкие выкладки, мы пока не будем рассматривать эти две вершины, а позже получим интересующую нас картину движения как побочный продукт геометрических рассуждений, которые помогут нам лучше понять, как происходит движение.

Таким образом, билиардный поток — это динамическая система с непрерывным временем в описанном выше пространстве касательных векторов. Вектор движется со скоростью, равной его длине, по прямой, содержащей его, и отражается от сторон. Достаточно очевидно, что можно рассматривать только векторы фиксированной длины, скажем единичной.

3. Развертка. Представим теперь, что при достижении точкой границы вместо отражения вектора скорости мы отражаем область T и рассматриваем движение с той же самой скоростью внутри отраженной копии (см. рис. 4.2.8). Эту операцию можно повторить снова, когда точка достигнет границы (отраженной копии), и т. д. Таким образом, вместо сложной ломаной линии внутри T мы получаем движение по прямой линии, которое переходит из одной копии треугольника T в другую. Естественно, первоначальное движение может быть восстановлено, если свернуть эти копии к исходному треугольнику. Различные начальные векторы ведут к различным последовательностям отражений и порождают различные ряды копий треугольника T вдоль разворачивающейся траектории. Поэтому мы должны попытаться рассмотреть все возможные последовательности отражений одновременно. Оказывается, для более общих областей, чем наш треугольник T , результат, вообще говоря, довольно беспорядочен

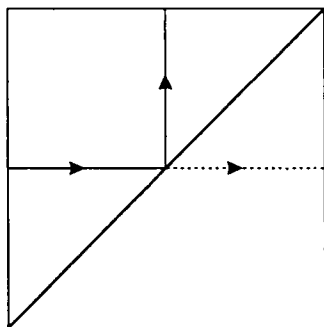


Рис. 4.2.8. Развертка треугольника

и копии исходной области могут пересекаться, но не совпадать с ней. Однако в нашем случае простое наблюдение показывает, что если копия пересекается с внутренностью исходной области, то она совпадает с ней. Таким образом, мы получаем *замощение* плоскости $\mathbb{R}^2$ копиями треугольника T , полученными при различных композициях отражений. Не все такие композиции сами являются отражениями. Например, композиция отражения относительно вертикальной стороны и отражения относительно горизонтальной стороны образа представляет собой вращение на угол π вокруг вершины прямого угла. И вновь тщательное наблюдение показывает, что элементы замощения могут иметь восемь различных ориентаций (см. рис. 4.2.9). Все треугольники в каждом из восьми классов по-

Рис. 4.2.9. Полная развертка

лучены друг из друга *параллельными переносами*. В качестве представителей каждого из восьми классов можно взять первоначальный треугольник T , его отражение относительно гипотенузы и треугольники, полученные отражениями единичного квадрата относительно координатных осей, а также их композиции (они коммутируют). Другими словами, восемь представителей точно заполняют квадрат $S = \{(x, y) : \max\{|x|, |y|\} \leq 1\}$, и любые другие элементы замощения могут быть получены параллельным переносом одного из этих восьми треугольников. Параллельные переносы, которые появляются в процессе, являются композициями итераций вертикальных и горизонтальных переносов на 2. Мы называем S *фундаментальной областью* для группы переносов замощения.

4. Факторпространство. Теперь заменим полную развертку орбиты билиардного потока на $\mathbb{R}^2$, описанного выше, частичной разверткой, заставляя ее оставаться в пределах квадрата S . Чтобы добиться этого, каждый раз, когда развернутая орбита достигает границы этого квадрата, вместо перемещения дальше к треугольнику T_1 мы будем двигаться в единственный треугольник T_2 внутри S , который получен из T_1 параллельным переносом (см. рис. 4.2.10). Это создает разрывы в орбите, если ее рассматривать на плоскости, а именно: перенос на вектор $(0, -2)$ для верхней стороны, $(0, 2)$ для нижней, $(2, 0)$ для левой и $(-2, 0)$ для правой. Но это в точности отождествления, описанные в п. 4.2.3 для линейного потока на торе, хотя и удвоенного размера. Заметим, что линейный поток, полученный таким образом, зависит от начального направления орбиты билиардного потока в T . Так, для исходных направлений с рациональными угловыми коэффициентами все орбиты замкнуты, а для иррациональных угловых коэффициентов они являются плотными и равномерно распределенными на торе.

Свернутая орбита в T состоит из частей, имеющих не более восьми различных направлений. Эти направления получаются из исходного с помощью композиций отражений, которые переводят исходный треугольник в восемь различных треугольников внутри квадрата S . Среди этих композиций есть четыре отражения: относительно горизонтального, вертикального и двух диагональных направлений; остальные преобразования — это повороты на углы, кратные $\pi/2$. Таким образом,

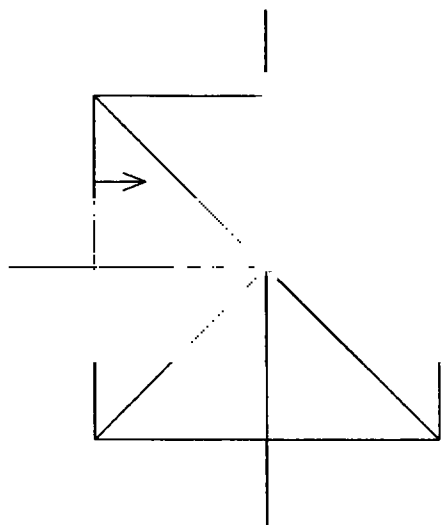


Рис. 4.2.10. Частичная развертка

если первоначальный угол с положительным горизонтальным направлением был равен α , допустимые углы для отрезков орбиты — это $\pm\alpha + \frac{k\pi}{2}$, $k = 0, 1, 2, 3$ (см. рис. 4.2.11).

Из плотности (соответственно периодичности) орбит на торе для начальных направлений с иррациональным (соответственно рациональным) коэффициентом наклона следует плотность (соответственно периодичность) билиардных орбит в треугольнике. То же применимо и к равномерному распределению орбит, для которых начальные направления имеют иррациональные угловые коэффициенты.

Заметим, что в ходе нашего анализа мы решили задачу продолжения движения для орбит, попадающих в угол. На торе такое продолжение вполне однозначно, и после свертки мы получаем точное описание: орбита, приходящая в начало координат под углом α с положительным направлением действительной оси,

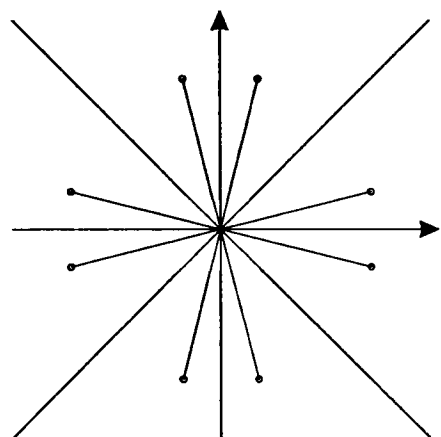


Рис. 4.2.11. Восемь допустимых углов

выходит из него под углом $\pi - \alpha$, т. е. движение будет осуществляться по той же траектории в обратном направлении, так же как и в случае, когда траектория исходила из вершины $(0, 1)$.

Переформулируем теперь наше заключение на языке первоначальной механической задачи. Угловой коэффициент фазового вектора в бильярде равен отношению скоростей двух материальных точек на отрезке. Таким образом, мы можем заключить, что начиная с любого данного отношения в процессе движения могут быть достигнуты только восемь различных значений этого отношения. Если начальное отношение рационально, тогда движение периодическое; если же оно иррационально, траектория в конечном счете подойдет произвольно близко к любой заранее заданной конфигурации. Кроме того, в последнем случае имеется естественное свойство равномерного распределения. Это, в частности, доказывает предложение 4.2.9.

5. *Случай неравных масс.* Имеются естественные обобщения, относящиеся к двум различным аспектам этой задачи, механическому и геометрическому. В исходной физической задаче массы материальных точек могут не быть равными. При столкновении изменение в скоростях обеих материальных точек обратно пропорционально их массам m_1 и m_2 . При переходе к новым координатам $q_1 = \sqrt{m_1}x_1$, $q_2 = \sqrt{m_2}x_2$ закон сохранения импульса и энергии превратится в закон бильярда на бильярдном столе $\{(q_1, q_2): 0 \leq q_1 \leq \sqrt{m_1/m_2}q_2 \leq \sqrt{m_1}\}$ (см. рис. 4.2.12). Если мы попробуем применить метод развертки для изучения бильярдного потока в этом треугольнике, то столкнемся с уже упомянутыми трудностями, состоящими в том, что развертки для различных последовательностей отражений могут пересекаться, не образуя какую-либо распознаваемую структуру. Это происходит неизбежно, если только не выполнено одно из условий $m_1 = m_2$, $m_1 = 3m_2$ или $m_2 = 3m_1$. Последние два случая соответствуют бильярду в прямоугольном треугольнике с наименьшим углом $\pi/6$.

Некоторые другие случаи также можно проанализировать с использованием аналогичных идей и их обобщений. Первый нетривиальный случай — это случай прямоугольного треугольника с острым углом $\pi/8$, которому соответствует отношение масс $3 + 4\sqrt{2}$. Подробнее об этом см. § 14.4 книги А. Катка и Б. Хаселблата «Введение в современную теорию динамических систем».

Теперь обратимся к геометрическому аспекту нашей задачи.

6. *Бильярд в многоугольнике.* Сейчас мы перечислим те многоугольники, для которых также можно применять наш подход к решению задачи для случая равных масс.

Рассмотрим бильярдные потоки внутри многоугольника P , обладающего тем свойством, что его образы при композициях отражений относительно его сторон

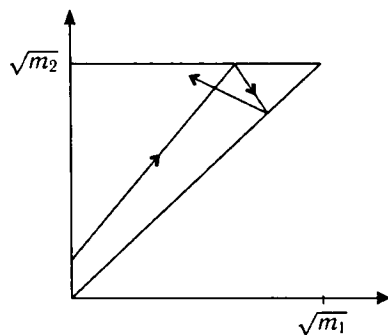


Рис. 4.2.12. Две неравные массы

замощают плоскость. Для таких многоугольников возможно развернуть одновременно все орбиты билиардного потока, подобно тому как мы это сделали для прямоугольного равнобедренного треугольника. После этого мы должны найти многоугольник, который служит фундаментальной областью группы переносов замощения, чтобы заменить полную развертку в бесконечную плоскость частичной, ограниченной компактным множеством.

Список многоугольников, удовлетворяющих этому свойству, достаточно короткий. Кроме прямоугольного равнобедренного треугольника он включает прямоугольники, равносторонний треугольник и прямоугольный треугольник с углом $\pi/6$. Чтобы найти фундаментальную область для этих случаев, мы должны рассмотреть в каждом замощении классы областей, которые можно получить друг из друга параллельным переносом, и затем попытаться выбрать по одному представителю от каждого класса. Результаты представлены на рис. 4.2.13.

Рис. 4.2.13. Замощения

Имеются четыре класса в первом случае, шесть в втором и двенадцать в последнем. Естественные фундаментальные области — увеличенный вдвое прямоугольник R в первом случае и правильные шестиугольники в двух других (рис. 4.2.14).

Рис. 4.2.14. Фундаментальные области

Будем изучать эти системы по отдельности. Наш анализ билиардного потока в треугольнике T переносится на случай прямоугольника в сущности дословно с небольшим упрощением. Действительно, двойной прямоугольник R естественно

отождествляется с тором таким образом, что пары противоположных сторон отождествляются посредством соответствующих параллельных переносов, и орбита бильярда в первоначальном прямоугольнике разворачивается в линейный поток на этом торе.

Два других случая могут быть также сведены к линейному потоку на торе, поскольку тор можно получить отождествлением пар противоположных сторон посредством переноса из любого параллелограмма, а не только из прямоугольника. Для этого только следует произвести надлежащую линейную замену координат на плоскости. Заметим, что в этих двух случаях можно построить фундаментальную область группы параллельных переносов замощения, которая имеет форму параллелограмма с углом $\pi/3$ (рис. 4.2.15). Это показывает, что бильярдный поток в треугольниках сводится к линейным потокам на соответствующих областях.

Рис. 4.2.15. Параллелограмм

Замечание 4.2.10. Классический пример линейного потока на двумерном торе, возникающий в связи с изучением нелинейных систем дифференциальных уравнений, — это задача Кеплера о материальной точке в гравитационном поле, сила притяжения в котором обратно пропорциональна квадрату расстояния (см. п. 6.2.5). Такая материальная точка имеет три степени свободы, но закон сохранения кинетического момента (см. п. 6.2.7) позволяет ограничиться рассмотрением точки на плоскости, тем самым сводя число степеней свободы к двум. Поскольку движение описывается уравнением второго порядка, фазовое пространство четырехмерно. Закон сохранения энергии позволяет понизить размерность фазового пространства до трех, а закон сохранения величины кинетического момента — до двух. Эти две размерности можно параметризовать посредством параметра времени вдоль эллипса и угла перигелия. Таким образом, рассматриваемая система подобна математическому маятнику (см. п. 6.2.2), для которого мы получаем потоки на окружностях, т. е. одномерные инвариантные торы. Однако задача Кеплера для нескольких планет *без взаимодействия* приводит к рассмотрению линейных потоков на инвариантных торах более высокой размерности. В этом состоит основное свойство *вполне интегрируемых* гамильтоновых динамических систем.

Упражнения

4.2.1. Приведите детальное доказательство равенства (4.2.1).

4.2.2. Докажите непосредственно, что для любого фиксированного m сумма выражений вида $\lg(p+1) - \lg p$ по всем числам p , имеющим в точности m десятичных разрядов, равна 1, как это и следует из предложения 4.2.7.

4.2.3. Проверьте справедливость выкладок, с помощью которых предложение 4.2.7 выводится из предложения 4.1.7 или теоремы 4.2.3.

4.2.4. С помощью предложения 4.2.7 найдите $\lim_{n \rightarrow \infty} F_{10}^2(n)/n$ и определите асимптотические частоты, с которыми 0 и 9 являются *вторыми* цифрами степеней числа 2.

4.2.5. В обозначениях из доказательства предложения 4.2.8 предположите, что $\gamma \neq 0$, и замените сечение C_1 сечением $C_2 := \{x_2 = 0\}$. Докажите, что полученное в результате отображение является вращением, и выразите угол поворота через γ .

4.2.6. Проверьте с помощью непосредственного вычисления производных по времени, что функции $x_1^2 + x_2^2$ и $x_3^2 + x_4^2$ инвариантны относительно системы (4.2.5).

4.2.7. Сформулируйте естественное свойство равномерного распределения, которое упоминается в предложении 4.2.9 и доказывается в п. 4.2.5.4.

4.2.8. Докажите, что любая замкнутая собственная подгруппа Γ в $\mathbb{R}$ циклическая, т. е. $\Gamma = \{na\}_{n \in \mathbb{Z}}$ для некоторого $a \in \mathbb{R}$.

4.2.9. Сколько возможных углов с данным начальным направлением может образовывать билиардный поток на квадрате и на каждом из двух треугольников, и каковы эти углы?

4.2.10. Предположим, что луч света попадает в круглую комнату с зеркальными стенами. Опишите в каких случаях какие из частей комнаты будут освещены наилучшим образом.

4.2.11. Докажите, что полная развертка правильного пятиугольника покрывает каждую точку плоскости бесконечно много раз.

4.2.12. Выведите правило продолжения орбит в билиардном описании системы из двух материальных точек, рассматривая двойные столкновения как пределы последовательности простых столкновений.

§ 4.3. Обратимые отображения окружности

Успех, которого мы достигли в исследовании поворотов окружности, вызван в значительной степени тем, что все результаты следуют из теории линейных динамических систем, а именно из результатов относительно поворотов плоскости (§3.1). Из этого следует однородность структуры орбиты, что приводит к однородной плотности орбит, а также к свойству равномерного распределения. Однако другая составляющая нашего успеха, быть может менее очевидная,

заключается в простоте структуры самой окружности. Так же как и в случае гомеоморфизмов отрезка (п. 2.3.1), это дает возможность проводить довольно содержательный анализ структуры орбиты любого обратимого отображения окружности. Одномерность окружности имеет две (связанные между собой) особенности, благодаря которым отображения хорошо поддаются анализу: это (циклический) порядок точек и теорема о промежуточном значении. Эти два свойства настолько тесно связывают различные орбиты, что структура орбит становится относительно легкой для описания. Значение структуры порядка станет особенно очевидным при обращении к предложениям 4.3.11 и 4.3.15.

Отметим также, что для необратимых отображений отрезка или окружности порядок точек может и не сохраняться, а следовательно, и использование этого свойства невозможно, в то время как теорема о промежуточном значении остается в силе, до тех пор пока имеется непрерывность. Соответственно и структурные особенности намного более сложны, но все же в значительной степени поддаются анализу. О некоторых необратимых отображениях отрезка подробнее говорится в гл. 11.

Один принцип, который проявится в различных аспектах на протяжении всего этого параграфа, состоит в том, что, хотя структура орбиты обратимых отображений окружности не полностью однородна, асимптотическое поведение почти столь же однородно или по крайней мере когерентно, как и структура орбит вращения, и в конечном счете нелинейные отображения оказываются похожими на поворот.

В этом параграфе рассматриваются два вида гомеоморфизмов окружности (см. определение A.1.16), в зависимости от того, имеют они периодические точки или нет. Каждая орбита имеет определенный тип асимптотического поведения, в первом случае соответствующий орбите рационального поворота, а во втором — иррационального. Мы введем параметр, который позволяет определить скорость асимптотического вращения, а также выяснить, является оно рациональным или нет, в зависимости от наличия периодических точек.

4.3.1. Поднятие и степень. Напомним соотношение между окружностью $S^1 = \mathbb{R}/\mathbb{Z}$ и прямой $\mathbb{R}$ (см. п. 2.6.2). Имеется проекция $\pi: \mathbb{R} \rightarrow S^1$, $x \mapsto [x]$, где $[x]$ — класс эквивалентности элемента x в $\mathbb{R}/\mathbb{Z}$, как описано в п. 2.6.2. Здесь $[\cdot]$ обозначает класс эквивалентности, в то время как целая часть числа обозначается $[\cdot]$. Мы также используем фигурные скобки $\{\cdot\}$ для обозначения дробной части.

Предложение 4.3.1. *Если отображение $f: S^1 \rightarrow S^1$ непрерывно, то существует такое непрерывное отображение $F: \mathbb{R} \rightarrow \mathbb{R}$, называемое поднятием отображения f на $\mathbb{R}$, что*

$$F \circ \pi = \pi \circ F, \quad (4.3.1)$$

т.е. $f([z]) = [F(z)]$. Такое поднятие единственно с точностью до аддитивной целой постоянной, и $\deg(f) := F(x+1) - F(x)$ — целое число, не зависящее от $x \in \mathbb{R}$ и поднятия F . Оно называется степенью отображения f . Если f — гомеоморфизм, то $|\deg(f)| = 1$.

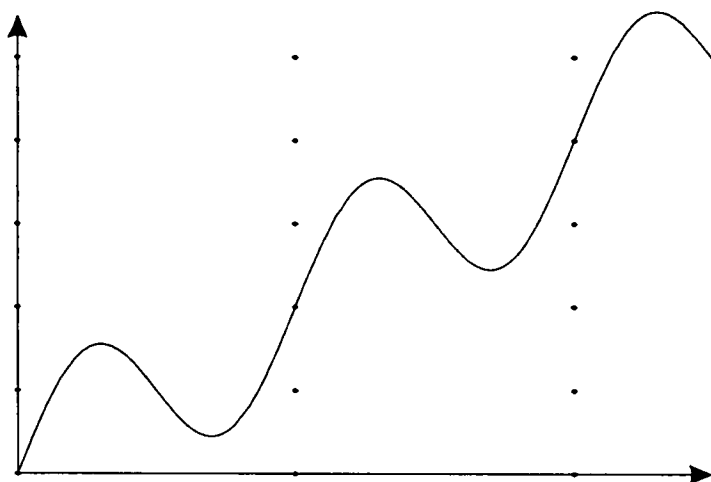


Рис. 4.3.1. Поднятие и степень

Доказательство. Существование. Возьмем точку $p \in S^1$. Тогда $p = [x_0]$ для некоторой точки $x_0 \in \mathbb{R}$ и $f(p) = [y_0]$ для некоторой точки $y_0 \in \mathbb{R}$. Исходя из выбора x_0 и y_0 определим $F: \mathbb{R} \rightarrow \mathbb{R}$, потребовав, чтобы отображение F было непрерывно и выполнялись равенства $F(x_0) = y_0$ и $f([z]) = [F(z)]$ для всех $z \in \mathbb{R}$. Такое отображение F можно построить, если непрерывно изменять начальную точку p , в результате чего $f(p)$ также будет изменяться непрерывно. Тогда мы можем непрерывно изменять x и y , не получая при этом неоднозначности, и, таким образом, $F(x) = y$ определяет непрерывное отображение³.

Единственность. Предположим, что $\tilde{F}$ — другое поднятие. Тогда $[\tilde{F}(x)] = f([x]) = [F(x)]$ для всех x , а следовательно, значение отображения $\tilde{F} - F$ — всегда целое число. Поскольку отображение непрерывно, это число должно быть постоянным.

Существование степени. Разность $F(x+1) - F(x)$ — целое число (теперь очевидно, что оно не зависит от выбора поднятия), поскольку $[F(x+1)] = f([x+1]) = f([x]) = [F(x)]$. В силу непрерывности величина $F(x+1) - F(x) =: \deg(f)$ должна быть постоянной.

Обратимость. Если $\deg(f) = 0$, то $F(x+1) = F(x)$ и, таким образом, отображение F не монотонно. Тогда отображение f необратимо, потому что оно не может быть монотонным. Если $|\deg(f)| > 1$, то $|F(x+1) - F(x)| > 1$ и по теореме о промежуточном значении существует такая точка $y \in (x, x+1)$, что $|F(y) - F(x)| = 1$, следовательно, $f([y]) = f([x])$ и $[y] \neq [x]$, так что отображение f необратимо. $\square$

³Более точно, выберем такое $\delta > 0$ что если $d([x], [x']) \leq \delta$, то $d(f([x]), f([x'])) < 1/2$, и определим F на отрезке $[x_0 - \delta, x_0 + \delta]$ следующим образом: если $|x - x_0| \leq \delta$, то $d(f([x]), q) < 1/2$ и имеется единственная такая точка $y \in (y_0 - 1/2, y_0 + 1/2)$, что $[y] = f([x])$. Определим $F(x) = y$. Аналогичным образом мы расширяем область определения на величину δ , до тех пор пока не определим F на единичном отрезке. Тогда равенство $f([z]) = [F(z)]$ определяет отображение F на всей прямой $\mathbb{R}$.

Определение 4.3.2. Пусть f — обратимое отображение. Если $\deg(f) = 1$, то мы говорим, что f *сохраняет ориентацию*, если $\deg(f) = -1$, то f *обращает ориентацию*.

Замечание 4.3.3. Заметим, что функция $F(x) - x \deg(f)$ периодическая, потому что

$$F(x+1) - (x+1) \deg(f) = F(x) + \deg(f) - (x+1) \deg(f) = F(x) + x \deg(f)$$

для всех x . В частности, если f — сохраняющий ориентацию гомеоморфизм, то функция $F(x) - x$ периодическая. Поэтому, в этом случае отображение $F - \text{Id}$ ограничено. Вскоре нам понадобится следующее несколько более сильное утверждение.

Лемма 4.3.4. Если f — сохраняющий ориентацию гомеоморфизм окружности и F — поднятие, то $F(y) - y \leq F(x) - x + 1$ для всех $x, y \in \mathbb{R}$.

Доказательство. Положим $k = \lfloor y - x \rfloor$. Тогда

$$\begin{aligned} F(y) - y &= F(y) + F(x+k) - F(x+k) + (x+k) - (x+k) - y = \\ &= (F(x+k) - (x+k)) + (F(y) - F(x+k)) - (y - (x+k)). \end{aligned} \quad (4.3.2)$$

Теперь заметим, что $F(x+k) - (x+k) = F(x) - x$ и при соответствующем выборе k выполнены неравенства $0 \leq y - (x+k) < 1$, поэтому $F(y) - F(x+k) \leq 1$. Таким образом, правая часть приведенного выше равенства не превосходит $F(x) - x + 1 - 0$. $\square$

4.3.2. Число вращения. Определить наличие или отсутствие периодических точек можно при помощи одного параметра, называемого числом вращения. Этот параметр также позволяет определить, с каким поворотом следует сравнивать гомеоморфизм окружности.

Предложение 4.3.5. Пусть $f: S^1 \rightarrow S^1$ — сохраняющий ориентацию гомеоморфизм и $F: \mathbb{R} \rightarrow \mathbb{R}$ — поднятие гомеоморфизма f . Тогда величина

$$\rho(F) := \lim_{|n| \rightarrow \infty} \frac{1}{n} (F^n(x) - x) \quad (4.3.3)$$

существует для всех $x \in \mathbb{R}$, не зависит от x и определена с точностью до целого числа, т. е. если $\tilde{F}$ — другое поднятие гомеоморфизма f , тогда $\rho(F) - \rho(\tilde{F}) = F - \tilde{F} \in \mathbb{Z}$. Число $\rho(F)$ рационально тогда и только тогда, когда f имеет периодическую точку.

Тот факт, что предел в (4.3.3) не зависит от точки, является первым проявлением ожидаемого когерентного асимптотического поведения орбит. Это предположение оправдывает следующую терминологию.

Определение 4.3.6. Класс эквивалентности $\rho(f) := [\rho(F)]$ называется *числом вращения*.

Последовательность, удовлетворяющая неравенству $a_{n+m} \leq a_n + a_m$, называется *субаддитивной*. Существование числа вращения следует из того, что выражения в правой части (4.3.3) обладают таким свойством.

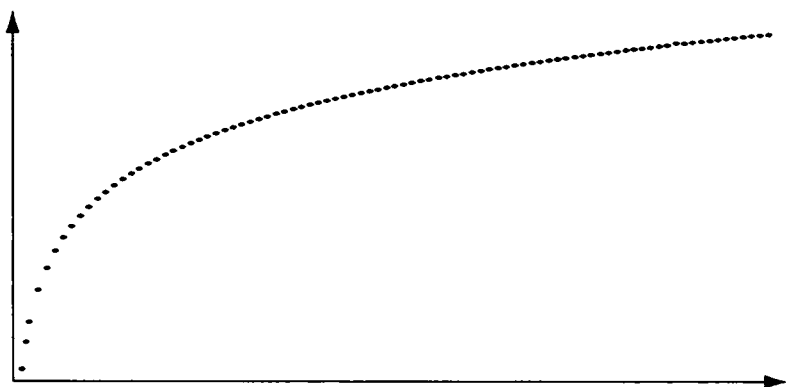


Рис. 4.3.2. Субаддитивность

Лемма 4.3.7. Если последовательность $(a_n)_{n \in \mathbb{N}}$ удовлетворяет неравенству $a_{m+n} \leq a_m + a_{n+k} + L$ для всех $m, n \in \mathbb{N}$ и некоторых k и L , то $\lim_{n \rightarrow \infty} a_n/n \in \mathbb{R} \cup \{-\infty\}$ существует.

Доказательство. Заметим сначала, что $a_{m+k} \leq a_m + a_{2k} + L$ и, следовательно,

$$a_{m+n} \leq a_m + a_n + a_{2k} + 2L = a_m + a_n + L',$$

так что мы можем взять $k = 0$. Пусть $a := \lim_{n \rightarrow \infty} a_n/n \in \mathbb{R} \cup \{-\infty\}$.

Если числа $a < b < c$ и $n > 2L/(c - b)$ таковы, что $a_n/n < b$, то для любого $l \geq n$, которое удовлетворяет неравенству $l(c - b) > 2 \max_{r < n} a_r$, мы можем записать $l = nk + r$, $r < n$. Тогда $a_l/l \leq (ka_n + a_r + kL)/l \leq a_n/n + a_r/l + (L/n) < c$. Это показывает, что $\lim_{l \rightarrow \infty} a_l/l \leq c$. Так как $c > a$ выбрано произвольно, это доказывает лемму. $\square$

Доказательство предложения 4.3.5. Докажем независимость числа вращения от точки x . Из замечания 4.3.3 следует, что $F(x + 1) = F(x) + 1$. Таким образом, если $|x - y| < 1$, то $|F(y) - F(x)| < 1$ и

$$\left| \frac{1}{n} |F^n(x) - x| - \frac{1}{n} |F^n(y) - y| \right| \leq \frac{1}{n} (|F^n(x) - F^n(y)| + |x - y|) \leq \frac{2}{n}.$$

Следовательно, числа вращения точек x и y совпадают, если одно из них существует.

Докажем существование. Возьмем $x \in \mathbb{R}$ и $a_n := F^n(x) - x$. Чтобы показать, что последовательность a_n/n сходится, заметим сначала, что, применяя лемму 4.3.4 к отображениям f^m и F^m , мы получим

$$a_{m+n} = F^{m+n}(x) - x = F^m(F^n(x)) - F^n(x) + a_n \leq a_m + 1 + a_n.$$

Таким образом, лемма 4.3.7 показывает, что последовательность a_n/n сходится (возможно, к $-\infty$). Но так как, кроме того, мы имеем

$$\frac{a_n}{n} = \frac{1}{n} \sum_{i=0}^{n-1} (F^{i+1}(x) - F^i(x)) = \frac{1}{n} \sum_{i=0}^{n-1} (F(x_i) - x_i) \geq \min F(y) - y,$$

предел есть вещественное число $\rho(F)$.

Заметим также, что $\rho(F + m) = \lim_{|n| \rightarrow \infty} (1/n)(F^n(x) + nm - x) = \rho(F) + m$ для $m \in \mathbb{Z}$, т. е. значение $\rho(F)$ корректно определено (mod 1).

Докажем утверждение о периодических точках. Если функция f имеет q -периодическую точку, тогда для ее поднятия x и некоторого $p \in \mathbb{Z}$ выполняется равенство $F^q(x) = x + p$. Таким образом, для $m \in \mathbb{N}$ мы имеем

$$\frac{F^{mq}(x) - x}{mq} = \frac{1}{mq} \sum_{i=0}^{m-1} F^q(F^{iq}(x)) - F^{iq}(x) = \frac{mp}{mq} = \frac{p}{q},$$

и $\rho(F) = p/q$.

Обратно, для любого поднятия F определение числа вращения приводит к соотношению

$$\rho(F^m) = \lim_{n \rightarrow \infty} \frac{1}{n} ((F^m)^n(x) - x) = M \lim_{n \rightarrow \infty} \frac{1}{mn} (F^{mn}(x) - x) = m\rho(F),$$

так что если $\rho(f) = p/q \in \mathbb{Q}$, то $\rho(f^q) = 0$, так как число вращения определено с точностью до целого числа. Таким образом, достаточно показать, что если $\rho(f) = 0$, то отображение f имеет неподвижную точку.

Предположим, что отображение f не имеет неподвижных точек, и пусть F — такое поднятие, что $F(0) \in [0, 1)$. Тогда $F(x) - x \notin \mathbb{Z}$ для всех $x \in \mathbb{R}$, поскольку условие $F(x) - x \in \mathbb{Z}$ означало бы, что $[x]$ — неподвижная точка для f . Поэтому $0 < F(x) - x < 1$ для всех $x \in \mathbb{R}$. Поскольку функция $F - \text{Id}$ непрерывная и периодическая, она достигает своих минимума и максимума, и поэтому существует такое $\delta > 0$, что

$$0 < \delta \leq F(x) - x \leq 1 - \delta < 1$$

для всех $x \in \mathbb{R}$. В частности, взяв $x = F^i(0)$ и используя равенство $F^n(0) = F^n(0) - 0 = \sum_{i=0}^{n-1} F^{i+1}(0) - F^i(0)$, мы получим, что

$$n\delta \leq F^n(0) \leq (1 - \delta)n,$$

или

$$\delta \leq \frac{F^n(0)}{n} \leq 1 - \delta.$$

При $n \rightarrow \infty$ получаем $\rho(F) \neq 0$, и это противоречие завершает доказательство. $\square$

Покажем, что все периодические орбиты, если они существуют, должны иметь один и тот же период.

Предложение 4.3.8. Пусть $f: S^1 \rightarrow S^1$ — сохраняющий ориентацию гомеоморфизм. Тогда все периодические орбиты имеют один и тот же период.

Если $\rho(f) = [p/q]$, где числа $p, q \in \mathbb{Z}$ взаимно просты, то поднятие F отображения f , для которого $\rho(F) = p/q$, удовлетворяет соотношению $F^q(x) = x + p$, если $[x]$ — периодическая точка, т. е. множество периодических точек отображения f поднимается до множества неподвижных точек отображения $F^q - \text{Id} - p$.

Доказательство. Если $[x]$ — периодическая точка, то $F^r(x) = x + s$ для некоторых $r, s \in \mathbb{Z}$ и, следовательно,

$$\frac{p}{q} = \rho(F) = \lim_{n \rightarrow \infty} \frac{F^{nr}(x) - x}{nr} = \lim_{n \rightarrow \infty} \frac{ns}{nr} = \frac{s}{r}.$$

Это означает, что $s = mr$ и $r = tq$, и поэтому $F^{mq}(x) = x + mp$. Мы должны показать, что на самом деле $F^q(x) = x + p$. Для этого заметим, что если $F^q(x) - p > x$, то в силу монотонности

$$F^{2q}(x) - 2p = F^q(F^q(x) - p) - p \geq F^q(x) - p > x$$

и по индукции $F^{mq}(x) - mp > x$, что невозможно. Аналогично неравенство $F^q(x) - p < x$ невозможно, потому что из него следовало бы неравенство $F^{mq}(x) - mp < x$. Это доказывает утверждение. $\square$

4.3.3. Инвариантность при сопряжениях. Понятие топологического сопряжения является центральным для многих аспектов теории динамических систем и будет введено позже (см. определение 7.3.3). Как показывает следующий результат, число вращения представляет собой первый нетривиальный пример инварианта относительно сопряжений.

Предложение 4.3.9. Если $h: S^1 \rightarrow S^1$ — сохраняющий ориентацию гомеоморфизм, тогда $\rho(h^{-1}fh) = \rho(f)$.

Доказательство. Пусть F и H — поднятия гомеоморфизмов f и h соответственно, т. е. $\pi F = f\pi$ и $\pi H = h\pi$. Тогда $\pi H^{-1} = H^{-1}h\pi H^{-1} = h^{-1}\pi H H^{-1} = h^{-1}\pi$, поэтому H^{-1} — поднятие гомеоморфизма h^{-1} . Аналогично $H^{-1}FH$ является поднятием гомеоморфизма $h^{-1}fh$, так как $\pi H^{-1}FH = h^{-1}\pi FH = h^{-1}f\pi H = h^{-1}fh\pi$.

Пусть отображение H таково, что $H(0) \in [0, 1)$. Мы должны оценить величину

$$|H^{-1}F^n H(x) - F^n(x)| = |(H^{-1}FH)^n(x) - F^n(x)|.$$

1. Для $x \in [0, 1)$ мы имеем $0 - 1 < H(x) - x < H(x) < H(1) < 2$, и из периодичности следует, что

$$|H(x) - x| < 2 \quad \text{для } x \in \mathbb{R}.$$

Подобным же образом мы получаем

$$|H^{-1}(x) - x| < 2 \quad \text{для } x \in \mathbb{R}.$$

2. Если $|y - x| < 2$, то $|F^n(y) - F^n(x)| < 3$, так как $||y| - |x|| \leq 2$, и, таким образом,

$$\begin{aligned} -3 \leq |y| - |x| - 1 &= F^n(|y|) - F^n(|x| + 1) < F^n(y) - F^n(x) < \\ < F^n(|y| + 1) - F^n(|x|) = |y| + 1 - |x| \leq 3. \end{aligned}$$

Из этих двух оценок следует, что

$$|H^{-1}F^n H(x) - F^n(x)| \leq |H^{-1}F^n H(x) - F^n H(x)| + |F^n H(x) - F^n(x)| < 2 + 3,$$

а потому $|(H^{-1}FH)^n(x) - F^n(x)|/n < 5/n$ и $\rho(H^{-1}FH) = \rho(F)$ (см. (4.3.3)). $\square$

Исследование поведения числа вращения при обращающих ориентацию сопряжениях предлагается читателю в качестве упражнения (см. упражнение 4.3.6)

4.3.4. Гомеоморфизмы окружности с периодическими точками. Структура орбит гомеоморфизма окружности может быть описана достаточно полно. Сначала мы сделаем это для случая, когда имеются периодические точки.

Прежде всего покажем, что периодические орбиты упорядочены таким же образом, как и орбиты соответствующего поворота. Это означает, что периодические орбиты сохраняющего ориентацию гомеоморфизма окружности ведут себя подобно орбитам поворота с тем же самым числом вращения. Так что имеется не только внутренняя согласованность различных периодических орбит (см. предложение 4.3.8), но также и качественная совместимость с орбитами поворота, как и можно было предположить исходя из доказательства предложения 4.3.8.

Прежде чем приступить к доказательству, необходимо определить «упорядоченность» орбиты, т.е. последовательность, в которой мы встречаем точки орбиты при движении от начальной точки в положительном направлении. Формально упорядочение можно определить, используя поднятия.

Определение 4.3.10. Для данных точек $x_0, \dots, x_{n-1} \in S^1$ возьмем такие точки $\bar{x}_0, \dots, \bar{x}_{n-1} \in \mathbb{R}$, $\bar{x}_i = x_i$. Тогда упорядочением множества $(x_0, \dots, x_{n-1})$ на S^1 называется такая перестановка σ на множестве $\{1, \dots, n-1\}$, что $\bar{x}_0 < \bar{x}_{\sigma(1)} < \dots < \bar{x}_{\sigma(n-1)} < \bar{x}_0 + 1$.

Для начала найдем упорядочение σ для $\pi\left(\left\{0, \frac{p}{q}, \frac{2p}{q}, \dots, \frac{(q-1)p}{q}\right\}\right)$ на S^1 , с которым мы позже сравним упорядочение периодической орбиты. Определим $k \in \mathbb{N}$ из условий $0 < k < q$ и $kp \equiv 1 \pmod{q}$. Тогда k минимизирует дробную часть $\{ip/q\}$ для $0 < i < q$, и, следовательно, $k \equiv \sigma(1)$. По индукции отсюда заключаем, что $ki \equiv \sigma(i) \pmod{q}$. Это определяет порядок σ для $\pi\left(\left\{0, \frac{p}{q}, \frac{2p}{q}, \dots, \frac{(q-1)p}{q}\right\}\right)$.

Таким образом, мы хотим доказать следующий результат.

Предложение 4.3.11. Пусть $f: S^1 \rightarrow S^1$ — сохраняющий ориентацию гомеоморфизм. Предположим, что числа p и q взаимно просты, и пусть $x \in S^1$ — такая точка, что $f^q(x) = x$. Тогда упорядочение множества $\{x, f(x), f^2(x), \dots, f^{q-1}(x)\}$ на S^1 задается соотношением $\sigma(i) = ki \pmod{q}$, где $kp \equiv 1 \pmod{q}$.

Доказательство. Зафиксируем такую точку $\bar{x} \in \pi^{-1}(\{x\})$ и такое поднятие F гомеоморфизма f , что $F^q(\bar{x}) = \bar{x} + p$ (см. предложение 4.3.8). Тогда отрезок $[\bar{x}, \bar{x} + p]$ разбивается (с точностью до общих концевых точек) на $p \cdot q$ отрезков элементами множества $A := \pi^{-1}(\{x, f(x), f^2(x), \dots, f^{q-1}(x)\})$ и на q отрезков вида $I_i = [F^i(\bar{x}), F^{i+1}(\bar{x})]$, $i = 0 \dots q-1$. Так как F — взаимно однозначное соответствие между любыми двумя смежными отрезками I_i и I_{i+1} в этом разбиении и оно сохраняет A , каждый отрезок I_i содержит $p+1$ точку из A . Возьмем такие числа $k, r \in \mathbb{Z}$, что правый сосед точки $\bar{x}$ в A — это $\bar{x}_1 = F^k(\bar{x}) - r$. Так как функция $\bar{F} = F^k - r$ возрастает на $\mathbb{R}$ и сохраняет множество A , из того обстоятельства, что точка $\bar{x}_1 = \bar{F}(\bar{x})$ является самым ближайшим правым соседом точки $\bar{x}$ в A и что отрезок $[\bar{x}, F(\bar{x})]$ разбивается на p отрезков элементами множества A , следует, что $\bar{F}(\bar{x}) = F(\bar{x})$ и, таким образом, $F^k p(x) = f(x)$. Поэтому k — единственное такое целое число между 0 и $q-1$, что $kp \equiv 1 \pmod{q}$, и упорядочение орбиты $\{x, f(x), f^2(x), \dots, f^{q-1}(x)\}$ задается соотношением $ki \equiv \sigma(i) \pmod{q}$. $\square$

Следующее предложение утверждает, что для гомеоморфизмов окружности с рациональным числом вращения все непериодические орбиты стремятся к периодическим орбитам. Это позволяет получить полную классификацию возможных орбит с рациональными числами вращения.

Предложение 4.3.12. Пусть $f: S^1 \rightarrow S^1$ — сохраняющий ориентацию гомеоморфизм с рациональным числом вращения $p(f) = p/q \in \mathbb{Q}$. Тогда имеются два следующих возможных типа непериодических орбит для f .

1) Если гомеоморфизм f имеет в точности одну периодическую орбиту, тогда любая другая точка является гетероклинической для f^q к двум точкам на периодической орбите множества. Эти точки различны, если период больше единицы. (Если период равен 1, тогда все орбиты гомоклиничны к неподвижной точке, как показано на рис. 4.3.3.)

2) Если гомеоморфизм f имеет больше одной периодической орбиты, тогда каждая непериодическая точка является гетероклинической для f^q к двум точкам на различных периодических орбитах.

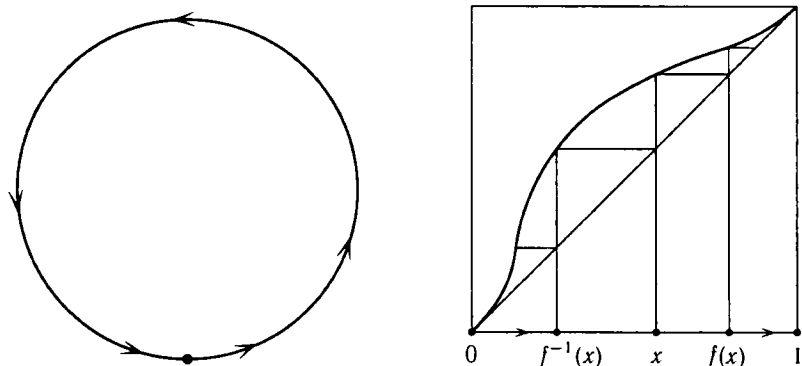


Рис. 4.3.3. Полуустойчивая точка

Доказательство. Заметим сначала, что можно отождествить отображение f^q с гомеоморфизмом отрезка, взяв поднятие z неподвижной точки отображения f^q и ограничив поднятие $F^q(\cdot) - p$ отображения f на $[z, z + 1]$. Но тогда мы получим почти все нужные нам утверждения, если применим к этому отображению предложение 2.3.5. Остается доказать только последнюю часть утверждения 2), а именно убедиться в том, что две предельные периодические орбиты различны. Но если имеется такой отрезок $I = [a, b] \subset \mathbb{R}$, что a и b являются смежными нулями отображения $F^q - \text{Id} - p$ и проектируются в одну и ту же периодическую орбиту, то гомеоморфизм f имеет только одну периодическую орбиту, потому что если $[a] = x \in S^1$, $[b] = f^k(x) \in S^1$, то множество $\bigcup_{n=0}^{q-1} f^{nk}(\pi((a, b)))$

покрывает дополнение к множеству $\{f^n(x)\}_{n=0}^{q-1}$ в S^1 и не содержит никаких периодических точек. В силу инвариантности множество $f^{nk}(\pi((a, b)))$ также не содержит периодических точек. $\square$

Замечание 4.3.13. Это означает, что асимптотическое поведение очень схоже для всех орбит, не только для периодических, и также согласовано со структурой соответствующего поворота.

В частности, если имеется только одна периодическая орбита, то она полуустойчива, т.е. «отталкивает с одной стороны и притягивает с другой», как, например, точка $x = 0$ для гладкого гомеоморфизма $f: S^1 \rightarrow S^1$,

$$x \mapsto x + \frac{1}{4} \sin^2 \pi x \pmod{1}.$$

Заметим, что не только индивидуальные непериодические орбиты притягиваются к периодическим, но еще и поведение различных орбит согласовано. Так, для непериодической точки x все точки $x, f(x), \dots, f^{q-1}(x)$ стремятся к соответствующим итерациям $y, f(y), \dots, f^{q-1}(y)$ периодической точки, и они движутся в том же самом направлении. Это немедленно следует из монотонности (ср. с леммой 2.3.2).

Лемма 4.3.14. Если $I \subset \mathbb{R}$ — отрезок, концы которого являются смежными нулями функции $F^q - \text{Id} - p$, то функция $F^q - \text{Id} - p$ имеет один и тот же знак внутри I и $F(I)$.

Доказательство. Если $F^q - \text{Id} - p > 0$ на I , то $F^q(x) > x + p$ для всех $x \in I$. Поскольку отображение F монотонно, $F^q(F(x)) = F(F^q(x)) > F(x + p) = F(x) + p$ для всех $x \in I$, и поэтому $F^q - \text{Id} - p > 0$ на $F(I)$.

Случай $F^q - \text{Id} - p > 0$ рассматривается аналогично. $\square$

Таким образом, для любого гомеоморфизма окружности с периодической точкой все точки асимптотически стремятся к периодическим орбитам одного и того же периода, и это происходит согласованным образом.

4.3.5. Гомеоморфизмы окружности без периодических точек. Аналогично предложению 4.3.11 мы докажем, что орбиты отображения окружности, не имеющего периодических точек, упорядочены таким же образом, как для соответствующего поворота.

Предложение 4.3.15. Пусть $F: \mathbb{R} \rightarrow \mathbb{R}$ — поднятие сохраняющего ориентацию гомеоморфизма $f: S^1 \rightarrow S^1$ с числом вращения $\rho := \rho(F) \notin \mathbb{Q}$, и пусть $n_1, n_2, m_1, m_2 \in \mathbb{Z}$ и $x \in \mathbb{R}$. Тогда

$$n_1 \rho + m_1 < n_2 \rho + m_2 \quad \text{тогда и только тогда, когда} \quad F^{n_1}(x) + m_1 < F^{n_2}(x) + m_2.$$

Левое из этих неравенств представляет собой частный случай правого (когда f — поворот на угол ρ).

Доказательство. Заметим, что равенство в правой части не может выполняться ни для какого x , потому что в этом случае выполнялось бы включение $F^{n_1}(x) - F^{n_2}(x) \in \mathbb{Z}$ и, следовательно, имела бы периодическая точка $[x]$. Таким образом, для данных $n_1, n_2, m_1, m_2 \in \mathbb{Z}$ непрерывная функция $F^{n_1}(x) + m_1 - F^{n_2}(x) - m_2$ никогда не изменяет знак, и второе неравенство не зависит от x .

Предположим теперь, что $F^{n_1}(x) + m_1 < F^{n_2}(x) + m_2$ для всех x . Замена $y := F^{n_2}(x)$ показывает, что это условие эквивалентно неравенству

$$F^{n_1 - n_2}(y) - y < m_2 - m_1 \quad \text{для всех} \quad y \in \mathbb{R}.$$

Подстановка $y = 0$ и $y = F^{n_1 - n_2}(0)$ приводит соответственно к неравенствам

$$F^{n_1 - n_2}(0) < m_2 - m_1$$

и

$$F^{2(n_1 - n_2)}(0) < (m_2 - m_1) + F^{n_1 - n_2}(0) < 2(m_2 - m_1).$$

По индукции $F^{n(n_1 - n_2)}(0) < n(m_2 - m_1)$ и

$$\rho = \lim_{n \rightarrow \infty} \frac{F^{n(n_1 - n_2)}(0)}{n(n_1 - n_2)} < \lim_{n \rightarrow \infty} \frac{n(m_2 - m_1)}{n(n_1 - n_2)} = \frac{m_2 - m_1}{n_1 - n_2}$$

(неравенство строгое, поскольку $\rho \notin \mathbb{Q}$). Поэтому

$$n_1 \rho + m_1 < n_2 \rho + m_2.$$

Это доказывает необходимость. Следуя этим неравенствам в обратном порядке, получаем достаточность. $\square$

Это предложение имеет некоторое сходство с предыдущим результатом, устанавливающим, что в случае рационального числа вращения *периодические* орбиты упорядочиваются подобно орбитам соответствующего поворота. Однако данный результат более сильный, поскольку он применим к каждой орбите, а не к некоторому выделенному подмножеству. Это обстоятельство поможет нам при изучении асимптотического поведения орбит гомеоморфизмов, не имеющих периодических точек.

Лемма 4.3.16. Пусть $f: S^1 \rightarrow S^1$ — сохраняющий ориентацию гомеоморфизм, $\tau(f) \in \mathbb{R} \setminus \mathbb{Q}$, $m, n \in \mathbb{Z}$, $m \neq n$, $x \in S^1$ и $I \subset S^1$ — отрезок с концами $f^m(x)$ и $f^n(x)$. Тогда каждая полуорбита имеет непустое пересечение с I .

Замечание 4.3.17. Для $x, y \in S^1$, $x \neq y$, имеются в точности два отрезка в S^1 с концами x и y . Лемма справедлива для любого из них. Поскольку точка x не периодическая, отрезок I никогда не вырождается в точку.

Доказательство. Рассмотрим положительные полуорбиты $(f^k(y))_{k \in \mathbb{N}}$. Доказательство для отрицательных полуорбит в точности такое же. Чтобы доказать лемму, достаточно убедиться, что обратные итерации отрезка I покрывают S^1 , т. е. $S^1 \subset \bigcup_{k \in \mathbb{N}} f^{-k}(I)$.

Положим $I_k := f^{-k(n-m)}(I)$ и заметим, что все эти отрезки смежные: если $k \in \mathbb{N}$, то I_k и I_{k-1} имеют общую концевую точку. Следовательно, если $S^1 \neq \bigcup_{k \in \mathbb{N}} I_k$, то последовательность концевых точек сходится к некоторой точке $z \in S^1$. Но тогда точка

$$\begin{aligned} z &= \lim_{k \rightarrow \infty} f^{-k(n-m)}(f^m(x)) = \lim_{k \rightarrow \infty} f^{(-k+1)(n-m)}(f^m(x)) = \\ &= \lim_{k \rightarrow \infty} f^{(n-m)}(f^{-k(n-m)}(f^m(x))) = f^{(n-m)}\left(\lim_{k \rightarrow \infty} f^{-k(n-m)}(f^m(x))\right) = \\ &= f^{(n-m)}(z) \end{aligned}$$

является периодической, что противоречит предположению. $\square$

В случае наличия периодических точек они составляют множество всех предельных точек орбит. Теперь мы посмотрим, какое множество играет эту роль в случае, когда число вращения является иррациональным.

Определение 4.3.18. Множество $\omega(x) := \bigcap_{n \in \mathbb{N}} \overline{\{f^i(x) : i \geq n\}}$ предельных точек положительной полуорбиты точки x называется ω -предельным множеством точки x .

Если имеются периодические точки, то все ω -предельные множества — это периодические орбиты. Но даже если периодических точек нет, ω -предельные множества для разных орбит выглядят одинаково, и на самом деле они совпадают.

Предложение 4.3.19. Пусть $f: S^1 \rightarrow S^1$ — сохраняющий ориентацию гомеоморфизм окружности S^1 без периодических точек. Тогда $\omega(x)$ не зависит от x , а множество $E := \omega(x)$ совершенно и либо совпадает с S^1 , либо нигде не плотно (см. определение A.1.5).

Как следует из предложения A.1.7, совершенные нигде не плотные множества являются канторовыми множествами, т.е. они гомеоморфны стандартному канторову множеству. Таким образом, наш результат позволяет получать канторовы множества непосредственно с использованием динамики отображения окружности, по крайней мере, когда мы приведем пример, где эта возможность действительно реализуется.

Доказательство. Установим независимость от x . Мы должны показать, что $\omega(x) = \omega(y)$ для $x, y \in S^1$. Пусть $z \in \omega(x)$. Тогда имеется такая последовательность $l_n \in \mathbb{N}$, что $f^{l_n}(x) \rightarrow z$. Если $y \in S^1$, то по лемме 4.3.16 найдутся такие числа $k_m \in \mathbb{N}$, что $f^{k_m}(y) \in I_m := [f^{l_m}(x), f^{l_m+1}(x)]$. Но тогда $\lim_{m \rightarrow \infty} f^{k_m}(y) = z$ и, таким образом, $z \in \omega(y)$.

Поэтому $\omega(x) \subset \omega(y)$ для любых $x, y \in S^1$, и в силу симметрии $\omega(x) = \omega(y)$ для любых $x, y \in S^1$.

Докажем теперь, что множество $E := \omega(x)$ либо совпадает с S^1 , либо нигде не плотно. Сначала покажем, что E — наименьшее замкнутое непустое f -инвариантное множество. Если $A \subset S^1$ — произвольное замкнутое f -инвариантное множество и $A \neq \emptyset$, то для любой точки $x \in A$ мы имеем $\{f^k(x)\}_{k \in \mathbb{Z}} \subset A$, так как множество A инвариантно. Кроме того, $E = \omega(x) \subset \overline{\{f^k(x)\}_{k \in \mathbb{Z}}} \subset A$, так как множество A замкнуто. Таким образом, любое непустое замкнутое инвариантное множество A содержит множество E . В частности, $\emptyset$ и E — единственные замкнутые инвариантные подмножества множества E . Поскольку множество E замкнуто, оно содержит свою границу, которая сама является замкнутым множеством (см. упражнение 2.6.6). Граница также инвариантна, потому что по определению граничная точка — это точка, для любой окрестности U которой выполняются условия $U \cap E \neq \emptyset$ и $U \setminus E \neq \emptyset$, а это свойство сохраняется при гомеоморфизме. Таким образом, граница ∂E — замкнутое инвариантное подмножество множества E , и поэтому либо $\partial E = \emptyset$, а следовательно, $E = S^1$, либо $\partial E = E$, откуда следует, что множество E нигде не плотно (см. упражнение 2.6.6).

Остается показать, что множество E совершенно. Возьмем $x \in E$. Поскольку $E = \omega(x)$, имеется такая последовательность $(k_n)_{n \in \mathbb{N}}$, что $\lim_{n \rightarrow \infty} f^{k_n}(x) = x$. Далее, поскольку периодических орбит нет, $f^{k_n}(x) \neq x$ для всех n . Следовательно, x — предельная точка множества E , так как $f^{k_n}(x) \in E$ для всех n в силу инвариантности. $\square$

4.3.6. Сравнение и классификация. Как в предложении 4.3.12, так и в предложении 4.3.19 имеется выделенное множество орбит (или периодических, или лежащих в E), к которому все другие орбиты асимптотически стремятся. Это выделенное множество тесно связано с поворотом, имеющим такое же число вращения (для случая иррационального числа вращения это станет понятно из теоремы 4.3.20). Таким образом, в случае наличия периодических точек имеется «остаток» от поворота, который может являться всего-навсего одной периодической орбитой или конечным множеством таковых. А в случае иррационального числа вращения соответствующий остаток — по крайней мере канторово множество. В этом и заключается различие — предложение 4.3.19 показывает, что структура орбит отображений без периодических точек существенно отличается от структуры орбит для отображений с периодическими точками. В случае наличия периодических точек все орбиты или являются периодическими, или стремятся к периодической орбите, а для отображений без периодических точек либо все орбиты плотны, либо все орбиты асимптотически стремятся к канторову множеству или содержатся в нем. Более того, дальнейшее различие проявляется, когда мы сравниваем структуру орбиты отображения окружности со структурой орбиты поворота с тем же самым

числом вращения. Подавляющее большинство отображений окружности с периодическими точками содержит непериодические орбиты — предложение 4.4.10 и лемма 4.4.12 показывают, что наличие непериодических орбит наблюдается на целых интервалах значений параметра в семействах отображений, тогда как периодичность всех орбит наблюдается лишь случайно. (Кроме того, подобные рассуждения показывают, что даже наличие бесконечного числа периодических точек неустойчиво, а следовательно, редко.) Таким образом, наличие непериодических орбит является наиболее общим поведением для отображений с рациональным числом вращения, и в этом состоит принципиальное качественное отличие таких отображений от рациональных поворотов.

Для иррационального числа вращения картина иная. Наибольшее качественное подобие иррациональному повороту наблюдается в случае $E = S^1$ в предложении 4.3.19. В этом случае все орбиты плотны ($\omega(x) = S^1$ для всех $x \in S^1$), т.е. возникает такая же ситуация, как и для иррациональных поворотов. В отличие от случая рационального числа вращения нет никаких оснований ожидать, что другая ситуация (E — канторово множество) встречается чаще (на самом деле для отображений класса C^2 такого вообще не может быть). Действительно, отображение с иррациональным числом вращения ρ или эквивалентно вращению, или в определенном смысле «содержит» его с точностью до некоторого искажения в зависимости от того, являются орбиты плотными или нет.

Теорема 4.3.20 (классификация Пуанкаре). Пусть $f: S^1 \rightarrow S^1$ — сохраняющий ориентацию гомеоморфизм с иррациональным числом вращения ρ . Тогда имеется такое непрерывное монотонное отображение $h: S^1 \rightarrow S^1$, что $h \circ f = R_\rho \circ h$, причем

- 1) если гомеоморфизм f транзитивен, то h — гомеоморфизм;
- 2) если гомеоморфизм f не транзитивен, то отображение h необратимо.

Отображение h здесь играет роль замены переменной или сопряжения, с которыми мы встретились в п. 1.2.9.3 и 3.1.3, за исключением того, что оно может и не быть обратимым.

Доказательство. Сначала мы построим поднятие отображения h только на поднятии одной орбиты и покажем, что оно монотонно. Тогда мы можем продолжить отображение, полученное таким образом, на замыкание этого поднятия и, используя монотонность, «заполнить» все пробелы, которые могут оставаться. Наконец, мы определим h как проекцию.

Рассмотрим поднятие $F: \mathbb{R} \rightarrow \mathbb{R}$ гомеоморфизма f и выберем точку $x \in \mathbb{R}$. Пусть $B := \{F^n(x) + m\}_{n,m \in \mathbb{Z}}$ — полное поднятие орбиты элемента $[x]$. Определим $H: B \rightarrow \mathbb{R}$, $F^n(x) + m \mapsto nr + m$, где $\rho := \rho(F)$. Из предложения 4.3.15 следует, что это отображение монотонно. Кроме того, множество $H(B)$ плотно в $\mathbb{R}$ согласно предложению 4.1.1. Если мы обозначим отображение $\mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto x + \rho$ через $\bar{R}_\rho$, то $H \circ F = \bar{R}_\rho \circ H$ на B , поскольку

$$H \circ F(F^n(x) + m) = H(F^{n+1}(x) + m) = (n+1)r + m$$

и

$$\bar{R}_\rho \circ H(F^n(x) + m) = \bar{R}_\rho(nr + m) = (n+1)r + m.$$

Лемма 4.3.21. Отображение H имеет непрерывное продолжение на замыкание $\bar{B}$ множества B .

Доказательство. Если $y \in \bar{B}$, то имеется такая последовательность $(x_n)_{n \in \mathbb{N}}$ в B , что $y = \lim_{n \rightarrow \infty} x_n$. Чтобы показать, что $H(y) := \lim_{n \rightarrow \infty} H(x_n)$ существует и не зависит от выбора последовательности, сходящейся к y , заметим сначала, что левый и правый пределы существуют и не зависят от последовательности, так как отображение H монотонно. Если левый и правый пределы не совпадают, то множество $\mathbb{R} \setminus H(B)$ содержит интервал, что противоречит плотности множества $H(B)$. $\square$

Теперь отображение H легко может быть продолжено на $\mathbb{R}$: поскольку $H: \bar{B} \rightarrow \mathbb{R}$ монотонно и сюръективно (потому что H монотонно и непрерывно на B , множество $\bar{B}$ замкнуто, и множество $H(B)$ плотно в $\mathbb{R}$), на интервалах, дополнительных к $\bar{B}$, отображение H определяется однозначно: на таких интервалах мы полагаем $H = \text{const}$, выбирая постоянную равной значениям в концах интервала. В результате мы получаем такое отображение $H: \mathbb{R} \rightarrow \mathbb{R}$, что $H \circ F = \bar{R}_\rho \circ H$, и, таким образом, искомое отображение $h: S^1 \rightarrow S^1$, так как для $z \in B$ справедливо равенство

$$H(z+1) = H(F^n(x) + m + 1) = nr + m + 1 = H(z) + 1$$

и это свойство сохраняется при непрерывном продолжении.

Чтобы исследовать вопрос об обратимости, заметим, что в транзитивном случае мы начинаем с плотной орбиты и поэтому $\bar{B} = \mathbb{R}$ и h — взаимно однозначное соответствие. В нетранзитивном случае отображение h постоянно на интервалах вне замыкания орбиты. $\square$

Замечание 4.3.22. В транзитивном случае теоремы 4.3.20, когда отображение h обратимо, мы говорим, что h сопрягает гомеоморфизм f с R_p , в случае необратимого отображения h мы говорим, что h является *полусопряжением* или что R_p — *фактор* гомеоморфизма f посредством h . Эти понятия изучаются в гл. 7 (определение 7.3.3).

Упражнения

4.3.1. Для каких значений параметра a функция $F(x) = 2x + a$ определяет поднятие отображения окружности?

4.3.2. Докажите с использованием соотношения (4.1.1), что $\rho(R_\alpha) = [\alpha]$.

4.3.3. Определите, является ли отображение $F(x) := x + (1/2) \sin x$ поднятием гомеоморфизма окружности.

4.3.4. Определите, является ли отображение $F(x) := x + (1/4\pi) \sin 2\pi x$ поднятием гомеоморфизма окружности, и если является, то сохраняет ли оно ориентацию. В случае утвердительного ответа найдите число вращения.

4.3.5. Пусть $f: S^1 \rightarrow S^1$ — монотонное (но не обязательно обратимое) отображение степени один, т. е. его поднятие — такая монотонная функция $F: \mathbb{R} \rightarrow \mathbb{R}$, что $F(x+1) = F(x) + 1$. Докажите, что утверждения предложений 4.3.5, 4.3.8 и 4.3.9 справедливы для f .

4.3.6. Что случится с числом вращения в условиях предложения 4.3.9 при применении обращающего ориентацию сопряжения?

4.3.7. Пусть $f: S^1 \rightarrow S^1$ — непрерывное отображение степени один (не обязательно монотонное) и $F: \mathbb{R} \rightarrow \mathbb{R}$ — его поднятие. Докажите, что оба предела

$$\rho^+(F) := \lim_{n \rightarrow \infty} \max_{x \in S^1} \frac{F^n(x) - x}{n} \quad \text{и} \quad \rho^-(F) := \lim_{n \rightarrow \infty} \min_{x \in S^1} \frac{F^n(x) - x}{n}$$

существуют.

Задачи для углубленного изучения

4.3.8. В обозначениях предыдущего упражнения назовем множество

$$R(F) := \left\{ \rho \in \mathbb{R} : \exists x \in \mathbb{R} \lim_{n \rightarrow \infty} \frac{F^n(x) - x}{n} = \rho \right\}$$

множеством вращения отображения F . Докажите, что $R(F) \neq \emptyset$.

4.3.9. Докажите, что для гомеоморфизма окружности с конечным числом неподвижных точек и с притягивающей неподвижной точкой существует также и отталкивающая неподвижная точка.

Покажите, что существует гомеоморфизм окружности, обладающий притягивающей неподвижной точкой, но не имеющий отталкивающих неподвижных точек.

§ 4.4. Феномен Кантора

Канторово множество естественным образом возникает в предложении 4.3.19 для некоторых гомеоморфизмов окружности без периодических точек. Канторовы множества и связанные с ними структуры могут появляться в том же контексте и в других ситуациях. Рассмотренное выше сопряжение возникает в нетранзитивном случае. Зависимость числа вращения от параметра представляет собой пример, имеющий интересные физические приложения.

4.4.1. Канторова лестница. В нетранзитивном случае отображение h из теоремы 4.3.20 служит примером следующего интересного явления.

Определение 4.4.1. Монотонная непрерывная функция $\phi: [0, 1] \rightarrow \mathbb{R}$ (или $\phi: [0, 1] \rightarrow S^1$) называется *канторовой лестницей*, если существует такое семейство $\{I_\alpha\}_{\alpha \in A}$ непересекающихся отрезков ненулевой длины с плотным объединением, лежащих в $[0, 1]$, что ϕ принимает различные постоянные значения на этих отрезках (см. рис. 4.4.1).

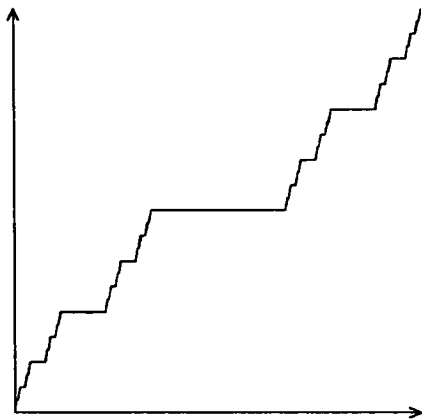


Рис. 4.4.1. Канторова лестница

Пример 4.4.2. Канторову лестницу можно построить довольно явным образом. Для $x = 0, \alpha_1 \alpha_2 \alpha_3 \dots = \sum_{i=1}^{\infty} \alpha_i 3^{-i}$ ($\alpha_i \neq 1$) в троичном множестве Кантора C определим $f(x) := \sum_{i=1}^{\infty} \alpha_i 2^{-i-1} \in [0, 1]$, как в лемме 2.7.3. Как было показано в п. 2.7.1, f — сюръективная и неубывающая функция, и два конца дополнительного интервала отображаются в одну и ту же точку. Нетрудно видеть, что функция f является непрерывной (это обстоятельство используется в упражнении 2.7.6). Мы можем продолжить f до неубывающего непрерывного отображения на $[0, 1]$, считая его постоянным на каждом дополнительном интервале, где постоянные выбираются равными значениям функции f в концевых точках. Полученное отображение будет канторовой лестницей.

График этой функции имеет некоторое самоподобие: преобразование плоскости, заданное матрицей $\begin{pmatrix} 1/3 & 0 \\ 0 & 1/2 \end{pmatrix}$, отображает график на его собственное подмножество, поскольку $f(x) = f(3x)/2$ на $[0, 1/3]$.

В англоязычной литературе канторову лестницу называют «лестницей дьявола». Этот термин возник из-за необычной ситуации — график этой функции полностью состоит из «ступенек», а именно горизонтальных отрезков на дополнительных интервалах, но все же разрывов эта функция не имеет — она непрерывна. Таким образом, сами ступеньки на лестнице имеются, но вот их боковые грани отсутствуют. Тем самым мы получаем удивительный пример функции с некоторыми экзотическими свойствами, но, как мы видели, в динамике такие функции возникают довольно естественно.

Обратимся еще раз к приведенному выше построению отображения h , для того чтобы лучше понять нетранзитивный случай. Так как множество $\bar{B}$ из доказательства теоремы 4.3.20 проектируется на замыкание орбиты элемента $[x]$, оно содержит ω -предельное множество $E = \omega([x])$ элемента $[x]$, и, выбирая $x \in \pi^{-1}(E)$, мы получаем $\pi(\bar{B}) = E$, где E — универсальное ω -предельное множество, рассмотренное выше. В транзитивном случае $\bar{B} = \mathbb{R}$ и $E = S^1$, но в нетранзитивном случае мы получаем, что если $x \in \pi^{-1}(E)$, то $\pi(\bar{B}) = E$ — канторово множество. Следовательно, в нетранзитивном случае h является взаимно однозначным соответствием между факторпространством $E/\sim$ (идентифицируются две концевые точки каждого дополнительного интервала) и S^1 . Оно сопрягает $f|_{E/\sim} \subset R_{\mathbb{R}(f)}$. Все орбиты отображения f в E плотны в E (по определению множества E). С другой стороны, из построения множества $E = \omega(x)$ следует, что все точки вне E притягиваются к E и при положительных, и при отрицательных итерациях, поскольку итерации такой точки должны оставаться внутри непересекающихся дополнительных интервалов к множеству E , длина которых стремится к нулю.

4.4.2. Блуждающие области. Можно представлять нетранзитивное отображение и иначе — как полученное из иррационального поворота «раздуванием» точек некоторых орбит в интервалы, объединение которых тогда составляет дополнение к E . Эти дополнительные интервалы переставляются подобно точкам в орбите иррационального поворота. Все внутренние точки в этих интервалах — «блуждающие» в смысле приводимого ниже определения, так как они остаются в пределах этих интервалов, образы которых попарно не пересекаются. В следующем пункте мы приведем явное построение такого примера.

Определение 4.4.3. Точка называется *блуждающей*, если она имеет окрестность, все образы и прообразы которой попарно не пересекаются.

Подобное поведение абсолютно противоположно рекуррентности (см. определение 6.1.8).

Возвращаясь к сравнению нашей ситуации со случаем рационального числа вращения, мы заметим, что отображение f сопряжено с поворотом только в том случае, если все орбиты периодические с одинаковым периодом и, следовательно, $f^q = \text{Id}$ для некоторого $q \in \mathbb{Z}$. Кроме того, рациональный поворот может быть фактором только в том случае, когда имеется бесконечно много периодических точек, т. е. в ситуации, которая, как было замечено ранее, неустойчива.

4.4.3. Пример Данжуа. В этом пункте мы приведем пример нетранзитивного диффеоморфизма окружности без периодических точек. Мы начнем построение с иррационального поворота и заменим точки одной орбиты подходящим образом выбранными отрезками. Полученное в результате отображение не является транзитивным. Арно Данжуа доказал следующий результат.

Предложение 4.4.4. Если $\rho \in \mathbb{R} \setminus \mathbb{Q}$, то существует такой нетранзитивный C^1 -диффеоморфизм $f: S^1 \rightarrow S^1$, что $\rho(f) = \rho$.

Доказательство. Пусть $l_n := (|n| + 3)^{-2}$, $c_n := 2((l_{n+1}/l_n) - 1) \geq -1$. Тогда

$$\sum_{n \in \mathbb{Z}} l_n < 2 \sum_{n=0}^{\infty} l_n = 2 \sum_{n=3}^{\infty} \frac{1}{n^2} < 2 \int_2^{\infty} \frac{1}{x^2} dx = 1.$$

Чтобы «вклеить» в орбиты $x_n = R_p^n x$ иррационального поворота R_p отрезки I_n длины l_n , добавим к окружности S^1 интервалы I_n , упорядоченные точно так же, как точки x_n , таким образом, чтобы расстояние между любыми двумя такими отрезками I_m и I_n равнялось

$$\left(1 - \sum_{n \in \mathbb{Z}} l_n\right) d(x_m, x_n) + \sum_{x_k \in (x_m, x_n)} l_k.$$

(Это сумма длин добавленных отрезков l_k и длины дуги окружности между точками x_m и x_n , причем масштаб выбирается таким образом, чтобы общая длина всех составляющих множества $S^1 \setminus \bigcup_{n \in \mathbb{Z}} I_n$ равнялась $1 - \sum_{n \in \mathbb{Z}} l_n$.) Чтобы найти такой гомеоморфизм окружности f , что $f(I_n) = I_{n+1}$ и ограничение $f|_{S^1 \setminus \bigcup_{n \in \mathbb{Z}} I_n}$ является полусопряженным с поворотом, достаточно определить производную $f'(x)$,

поскольку тогда мы сможем восстановить f при помощи интегрирования.

Определим на отрезке $[a, a + l]$ функцию

$$h(a, l, x) := 1 - \frac{1}{l} |2(x - a) - l|.$$

Нетрудно видеть, что $h(a, l, a + l/2) = 1$ и $\int_a^{a+l} h(a, l, x) dx = l/2$. Обозначим левый конец отрезка I_n через a_n и положим

$$f'(x) = \begin{cases} 1, & \text{если } x \in S^1 \setminus \bigcup_{n \in \mathbb{Z}} I_n, \\ 1 + c_n h(a_n, l_n, x), & \text{если } x \in I_n. \end{cases}$$

Выбрав $c_n = 2((l_{n+1}/l_n) - 1) = 2(l_{n+1} - l_n)/l_n$, мы получим

$$\int_{I_n} f'(x) dx = \int_{I_n} (1 + c_n h(a_n, l_n, x)) dx = l_n + \frac{l_n}{2} c_n = l_{n+1}.$$

а следовательно, $f(I_n) = I_{n+1}$, как и требовалось. $\square$

Внимательно изучив это доказательство, можно заметить, что если мы хотим, чтобы отрезки попали в промежутки между элементами универсального канторова множества, производная функции f должна вести себя не совсем регулярным образом. Систематический тщательный анализ показывает, что для достаточно гладких гомеоморфизмов окружности такое явление не наблюдается.

Диффеоморфизм $f: S^1 \rightarrow S^1$ класса C^2 с иррациональным числом вращения $\rho(f) \in \mathbb{R} \setminus \mathbb{Q}$ транзитивен и, следовательно, топологически сопряжен с поворотом $R_{\rho(f)}$. На самом деле можно немного ослабить предположение о регулярности. Наиболее естественно допустить, что производная имеет ограниченную вариацию. Функция $g: S^1 \rightarrow \mathbb{R}$ имеет ограниченную вариацию, если ее полная вариация $\text{Var}(g) := \sup \sum_{k=1}^n |g(x_k) - g(x'_k)|$ конечна. Здесь точная верхняя грань берется по всем таким конечным наборам $\{x_k, x'_k\}_{k=1}^n$, что x_k, x'_k — концы интервала I_k и $I_k \cap I_j = \emptyset$ при $k \neq j$. Каждая липшицева функция, а следовательно, каждая непрерывно дифференцируемая функция имеет ограниченную вариацию.

4.4.4. Зависимость числа вращения от параметра. Теперь мы исследуем зависимость числа вращения от отображения при изменении отображения. Для начала покажем, что число вращения непрерывно и монотонно.

Предложение 4.4.5. *Отображение $\rho(\cdot)$ непрерывно в равномерной топологии.*

Доказательство. Если $\rho(f) = \rho$, возьмем такие числа $p'/q', p/q \in \mathbb{Q}$, что $p'/q' < \rho < p/q$. Выберем такое поднятие F отображения f , что $-1 < F^q(x) - x - p \leq 0$ для некоторого $x \in \mathbb{R}$. Тогда $F^q(x) < x + p$ для всех $x \in \mathbb{R}$, так как иначе по теореме о промежуточном значении для некоторого $x \in \mathbb{R}$ выполнялось бы равенство $F^q(x) = x + p$ и мы бы получили, что $\rho = p/q$. Так как функция $F^q - \text{Id}$ периодическая и непрерывная, она достигает своего максимума. Таким образом, существует такое $\delta > 0$, что $F^q(x) < x + p - \delta$ для всех $x \in \mathbb{R}$. Из этого следует, что любое достаточно малое возмущение $\tilde{F}$ поднятия F в равномерной топологии также удовлетворяет условию $\tilde{F}^q(x) < x + p$ для всех $x \in \mathbb{R}$, и, таким образом, $\rho(\tilde{F}) < p/q$. Аналогичное рассуждение для p'/q' завершает доказательство. $\square$

Из определения числа вращения следует также и монотонность: если $F_1 > F_2$, то $\rho(F_1) \geq \rho(F_2)$. Это приводит к следующим понятиям упорядочения на окружности и для отображений окружности.

Определение 4.4.6. Определим отношение « $\prec$ » на S^1 следующим образом: $[x] < [y] :\Leftrightarrow y - x \in (0, 1/2) \pmod{1}$ и определим частичное упорядочение « $\prec$ » на множестве сохраняющих ориентацию гомеоморфизмов окружности: $f_0 \prec f_1 :\Leftrightarrow f_0(x) < f_1(x)$ для всех $x \in S^1$.

Заметим, что ни одно из этих отношений не транзитивно. Действительно, $[0] < [1/3] < [2/3] < [0]$, и соответственно $R_0 \prec R_{1/3} \prec R_{2/3} \prec R_0$, где R_α — поворот, определенный в § 4.1. Теперь из определения числа вращения немедленно следует такой результат.

Предложение 4.4.7. *Отображение $\rho(\cdot)$ монотонно: если $f_1 \prec f_2$, то $\rho(f_1) \leq \rho(f_2)$.*

Замечание 4.4.8. В частности, если $\{f_t\}$ — такое семейство сохраняющих ориентацию гомеоморфизмов окружности, что $f_t(x)$ возрастает по t для каждого $x \in \mathbb{R}$, то $\rho(f_t)$ не убывает по t .

Для иррациональных значений число вращения строго возрастает.

Предложение 4.4.9. *Если $f_0 \prec f_1$ и $\rho(f_0) \notin \mathbb{Q}$, то $\rho(f_0) < \rho(f_1)$.*

Доказательство. Если F_0 и F_1 — такие поднятия, что $0 < F_1(x) - F_0(x) < 1/2$ для всех $x \in \mathbb{R}$, то в силу непрерывности и периодичности $F_1(x) - F_0(x) > \delta$ для некоторого $\delta > 0$ и всех $x \in \mathbb{R}$. Возьмем такое число $p/q \in \mathbb{Q}$, что $p/q - \delta/q < \rho(F_0) < p/q$. Тогда существует такое $x_0 \in \mathbb{R}$, что $F_0^q(x_0) - x_0 > p - \delta$, так как в противном случае мы бы получили, что $\rho(F_0) = \lim_{n \rightarrow \infty} \frac{F_0^{nq}(x) - x}{nq} \leq \lim_{n \rightarrow \infty} \frac{n(p - \delta)}{nq} = p/q - \delta/q$. Поскольку

$$F_1^q(x_0) = F_1(F_1^{q-1}(x_0)) > F_0(F_1^{q-1}(x_0)) + \delta > F_0(F_0^{q-1}(x_0)) + \delta = F_0^q(x_0) + \delta > x_0 + p,$$

мы имеем или $F_1^q(x) > x + p$ для всех $x \in \mathbb{R}$, или же $F_1^q(x_1) = x_1 + p$ для некоторого $x_1 \in \mathbb{R}$. В любом случае $\rho(F_0) < p/q \leq \rho(F_1)$. $\square$

В то время как предложение 4.4.9 показывает, что наличие иррационального числа вращения не является устойчивым, ситуация с рациональными числами вращения иная.

Предложение 4.4.10. *Пусть $f: S^1 \rightarrow S^1$ — сохраняющий ориентацию гомеоморфизм с рациональным числом вращения $\rho(f) = p/q$, имеющий непериодические точки. Тогда все достаточно близкие возмущения $\tilde{f}$ гомеоморфизма f , удовлетворяющие условию $\tilde{f} \prec f$, или все достаточно близкие возмущения $\bar{f}$, удовлетворяющие условию $f \prec \bar{f}$ (или и те, и другие), имеют число вращения p/q .*

Основной вопрос здесь состоит в том, имеет ли график функции $F^q - p$ части выше и ниже диагонали, так как в этом случае при любых малых возмущениях мы не можем избавиться от пересечений с диагональю (рис. 4.4.2). Пограничный случай, когда график полностью находится с одной стороны от диагонали, соответствует ситуации, когда в динамике происходят бифуркации (см. также рис. 2.3.2).

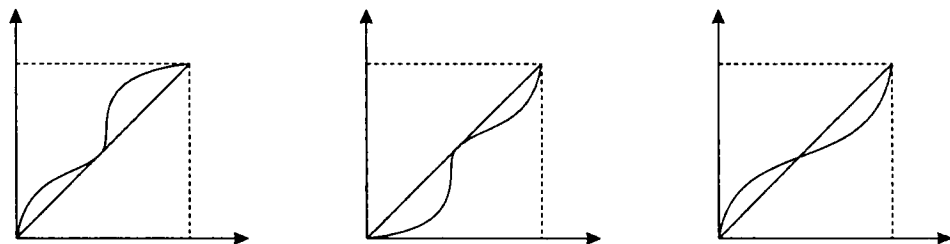


Рис. 4.4.2. Односторонняя и двусторонняя устойчивость

Доказательство. Поскольку гомеоморфизм f имеет непериодические точки, функция $F^q - \text{Id} - p$ не обращается тождественно в нуль для любого поднятия F отображения f . (Она имеет нули по предположению.) Если существует такое число $x \in \mathbb{R}$, что $F^q(x) - x - p > 0$, то для любого достаточно малого возмущения $\tilde{f} < f$ его поднятие $\tilde{F}$ таково, что $\tilde{F}^q(x) - x - p > 0$, и, следовательно, $\rho(\tilde{f}) \geq p/q$, а значит, $\rho(f) = p/q$ в соответствии с предложением 4.4.7.

В противном случае то же самое справедливо для таких возмущений $\tilde{f}$ гомеоморфизма f , что $f < \tilde{f}$. $\square$

Замечание 4.4.11. Как видно из приведенного доказательства, отображения окружности, которые имеют притягивающую или отталкивающую периодическую орбиту (т.е. такую орбиту, что в некоторой точке поднятия функция $F^q - \text{Id} - p$ меняет знак), могут быть возмущены (в любом направлении) без изменения числа вращения.

С другой стороны, если функция $F^q - \text{Id} - p$ не изменяет знак, например $F^q - \text{Id} - p \geq 0$, то любое такое возмущение $\tilde{f}$ отображения f , что $f < \tilde{f}$, будет иметь число вращения $\rho(\tilde{f}) > p/q$, поскольку $\tilde{F}^q - \text{Id} - p \geq \delta > 0$. В этом случае нули функции $F^q - \text{Id} - p$ проектируются на «параболические» или *полуустойчивые* периодические орбиты. Это орбиты p , которые притягивают с одной стороны и отталкивают с другой, т.е. имеется такая открытая окрестность U точки p , что для всех точек x , лежащих в одной компоненте множества $U \setminus \{p\}$, мы имеем $\lim_{n \rightarrow \infty} d(f^n(x), f^n(p)) = 0$, а для всех точек x , лежащих в другой компоненте, $\lim_{n \rightarrow -\infty} d(f^n(x), f^n(p)) = 0$ (см. рис. 4.3.3).

Рассмотрим следующий крайний случай.

Лемма 4.4.12. Если все точки отображения $f: S^1 \rightarrow S^1$ периодические, то число вращения строго возрастает по f .

Чтобы убедиться, что число вращения зависит от f очень негладким образом, переформулируем эти заключения следующим образом: число вращения как функция параметра может являться канторовой лестницей (см. определение 4.4.1), и, более того, обычно так и происходит.

Предложение 4.4.13. Предположим, что $\{f_t\}_{t \in [0,1]}$ — такое монотонное непрерывное семейство сохраняющих ориентацию гомеоморфизмов окружности, что отображение $\rho: t \mapsto \rho(f_t)$ непостоянно. Если существует такое плотное множество $S \subset \mathbb{Q}$, что для всякого отображения f_t либо $\rho(f_t) \notin S$, либо f_t имеет непериодические точки, то ρ является канторовой лестницей.

Доказательство. Согласно предложениям 4.4.7 и 4.4.5 отображение ρ монотонно и непрерывно, откуда с учетом предложения 4.4.10 мы заключаем, что $\rho^{-1}(S)$ — объединение непересекающихся отрезков положительной длины.

Нам нужно показать, что множество $\rho^{-1}(S)$ плотно. Увеличивая в случае необходимости множество S , мы можем считать, что если $\rho(f_t) = p/q \in \mathbb{Q} \setminus S$, то f_t имеет только периодические точки. Тогда из предложения 4.4.9 и леммы 4.4.12 следует, что отображение ρ строго монотонно в точках $t \in \rho^{-1}([0, 1] \setminus S)$. Таким образом, для $t \in [0, 1] \setminus \rho^{-1}(S)$ и $\epsilon > 0$ мы имеем $\rho(t) \neq \rho(t + \epsilon)$, и, следовательно, так как множество S плотно, а отображение ρ непрерывно, по теореме о промежуточном значении существует точка $t_1 \in \rho^{-1}(S) \cap [t, t + \epsilon]$. Это доказывает, что множество $\rho^{-1}(S)$ плотно. $\square$

В завершение мы отметим, что приведенные в этом параграфе результаты зависят от монотонности и непрерывности отображения f , но не от обратимости. Таким образом, достаточно предположить, что $f: S^1 \rightarrow S^1$ — непрерывное сохраняющее упорядочение отображение степени один, т.е. что его поднятие F неубывающее и $F(x+1) = F(x) + 1$ (см. упражнение 4.3.5). Такое отображение может принимать постоянные значения на конечном или счетном множестве отрезков.

4.4.5. Синхронизация частот. С использованием результатов относительно зависимости числа вращения от параметра, полученных в предыдущем пункте, можно также сделать некоторые выводы относительно потоков на двумерном торе и, в частности, относительно некоторых систем дифференциальных уравнений, возникающих в приложениях. Теперь мы можем обратить внимание на то

явление, что спаренные осцилляторы имеют тенденцию к *синхронизации*, т. е. их частоты со временем совпадут или по крайней мере будут связаны рациональным соотношением.

Вспомним, что в п. 1.2.10 мы рассматривали задачи о свечении светлячков и о суточных ритмах как примеры ситуаций, которые можно моделировать как спаренные осцилляторы. В то же время мы можем попросту моделировать эти биологические часы как гармонический осциллятор или его близкий аналог. Действительно, как мы увидим в п. 6.2.2, при линеаризации гармонический осциллятор служит хорошей отправной точкой.

Как мы видели в п. 4.2.4, два независимых гармонических осциллятора на совместном множестве уровня дают линейный поток на торе. Этот линейный поток на $\mathbb{T}^2$ может быть описан как результат решения системы дифференциальных уравнений

$$\dot{x}_1 = \omega_1,$$

$$\dot{x}_2 = \omega_2.$$

Теперь мы хотим понять, что произойдет при спаривании двух осцилляторов. Для этого несколько изменим приведенные выше дифференциальные уравнения, включив в них «смешанные» члены:

$$\begin{aligned}\dot{x}_1 &= \omega_1 + c_1 \sin 2\pi(x_2 - x_1), \\ \dot{x}_2 &= \omega_2 + c_2 \sin 2\pi(x_1 - x_2).\end{aligned}\tag{4.4.1}$$

Эта модель отличается от спаривания исходных уравнений второго порядка из п. 4.2.4, но, тем не менее, она позволяет получить хорошее представление о системе.

Некоторая тонкость состоит в выборе синусов для получения смешанных членов, поскольку обе переменные определены только по модулю 1. Постоянные c_1 и c_2 указывают силу сцепления осцилляторов. Когда они равны нулю, нет никакого сцепления и мы вновь приходим к линейному потоку на двумерном торе. Если они положительны, то правая часть действует так, чтобы увеличить более медленную из частот ω_1 и ω_2 и уменьшить более быструю, что может привести к синхронизации.

В п. 4.2.3 мы изучали потоки на двумерном торе, рассматривая отображение последования потока системы (4.4.1), скажем, для $x_2 = 0$. В отсутствии членов, отвечающих за сцепление, т. е. при $c_1 = c_2 = 0$, это отображение последования представляет собой попросту поворот с числом вращения ω_1/ω_2 . Таким образом, для малых значений постоянных сцепления отображение последования есть возмущение этого поворота. В «большинстве» случаев это возмущение имеет рациональное число вращения, поскольку это довольно устойчивая ситуация. И если число вращения рационально, то асимптотическое поведение обязательно периодическое (с тем же периодом).

Для более детального исследования предположим, что частоты ω_1 и ω_2 близки к друг другу. Сначала положим $\omega_1 = \omega_2 =: \omega$. В этом случае система (4.4.1) имеет решение $x(t) = y(t) = \omega t$ независимо от значений c_1 и c_2 , так что отображение по-

следования всегда имеет неподвижную точку, и, следовательно, число вращения равно 0.

Если $(c_1, c_2) \neq (0, 0)$, то получающиеся в результате отображения не сопряжены с поворотами, и поэтому их число вращения сохраняется при малых возмущениях в силу предложения 4.4.10. В частности, если зафиксировать c_1 и c_2 , то для малых значений $\omega_1 - \omega_2$ поток системы (4.4.1) имеет отображение последования с неподвижной точкой, все орбиты которого асимптотически стремятся к неподвижной (или сами неподвижны). Это означает, что все соответствующие решения системы (4.4.1) асимптотически стремятся к периодическому решению с равными частотами, т. е. на практике две априорно различные частоты осцилляторов при спаривании переходят в общую компромиссную частоту. Таким образом, два осциллятора синхронизируются, если их частоты достаточно близки друг к другу.

Заметим, кстати, что должен иметься сдвиг по фазе (т. е. $x - y$), величина которого сравнима с разностью между частотами.

Упражнения

4.4.1. Рассмотрим множество $\mathcal{C}$ таких монотонных функций $f: [0, 1] \rightarrow [0, 1]$, что $f(0) = 0$ и $f(1) = 1$, и определим на этом множестве следующее отображение T . Зафиксируем $f \in \mathcal{C}$ и обозначим график функции f через G . Пусть G_1 — образ графика G при преобразовании плоскости, заданном матрицей $\begin{pmatrix} 1/3 & 0 \\ 0 & 1/2 \end{pmatrix}$, и пусть $G_2 = G_1 + (2/3, 1/2)$. Обозначим через G' объединение множеств G_1, G_2 и отрезка прямой, соединяющего точки $(1/3, 1/2)$ и $(2/3, 1/2)$. Докажите, что G' является графиком некоторой функции $f' \in \mathcal{C}$, и положите $T(f) = f'$. Покажите также, что отображение T является сжимающим на множестве $\mathcal{C}$ относительно нормы равномерной сходимости. (Поскольку $\mathcal{C}$ является замкнутым подмножеством полного пространства, можно применять принцип сжимающих отображений. Так как канторова лестница является неподвижной точкой, мы получаем эффективную аппроксимационную процедуру, которая, в частности, была использована при создании рис. 4.4.1.)

4.4.2. Докажите лемму 4.4.12.

Задачи для углубленного изучения

Следующие несколько упражнений составляют краткое введение в теорию, позволяющую более детально объяснить, как происходит синхронизация частот для типичных примеров.

4.4.3. (Языки Арнольда.) Пусть $a, b \in [0, 1]$. Определим отображения $f_{a,b}: S^1 \rightarrow S^1$, $x \mapsto x + a + b \sin 2\pi x \pmod{1}$. Покажите, что для $p/q \in \mathbb{Q} \cap [0, 1]$ области $A_{p/q} := \{(a, b) \in [0, 1] \times [0, 1] : \rho(f_{a,b}) = p/q\}$ замкнуты. Покажите также, что эти области пересекают множество $[0, 1] \times \{0\}$ в точке p/q (см. рис. 4.4.3).

4.4.4. Покажите, что множество $A_{p/q}$ пересекает всякую линию $b = \text{const}$ по отрезку ненулевой длины во всех случаях, кроме случая $b = 0$.

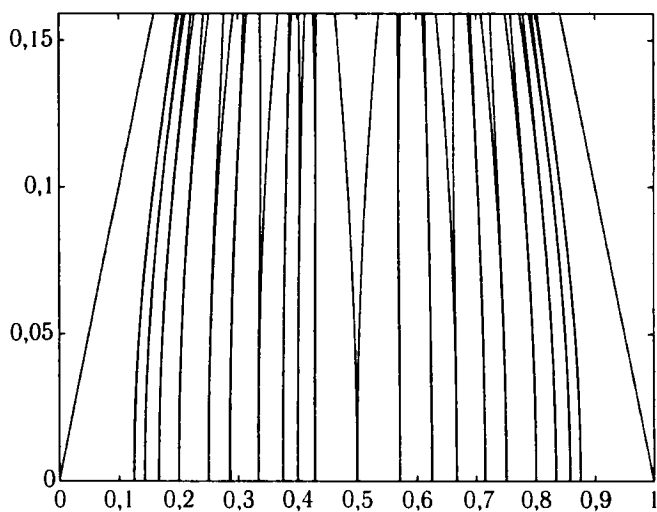


Рис. 4.4.3. Языки Арнольда

4.4.5. Связны ли множества $A_{p/q}$?

4.4.6. Покажите, что объединение множеств $A_{p/q}$ плотно в $[0, 1] \times [0, 1]$.

РЕКУРРЕНТНОСТЬ И РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ В ВЫСШИХ РАЗМЕРНОСТЯХ

§ 5.1. Сдвиги и линейные потоки на торе

5.1.1. Тор. Рассмотрим n -мерный тор

$$\mathbb{T}^n = \underbrace{S^1 \times \dots \times S^1}_{n \text{ раз}} = \mathbb{R}^n / \mathbb{Z}^n = \underbrace{\mathbb{R}/\mathbb{Z} \times \dots \times \mathbb{R}/\mathbb{Z}}_{n \text{ раз}},$$

который впервые упоминался в п. 2.6.4 для случая $n = 2$. Мы уже встречались с несколькими различными способами представления двумерного тора. Теперь рассмотрим более подробно n -мерный тор. Естественной фундаментальной областью для $\mathbb{R}^n / \mathbb{Z}^n$ является единичный куб

$$I^n = \{(x_1, \dots, x_n) \in \mathbb{R}^n : 0 \leq x_i \leq 1 \text{ для } i = 1, \dots, n\}.$$

Для получения тора необходимо отождествить противоположные грани куба I^n таким образом, чтобы точка $(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n)$ отождествлялась с $(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n)$ (см. рис. 5.1.1), поскольку эти две точки представляют один и тот же элемент тора. Подобно случаю окружности, на $\mathbb{T}^n$ имеются две удобные системы координат, а именно:

1) мультипликативная, в которой элементы тора $\mathbb{T}^n$ представлены как точки $(z_1, \dots, z_n)$, где $z_i \in \mathbb{C}$ и $|z_i| = 1$, $i = 1, \dots, n$, и

2) аддитивная, в которой элементы тора представляются как n -векторы $(x_1, \dots, x_n)$, где каждая координата определена по модулю 1.

Соответствие $(x_1, \dots, x_n) \mapsto (e^{2\pi i x_1}, \dots, e^{2\pi i x_n})$ задает изоморфизм между этими двумя представлениями. Очень часто эти системы координат называются мультипликативной и аддитивной соответственно, поскольку имеется групповая структура на торе, которая может рассматриваться как умножение или как сложение: для любых двух элементов $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$ имеется элемент $x + y$, определенный соотношением $x + y = (x_1 + y_1, \dots, x_n + y_n)$ (в аддитивной записи), и мы можем определить противоположные элементы относительно этого сложения точно так же, как в $\mathbb{R}^n$. В мультипликативной системе обозначений можно определить ту же самую структуру, если брать покомпонентные произведения, а обратные элементы в этом случае — это обычные обратные числа. Аддитивная структура представляет собой обычное сложение по модулю 1, «унаследованное» из $\mathbb{R}^n$ при отождествлении векторов по модулю 1. (Таким образом, тор — это «факторгруппа» аддитивной группы $\mathbb{R}^n$.)

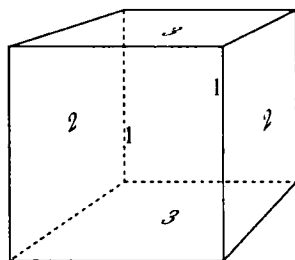


Рис. 5.1.1. Трехмерный тор

В аддитивной записи возьмем $\gamma = (\gamma_1, \dots, \gamma_n) \in \mathbb{T}^n$. В качестве естественного обобщения поворота на многомерный случай рассмотрим сдвиг T_γ , заданный соотношением

$$T_\gamma(x_1, \dots, x_n) = (x_1 + \gamma_1, \dots, x_n + \gamma_n) \pmod{1}.$$

Если все координаты вектора γ — рациональные числа, т. е., скажем, $\gamma_i = p_i/q_i$, где числа p_i и q_i взаимно просты для каждого $i = 1, \dots, n$, то отображение T_γ периодическое. Минимальный период равен наименьшему общему кратному знаменателей $q_1, \dots, q_n$.

Однако, в отличие от случаев окружности и линейных потоков на двумерном торе, минимальность не является единственной альтернативой периодичности. Например, если $n = 2$ и $\gamma = (\alpha, 0)$, где α — иррациональное число, тогда тор $\mathbb{T}^2$ распадается на инвариантные окружности $x_2 = \text{const}$, и каждая орбита остается на одной из этих окружностей и заполняет ее плотно.

5.1.2. Критерий минимальности. Условием минимальности (см. определение 4.1.4 и предложение 4.1.1) сдвига T_γ служит своего рода взаимная иррациональность координат вектора переноса γ . А именно, числа $\gamma_1, \dots, \gamma_n$ и 1 должны быть *рационально независимы*.

Определение 5.1.1. Множество $A \subset \mathbb{R}$ называется *рационально независимым*, если для любых точек $x_1, \dots, x_n \in A$ и любого вектора $(k_1, \dots, k_n) \in \mathbb{Z}^{n+1} \setminus \{0\}$ выполняется неравенство $\sum_{i=1}^n k_i \gamma_i \neq 0$.

Рациональная независимость чисел $\gamma_1, \dots, \gamma_n$ и 1 означает, что $k_0 + \sum_{i=1}^n k_i \gamma_i \neq 0$ для $(k_0, k_1, \dots, k_n) \in \mathbb{Z}^{n+1} \setminus \{0\}$, или, что эквивалентно, $\sum_{i=1}^n k_i \gamma_i \notin \mathbb{Z}$ для любого набора целых чисел $k_1, \dots, k_n$, кроме $k_1 = k_2 = \dots = k_n = 0$. Заметим, что в случае одного числа это понятие в точности совпадает с иррациональностью.

Предложение 5.1.2. Сдвиг T_γ на $\mathbb{T}^2$ минимален тогда и только тогда, когда числа γ_1, γ_2 и 1 рационально независимы, т. е. не существует отличной от нуля пары целых чисел k_1, k_2 , для которой $k_1 \gamma_1 + k_2 \gamma_2 \in \mathbb{Z}$.

Мы докажем этот критерий минимальности ниже (см. п. 5.1.5), поскольку используемые при доказательстве рассуждения хотя и элементарны, но все же довольно громоздки. Мы приведем доказательство только для размерности 2, дабы немного упростить обозначения и рассуждения и иметь возможность использовать наглядные геометрические рисунки.

Тем не менее, уже сейчас нетрудно понять, почему возникает такое ограничение на вектор сдвига. В п. 4.2.3 мы видели, что линейный поток $(T_\gamma^t)_{t \in \mathbb{R}}$ на двумерном торе минимален, если вектор сдвига γ имеет иррациональный коэффициент наклона. Поэтому сдвиг T_γ может быть минимален, только если $\gamma_1/\gamma_2 \notin \mathbb{Q}$, или $k_1 \gamma_1 + k_2 \gamma_2 \neq 0$ для любой пары $(k_1, k_2) \in \mathbb{Z}^2 \setminus \{0\}$. С другой стороны, это условие не является достаточным, поскольку если, например, число γ_1 рационально, то первая координата любой орбиты может принимать только

конечное множество значений и орбита не может быть плотной. Приведенное выше условие минимальности позволяет избежать таких проблем.

5.1.3. Линейные потоки. В п. 4.2.3 мы определили линейные потоки на двумерном торе. Линейные потоки на n -мерном торе задаются аналогично — как однопараметрическая группа сдвигов

$$T_{\omega}^t(x_1, \dots, x_n) = (x_1 + t\omega_1, \dots, x_n + t\omega_n) \pmod{1}.$$

Очевидно, что поток $\{T_{\omega}^t\}$ минимален, если для некоторого t_0 преобразование $T_{\omega}^{t_0}$ минимально. Это наблюдение вместе с предложением 5.1.2 позволяет нам установить критерий минимальности для этого случая.

Предложение 5.1.3. *Поток $\{T_{\omega}^t\}$ минимален тогда и только тогда, когда числа $\omega_1, \dots, \omega_n$ рационально независимы.*

Доказательство. Поскольку $T_{\omega}^t = T_{t\omega}$, минимальность будет следовать из предложения 5.1.2, если только мы покажем, что имеется такое действительное число t , что для любого целочисленного вектора $(k_1, \dots, k_n)$, отличного от нуля, сумма $\sum_{i=1}^n tk_i\omega_i$ не является целым числом. Чтобы убедиться в этом, заметим, что если $k \in \mathbb{Z}$, то из соотношения $s \sum_{i=1}^n k_i\omega_i = k$ следует, что $s = k / (\sum k_i\omega_i)$ (поскольку $\sum_{i=1}^n k_i\omega_i \neq 0$ в силу рациональной независимости). Имеется лишь счетное множество таких значений s , так что нам подойдет любое число $t \in \mathbb{R} \setminus \{k / \sum k_i\omega_i : k_1, \dots, k_n, k \in \mathbb{Z}, (\sum k_i\omega_i) \neq 0\}$.

С другой стороны, если $\sum_{i=1}^n k_i\omega_i = 0$ для некоторого ненулевого вектора $(k_1, \dots, k_n)$, тогда функция $\varphi(x) = \sin 2\pi \left(\sum_{i=1}^n k_i x_i \right)$ непрерывна, непостоянна и инварианта относительно потока $\{T_{\omega}^t\}$. Таким образом, поток не минимален, поскольку $\varphi^{-1}([0, 1])$ — замкнутое инвариантное множество (см. определение 4.1.4). $\square$

5.1.4. Равномерное распределение: элементарное доказательство. Так же как и в п. 4.1.4, мы можем рассмотреть частоты, с которыми орбита минимального сдвига посещает различные части тора. В одномерном случае мы использовали дуги (отрезки) как естественные «окна» для измерения частоты посещений. Естественным обобщением для n -мерного тора будут n -мерные параллелепипеды $\Delta = \Delta_1 \times \dots \times \Delta_n$, где $\Delta_1, \dots, \Delta_n$ — дуги. Для $n = 2$ естественно называть параллелепипед *прямоугольником*. Объем $\text{vol}(\Delta)$ параллелепипеда Δ определяется как произведение длин дуг $\Delta_1, \dots, \Delta_n$. Таким образом, мы получаем следующее естественное обобщение понятия равномерного распределения, которое было введено в замечании после предложения 4.1.7.

Определение 5.1.4. Последовательность $(x_m)_{m \in \mathbb{N}}$ в T^n называется *равномерно распределенной*, если

$$\lim_{m \rightarrow \infty} \frac{\text{card}\{k \in \{1, \dots, m\} : x_k \in \Delta\}}{m} = \text{vol}(\Delta)$$

для каждого n -мерного параллелепипеда $\Delta \subset T^n$.

При доказательстве предложения 4.1.7 используется только то обстоятельство, что иррациональный поворот минимален и изометричен. Точнее говоря, необходимо, чтобы дуги переходили в дуги той же самой длины и чтобы окружность можно было разложить в объединение произвольно малых изометричных дуг с непересекающимися внутренними частями. Сдвиги на торе являются изометриями, и аналог свойства разложения окружности справедлив для тора, если заменить дуги на параллелепипеды, так что не удивительно, что доказательство предложения 4.1.7 может быть использовано и в этой ситуации. Мы приведем детальное рассуждение в двумерном случае. В отличие от доказательства минимальности, где подобное ограничение приводит к сильному упрощению, в данном случае это просто вопрос простоты обозначений. Результат о том, что из минимальности следует равномерность распределения, можно самым непосредственным образом обобщить на случай сдвигов в произвольной размерности.

Теорема 5.1.5. *Если числа $\gamma_1, \gamma_2, 1$ рационально независимы, тогда каждая полуорбита сдвига $T_{(\gamma_1, \gamma_2)}$ равномерно распределена.*

Как и в п. 4.1.4, определим

$$F_{\Delta}(x, n) := \text{card}\{k \in \mathbb{Z}: 0 \leq k < n, T_{\gamma}^k(x) \in \Delta\}$$

для любой точки $x \in \mathbb{T}^2$ и любого прямоугольника Δ . Это определение распространяется на объединения непересекающихся прямоугольников. Положим $\gamma = (\gamma_1, \gamma_2)$. В силу предложения 5.1.2 сдвиг T_{γ} минимален. Это позволяет обобщить предложение 4.1.5.

Предложение 5.1.6. *Пусть $\Delta = \Delta_1 \times \Delta_2$, $\Delta' = \Delta'_1 \times \Delta'_2$ — два таких прямоугольника, что $\ell(\Delta_i) < \ell(\Delta'_i)$, $i = 1, 2$. Тогда существует такое число $N_0 \in \mathbb{N}$, зависящее от Δ, Δ' и γ , что для всех $x \in \mathbb{T}^2$, $N \geq N_0$ и $n \in \mathbb{N}$ выполняется неравенство*

$$F_{\Delta'}(x, n + N) \geq F_{\Delta}(x, n).$$

Доказательство. По предположению существует сдвиг T_{β} прямоугольника Δ , который лежит внутри Δ' . В силу минимальности сдвига T_{γ} мы можем найти число $N_0 \in \mathbb{N}$, для которого образы $T_{\gamma}^{N_0} \Delta$ и $T_{\beta} \Delta$ настолько близки, что $T_{\gamma}^{N_0} \Delta \subset \Delta'$. Таким образом, если $T_{\gamma}^n(x) \in \Delta$, то $T_{\gamma}^{n+N_0}(x) \in \Delta'$ и $F_{\Delta'}(x, n + N) \geq F_{\Delta'}(x, n + N_0) \geq F_{\Delta}(x, n)$ для $n \geq N_0$. $\square$

Доказательство теоремы 5.1.5. Теперь, подобно одномерному случаю, рассмотрим прямоугольник $\Delta = \Delta_1 \times \Delta_2$, где $\ell(\Delta_1) = \ell(\Delta_2) = 1/k$. Разделим тор $\mathbb{T}^2$ на $(k-1)^2$ непересекающихся прямоугольников, каждый из которых является произведением двух дуг длины $1/(k-1)$ (см. рис. 5.1.2), и применим предложение 5.1.6, так же как мы это сделали при доказательстве леммы 4.1.9. Мы получаем

$$\bar{f}(\Delta) := \overline{\lim}_{n \rightarrow \infty} \frac{F_{\Delta}(x, n)}{n} \leq 1/(k-1)^2. \quad (5.1.1)$$

Наконец, пусть $\Delta = \Delta_1 \times \Delta_2$ — произвольный прямоугольник. Зафиксируем $\varepsilon > 0$ и найдем такой прямоугольник $\Delta' = \Delta'_1 \times \Delta'_2$, что $\Delta_i \subset \Delta'_i$ для $i = 1, 2$,

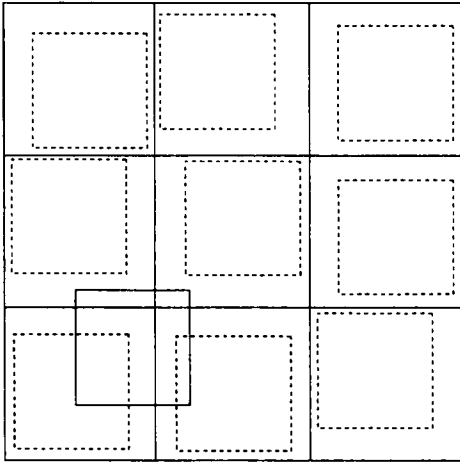


Рис. 5.1.2. Тор с прямоугольниками

длины отрезков Δ'_i равны l_i/k и $\text{vol } \Delta' < \text{vol } \Delta + \varepsilon$. Используя неравенство (5.1.1) и субаддитивность отображения $\bar{f}$, мы получаем

$$\bar{f}(\Delta) \leq \bar{f}(\Delta') \leq \left(\frac{k}{k-1}\right)^2 \text{vol } \Delta' < \left(\frac{k}{k-1}\right)^2 (\text{vol } \Delta + \varepsilon).$$

Поскольку ε произвольно мало, а k произвольно велико, из этого следует, что $\bar{f}(\Delta) \leq \text{vol } \Delta$ для любого прямоугольника Δ , а значит, в силу субаддитивности отображения $\bar{f}$, и для любого конечного объединения непересекающихся прямоугольников. В частности, так как $\mathbb{T}^2 \setminus \Delta$ — объединение трех непересекающихся прямоугольников, мы получаем

$$\underline{f}(\Delta) := \lim_{n \rightarrow \infty} \frac{F_{\Delta}(x, n)}{n} = 1 - \bar{f}(\mathbb{T}^2 \setminus \Delta) \geq 1 - \text{vol}(\mathbb{T}^2 \setminus \Delta) = \text{vol } \Delta,$$

откуда следует, что $\underline{f}(\Delta) = \bar{f}(\Delta) = \text{vol } \Delta$. □

Приведем очевидное обобщение теоремы 4.1.15 со случая поворотов окружности на сдвиги тора.

Теорема 5.1.7. Пусть $\gamma = (\gamma_1, \gamma_2)$ и φ — произвольная интегрируемая по Риману функция на $\mathbb{T}^2$. Если числа 1, γ_1 , γ_2 рационально независимы, то

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(T_{\gamma}^k(x_1, x_2)) = \int_{\mathbb{T}^2} \varphi(\theta_1, \theta_2) d\theta_1 d\theta_2$$

равномерно по $(x_1, x_2) \in \mathbb{T}^2$.

Доказательство. Переход от равномерного распределения для прямоугольников к равномерному распределению для непрерывных и, в более общей ситуации, интегрируемых по Риману функций осуществляется точно так же, как и в одномерном случае (см. предложение 4.1.14 и теорему 4.1.15). Для любого

прямоугольника Δ его объем определяется соотношением

$$\text{vol } \Delta = \int_{\mathbb{T}^2} \chi_{\Delta}(\theta_1, \theta_2) d\theta_1 d\theta_2,$$

и по определению функция φ интегрируема по Риману, если для любого $\varepsilon > 0$ существуют такие конечные линейные комбинации φ_1, φ_2 характеристических функций прямоугольников, что $\varphi_1 \leq \varphi \leq \varphi_2$ и

$$\int_{\mathbb{T}^2} \varphi_1(\theta_1, \theta_2) d\theta_1 d\theta_2 < \int_{\mathbb{T}^2} \varphi_2(\theta_1, \theta_2) d\theta_1 d\theta_2 + \varepsilon.$$

(В частности, любая непрерывная функция или любая ограниченная функция с конечным числом точек разрыва интегрируема по Риману.) $\square$

5.1.5. Доказательство критерия минимальности. Теперь мы докажем, что сдвиги на торе минимальны тогда и только тогда, когда вектор переноса «вполне иррациональный». Из этого условия следует иррациональность чисел γ_1 и γ_2 , а также и их отношения. В то же время оно является более сильным, в чем можно легко убедиться, если рассмотреть простой пример $\gamma_2 = 1 - \gamma_1$, где γ_1 — произвольное иррациональное число.

Доказательство будет значительно менее тривиальным, чем простое рассуждение в доказательстве предложения 4.1.1. Главная идея, однако, та же самая. Она заключается в следующем: если точки в орбите не выстроены определенным способом, они будут заполнять все окружающее пространство, откуда и будет следовать минимальность. Основное отличие от одномерного случая состоит в том, что в одномерной ситуации «специальное выстраивание» означало конечность орбиты, а следовательно, ее периодичность, в то время как теперь мы должны исследовать промежуточный случай и показать, что это происходит, только если орбиты находятся на параллельных рациональных прямых, накручивающихся на тор.

Доказательство предложения 5.1.2. Мы будем использовать аддитивную запись. Заметим сначала, что сдвиг минимален тогда и только тогда, когда орбита нуля плотна, поскольку для любой точки $x \in \mathbb{T}^2$ мы имеем

$$T_{\gamma}(x) = x + \gamma = 0 + \gamma + x = T_{\gamma}(0) + x \pmod{1},$$

т.е. орбита $\mathcal{O}(x)$ точки x получается как $T_x(\mathcal{O}(0))$, а поэтому она плотна тогда и только тогда, когда орбита $\mathcal{O}(0)$ плотна, так как T_x — гомеоморфизм. (Это рассуждение повторяет рассуждения, используемые при доказательстве более общего предложения 4.1.19.)

Выберем $\varepsilon > 0$ и рассмотрим множество D_{ε} всех итераций $T_{\gamma}^m(0)$, которые лежат в ε -шаре $B(0, \varepsilon)$ с центром в 0. Имеются две возможности:

1) для некоторого $\varepsilon > 0$ множество D_{ε} линейно зависимо (т.е. лежит на прямой);

2) для любого $\varepsilon > 0$ множество D_{ε} содержит 2 линейно независимых вектора.

Докажем три соответствующие леммы.

Лемма 5.1.8. *Из условия 2) следует минимальность.*

Лемма 5.1.9. *Из условия 1) следует рациональная зависимость.*

Лемма 5.1.10. *Из рациональной зависимости следует условие 1).*

Минимальность исключает случай 1) и, следовательно, влечет за собой выполнение условия 2), так что минимальность эквивалентна условию 2). Таким образом, минимальность $\iff$ 2) $\iff$ не 1) $\iff$ рациональная независимость.

Доказательство леммы 5.1.8. Это рассуждение подобно доказательству предложению 4.1.1, хотя и более сложное. Достаточно показать, что орбита точки 0 плотна. Возьмем $\varepsilon > 0$ и предположим, что элементы $v_1, v_2 \in D_{\varepsilon}$ линейно независимы. Это означает, что они порождают малый параллелограмм $\{av_1 + bv_2 : a, b \in [0, 1]\}$. Вершины этого параллелограмма являются частью орбиты $\mathcal{O}(0)$: это уже известно для точек 0, v_1 и v_2 , и для $v_1 + v_2$ в этом также легко убедиться, если представить v_1 и v_2 в виде $V_1 = 0 + m_1\gamma - k(m_1)$ и $V_2 = 0 + m_2\gamma - k(m_2)$

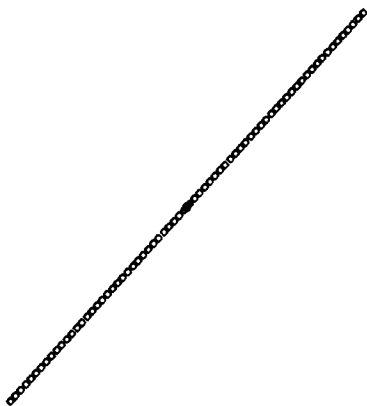


Рис. 5.1.3. Зависимость и независимость

в $\mathbb{R}^2$, где $k(m_1)$ и $k(m_2)$ — соответственно такие целочисленные векторы, для которых $\|V_1\| < \epsilon$ и $\|V_2\| < \epsilon$. Тогда $V_1 + V_2 = 0 + (m_1 + m_2)\gamma - (k(m_1) + k(m_2)) = T_Y^{m_1+m_2}(0) \pmod{1}$ и, следовательно, $v_1 + v_2 = T_Y^{m_1+m_2}(0)$.

Кроме того, орбита точки 0 содержит все целочисленные линейные комбинации элементов v_1 и v_2 , поскольку $kV_1 + lV_2 = T_Y^{km_1+lm_2}(0) \pmod{1}$. Поэтому рассмотрим замощение плоскости, определенное переносами области $R := \{aV_1 + bV_2 : a, b \in [0, 1]\}$ на целые кратные векторов V_1 и V_2 . Плоскость покрывается подобными параллелограммами, имеющими общие точки только на границах, и каждая точка плоскости находится на расстоянии не более ϵ от одной из вершин этого замощения (см. рис. 5.1.3). В частности, каждая точка квадрата $[0, 1] \times [0, 1]$ находится в ϵ -окрестности некоторой вершины, т. е. каждая точка тора $\mathbb{T}^2$ располагается на расстоянии не более ϵ от некоторой точки орбиты $O(0)$. Согласно предположению случая 2) это справедливо для любого $\epsilon > 0$, т. е. орбита $O(0)$ плотна в $\mathbb{T}^2$. $\square$

Доказательство леммы 5.1.9. Если точка 0 периодическая, то числа γ_1 и γ_2 рациональны и утверждение доказано.

Предположим, что орбита точки 0 бесконечна. Тогда для любого ϵ она содержит две такие точки $p = T_Y^m(0)$ и $q = T_Y^n(0)$, что $\|q - p\| < \epsilon$. Но тогда найдутся такие точки $P = m\gamma \in \mathbb{R}^2$ и $Q = n\gamma + k \in \mathbb{R}^2$, что $\epsilon > \|P - Q\| = \|m\gamma - n\gamma - k\| = \|(m - n)\gamma - k\|$, а это означает, что $T_Y^{m-n}(0) - k \in B(0, \epsilon)$ и $D_\epsilon \neq \{0\}$ для всех $\epsilon > 0$.

Если для $\epsilon > 0$ выполнено утверждение 1), то множество $\{0\} \neq D_{\epsilon'} \subset D_\epsilon$ линейно зависимо для всех $\epsilon' < \epsilon$. Таким образом, D_ϵ лежит на единственной прямой L , заданной уравнением $ax + by = 0$, проходящей через 0.

Утверждение. Орбита $O(0)$ плотна на проекции прямой L (см. рис. 5.1.4).

Поскольку $D_{\epsilon'} \neq \{0\}$ для всех $\epsilon' < \epsilon$, имеются точки $0 \neq p_{\epsilon'} \in D_{\epsilon'}$, а следовательно, и точки $P = n\gamma - k \in L \cap B(0, \epsilon')$ (где $n \in \mathbb{Z}, k \in \mathbb{Z}^2$). Но тогда множество $\{mP : m \in \mathbb{Z}\}$ является ϵ' -плотным в L и проектируется на $O(0)$.

Далее, a и b рационально зависимы, поскольку иначе коэффициент наклона прямой L был бы иррациональным, так что проекция этой прямой на $\mathbb{T}^2$ была бы плотна и по приведенному выше утверждению орбита $O(0)$ тоже была бы плотна. Таким образом, существует такая пара чисел $(k_1, k_2) \in \mathbb{Z}^2 \setminus \{0\}$, что $ak_1 - bk_2 = 0$. Если $a = 0$ (или $b = 0$), то $ax + by = 0 \Leftrightarrow y = 0$ (или $x = 0$). В противном случае, умножив уравнение $ax + by = 0$ на $k_1/b = k_2/a$, мы получим $k_2x + k_1y = 0$, т. е. можно считать, что $a, b \in \mathbb{Z}$. Если точка $n\gamma - k$ лежит на прямой $ax + by = 0$, то $an\gamma_1 - k_1 + bn\gamma_2 - k_2 = 0$ или $an\gamma_1 + bn\gamma_2 = k_1 + k_2$, откуда и следует рациональная зависимость. $\square$

Доказательство леммы 5.1.10. Предположим, что $k_1\gamma_1 + k_2\gamma_2 = N \in \mathbb{Z}$, и разделим это число на γ_1 , получим $\gamma_2/\gamma_1 = (N - k_1)/k_2 =: s \in \mathbb{Q}$ (если $k_2 \neq 0$), т. е. итерации $(n\gamma_1, n\gamma_2)$ точки 0 при

Рис. 5.1.4. Плотность

повторных переносах на вектор γ лежат на прямой $y = sx$ с рациональным коэффициентом наклона s . Она проектируется на тор как орбита линейного потока T_γ^t , которая, как мы выяснили в п. 4.2.3, замкнута и, следовательно, не плотна, если $\gamma_2/\gamma_1 \in \mathbb{Q}$. Поэтому орбита точки 0 при отображении T_γ также не плотна, откуда и следует условие 1). (Если $k_2 = 0$, то $k_1 \neq 0$ и то же самое рассуждение проходит после замены x на y .) □

□

5.1.6. Равномерное распределение: метод Кронекера—Вейля. Описанный в п. 4.1.6 метод Кронекера—Вейля доказательства равномерного распределения, который заключается в том, что сначала рассматриваются тригонометрические многочлены, затем непрерывные функции и, наконец, характеристические функции, также работает в более высокой размерности. Для упрощения обозначений мы снова рассмотрим двумерный случай, оставляя обобщения на произвольную размерность читателю.

Характеры (аналог которых для одномерного случая рассматривался в п. 4.1.6) теперь определяются как групповые «гомоморфизмы» тора $\mathbb{T}^2$ в S^1 , где мы рассматриваем $\mathbb{T}^2$ как аддитивную группу (как описано в начале этой главы) и S^1 рассматривается как группа комплексных чисел с абсолютным значением 1 с умножением в качестве групповой операции. Гомоморфизм — это отображение, сохраняющее структуру группы, т. е. образ суммы двух элементов есть произведение их образов. Точнее говоря, если использовать аддитивную систему обозначений для тора, то характеры имеют следующий вид:

$$c_{m_1, m_2}(x_1, x_2) = e^{2\pi i(m_1 x_1 + m_2 x_2)} = \cos 2\pi(m_1 x_1 + m_2 x_2) + i \sin 2\pi(m_1 x_1 + m_2 x_2),$$

где m_1, m_2 — произвольные целые числа. Конечные линейные комбинации характеров называются *тригонометрическими многочленами*, так как они также могут быть выражены как конечные линейные комбинации синусов и косинусов. Характеры являются *собственными функциями* для сдвигов, поскольку

$$c_{m_1, m_2}(T_\gamma(x_1, x_2)) = e^{2\pi i(m_1(x_1 + \gamma_1) + m_2(x_2 + \gamma_2))} = e^{2\pi i(m_1 \gamma_1 + m_2 \gamma_2)} c_{m_1, m_2}(x_1, x_2).$$

Важное наблюдение, которое полезно для наших целей, состоит в том, что, поскольку числа γ_1, γ_2 и 1 являются рационально независимыми, т. е. число $m_1 \gamma_1 + m_2 \gamma_2$ никогда не является целым, кроме

случая $m_1 = m_2 = 0$, собственное значение $e^{2\pi i(m_1\gamma_1 + m_2\gamma_2)}$ может быть равно 1 только в случае $m_1 = m_2 = 0$.

Тривиальный характер $c_{0,0} = 1$ не изменяется при усреднении. Для других характеров мы, как и в п. 4.1.6, используем формулу для суммы геометрической прогрессии и получаем

$$\begin{aligned} \left| \frac{1}{n} \sum_{k=0}^{n-1} c_{m_1, m_2}(T_Y^k(x_1, x_2)) \right| &= \left| \frac{1}{n} \sum_{k=0}^{n-1} e^{2\pi i k(m_1\gamma_1 + m_2\gamma_2)} \right| |c_{m_1, m_2}(x_1, x_2)| = \\ &= \left| \frac{1 - e^{2\pi i n(m_1\gamma_1 + m_2\gamma_2)}}{n(1 - e^{2\pi i(m_1\gamma_1 + m_2\gamma_2)})} \right| \leq \frac{2}{n(1 - e^{2\pi i(m_1\gamma_1 + m_2\gamma_2)})} \xrightarrow{n \rightarrow \infty} 0 = \int_{\mathbb{T}^2} c_{m_1, m_2}. \end{aligned}$$

Используя линейность интеграла, заключаем, что для любой функции φ , являющейся конечной линейной комбинацией характеров, т.е. тригонометрическим многочленом, выполняется неравенство

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(T_Y^k(x_1, x_2)) = \int_{\mathbb{T}^2} \varphi. \quad (5.1.2)$$

Теперь, применяя многомерную версию аппроксимационной теоремы Вейерштрасса, которая утверждает, что любая непрерывная функция на плоскости, периодическая с периодом 1 по обоим переменным, является равномерным пределом тригонометрических многочленов, мы можем заключить, что равенство (5.1.2) справедливо для любой непрерывной функции. Наконец, для прямоугольников, так же как и в одномерном случае, мы можем доказать непрерывность распределения, построив такие непрерывные функции φ_1, φ_2 , $\varphi_1 \leq \chi_{\Delta} \leq \varphi_2$, что $\int (\varphi_2 - \varphi_1) < \varepsilon$. Как уже было отмечено в конце п. 5.1.2, легко видеть, что если числа $1, \gamma_1$ и γ_2 рационально зависимы, то отображение T_Y не минимально. Если $m_1\gamma_1 + m_2\gamma_2 = k$, где $m_1, m_2, k \in \mathbb{Z}$ и $m_1^2 + m_2^2 > 0$, то $e^{2\pi i(m_1\gamma_1 + m_2\gamma_2)} = 1$ и значения непостоянного характера c_{m_1, m_2} не изменяются при переносе.

Заметим, что использование метода Кронекера—Вейля позволяет обойти довольно тонкое рассуждение, которое требовалось в п. 5.1.2 для доказательства условия минимальности. Применяя этот подход, мы можем доказать, что если числа γ_1, γ_2 и 1 рационально независимы, то распределение равномерно. Обобщение доказательства на случай произвольной размерности с использованием этого метода также совершенно рутинно.

Упражнения

5.1.1. Покажите, что числа $1, \sqrt{3}$ и $\sqrt{5}$ рационально независимы.

5.1.2. Предположим, что числа $n, m \in \mathbb{Z}$ таковы, что $1, \sqrt{n}$ и $\sqrt{m}$ рационально зависимы. Что в этом случае можно сказать о числах n и m ?

5.1.3. Опишите замыкания орбит для сдвига на вектор $(\alpha, 1/4 + 2\alpha)$ на $\mathbb{T}^2$, где $\alpha \notin \mathbb{Q}$.

5.1.4. (Из этого упражнения становится понятным происхождение термина «рациональная независимость».) Действительная прямая $\mathbb{R}$ может рассматриваться как линейное пространство над полем рациональных чисел $\mathbb{Q}$ (т.е. рациональные числа являются скалярами). Покажите, что множество чисел из $\mathbb{R}$ рационально независимо тогда и только тогда, когда оно рационально независимо в линейном пространстве $\mathbb{R}$ над $\mathbb{Q}$.

5.1.5. Покажите, что если все координаты вектора γ являются рациональными числами, т.е., например, $\gamma_i = p_i/q_i$, где числа p_i и q_i взаимно просты для всех $i = 1, \dots, n$, то сдвиг T_γ — периодическое отображение и его период равен наименьшему общему кратному знаменателей $q_1, \dots, q_n$.

5.1.6. Покажите, что замыкание орбиты неминимального сдвига на торе $\mathbb{T}^2$ либо конечно, либо представляет собой конечное объединение окружностей.

5.1.7. Покажите, что всякая замкнутая собственная подгруппа Γ группы $\mathbb{R}^2$ эквивалентна одной из следующих групп с точностью до линейной замены координат: $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{Z} \times \mathbb{Z}$, $\mathbb{Z} \times \mathbb{R}$.

Задачи для углубленного изучения

5.1.8. Обобщите упражнение 5.1.6 на $\mathbb{R}^n$.

5.1.9. Обобщите упражнение 5.1.7 на $\mathbb{R}^n$.

5.1.10. Приведите детальное доказательство аналога критерия минимальности из п. 5.1.5 для n -мерного случая.

5.1.11. Сформулируйте свойство равномерности распределения для сдвига на группе $\mathbb{Z}_2$ 2-адических целых чисел (упражнение 4.1.10) и докажите, что из минимальности следует, что распределение равномерно.

§ 5.2. Приложения сдвигов и линейных потоков

Приведем теперь несколько примеров динамических систем, в которых можно естественным образом проанализировать динамику, используя линейные потоки (или сдвиги) на торах.

5.2.1. Линейные отображения и потоки. Хорошее знание линейных отображений и потоков на торах позволяет описать динамику важного класса линейных систем, а именно отображений с собственными значениями, по модулю равными 1, и линейных дифференциальных уравнений с постоянными коэффициентами, матрица коэффициентов которых имеет чисто мнимые собственные значения (а следовательно, сдвиги за время T имеют собственные значения, по модулю равные 1). В п. 4.2.4 мы рассматривали частные примеры потоков в $\mathbb{R}^4$, возникающих для линейных дифференциальных уравнений с постоянными коэффициентами, и торы естественно получались как инвариантные множества, на которых мы наблюдаем линейный поток. Теперь обобщим этот пример, рассмотрев линейное отображение в $\mathbb{R}^{2m}$, собственные значения которого образуют m различных комплексно сопряженных пар $e^{\pm i\omega_j}$. Как и прежде, каждая пара соответствует двумерному инвариантному подпространству, в котором отображение действует как поворот в надлежащих координатах. Собственное подпространство и эти координаты получаются, если взять комплексный собственный вектор w_j , а затем выбрать действительные векторы $v_j = w_j + \bar{w}_j$ и $v'_j = i(w_j - \bar{w}_j)$ в качестве базиса. Прделав это для каждой пары собственных значений, мы получим базис в $\mathbb{R}^{2m}$, относительно которого отображение имеет блочно-диагональное матричное представление, причем каждый блок является (2×2) -матрицей, представляющей поворот. Это отображение оставляет инвариантными множества, заданные уравнениями $x_{2j-1}^2 + x_{2j}^2 = r_j^2$ для $j = 1, \dots, m$. Эти множества являются торами, размерность которых зависит от числа нулевых значений r_j . Точнее говоря, такой тор может быть параметризован полярными координатами $x_{2j-1} = r_j \cos \varphi_j$, $x_{2j} = r_j \sin \varphi_j$, и отображение действует поворотами, которые

переводят φ_j в $\varphi_j + v_j$. Ясно также, что если какое-то из значений r_j равно 0, это понижает размерность тора.

Поэтому, применяя наш критерий минимальности (предложение 5.1.2) к $\mathbb{T}^k$, мы заключаем, что ограничение потока на такой инвариантный тор является минимальным, если множество $\{v_j: r_j \neq 0\} \cup \{1\}$ рационально независимо.

Можно обобщить этот вывод и сделать заключения относительно действия линейного отображения внутри нейтрального пространства E^0 , определенного в (3.3.4), когда ограничение на это подпространство имеет различные собственные значения.

5.2.2. Движение свободной частицы на торе. Рассмотрим движение точечной массы на плоском торе $\mathbb{T}^n = \mathbb{R}^n / \mathbb{Z}^n$ без действия внешних сил. Движение описывается обыкновенным дифференциальным уравнением второго порядка $\ddot{x} = 0$, где вектор x определен по модулю $\mathbb{Z}^n$. Мы можем также записать

$$\dot{x} = v,$$

$$\dot{v} = 0,$$

откуда видно, что движение происходит вдоль прямых линий с постоянной скоростью, так как вектор v сохраняется. Это означает, что n компонент вектора v — *первые интегралы* (или константы) движения. Для любого заданного вектора v движение соответствует линейному потоку T'_v . Таким образом, фазовое пространство — это $\mathbb{R}^n \times \mathbb{T}^n$, причем динамика описывается следующим образом: торы $\{v\} \times \mathbb{T}^n$ инвариантны, а движение на $\{v\} \times \mathbb{T}^n$ задается как $\{v\} \times T'_v$. Этот поток также называется *геодезическим потоком* на $\mathbb{T}^n$. Геодезические — это пути, которые прокладываются на $\mathbb{T}^n$ орбитами. Они являются проекциями прямых в пространстве $\mathbb{R}^n$ на $\mathbb{T}^n$. В то время как для различных начальных скоростей v эти кривые могут быть плотны, периодичны или же не плотны и не периодичны, орбиты потока никогда не плотны в фазовом пространстве.

Один из способов изучения этого потока посредством динамической системы с дискретным временем — ограничиться рассмотрением векторов, имеющих основания на окружности $y = 0$ и направленных вверх. Каждый из этих векторов определяет орбиту потока, которая возвращается в это множество. Например, если $n = 2$ и α — котангенс угла наклона такого вектора, тогда возвращающее отображение определяется следующим образом: $(x, \alpha) \mapsto (x + \alpha, \alpha)$. Мы еще не раз встретимся с этим интегрируемым закручиванием (см. пример 6.1.2, п. 6.3.4.1).

5.2.3. Система многих частиц на отрезке. Теперь мы приведем еще одно непосредственное обобщение простой механической модели, которая обсуждалась в п. 4.2.5, а именно рассмотрим произвольное конечное число материальных точек с равными массами, перемещающихся на отрезке и упруго сталкивающихся между собой и с концами отрезка. Так как порядок материальных точек не может изменяться в ходе их движения, мы можем сделать вывод, что если обозначить положения частиц через $x_1, \dots, x_n$, то $0 \leq x_1 \leq \dots \leq x_n \leq 1$, т. е. конфигурационное пространство этой механической системы есть симплекс $T_n := \{(x_1, \dots, x_n): 0 \leq x_1 \leq \dots \leq x_n \leq 1\}$, а фазовое пространство — пространство касательных векторов с основаниями на T_n с необходимым согласованием на границе симплекса (см. рис. 5.2.1).

Рассматривая n -мерные аналоги геометрических рассуждений из п. 4.2.5, мы приходим к заключению, что эту систему можно описать как движение единственной материальной точки, которая



Рис. 5.2.1. Три материальные точки на прямой и их конфигурационное пространство

отражается от граней симплекса T_n в соответствии с n -мерным аналогом закона «угол падения равен углу отражения». Это означает, что можно определить продолжение траектории движения после удара о грань, если взять плоскость, порожденную входящей траекторией и нормальным вектором к грани, и применить двумерный закон отражения в этой плоскости. Это правило не позволяет нам изучать движения, которые включают в себя столкновения, например, с ребром или вершиной, т.е. кратные или одновременные столкновения.

Частичная развертка, которая позволила нам описать бильярд в треугольнике в терминах линейного потока на двумерном торе, применима и здесь, и фундаментальной областью является n -мерный куб с удвоенными линейными размерами, т.е. $\max |x_i| \leq 1$. Тогда $n!2^n$ отраженных копий симплекса T_n заполняют этот куб, а переносы этого куба покрывают $\mathbb{R}^n$. Таким образом, осуществляя полную развертку этого движения на T_n , мы получаем движение свободной материальной точки в $\mathbb{R}^n$. После редуцирования этого движения на фундаментальную область (куб, который мы отождествляем с n -мерным тором) мы получаем движение свободной материальной точки на n -мерном торе. Таким образом, мы можем описать это движение в терминах линейного потока на n -мерном торе.

Механический аналог геометрической развертки состоит в том, что после столкновения любые две частицы обмениваются импульсами и поэтому можно рассматривать только перенос импульсов, что позволяет считать, что материальные точки проходят сквозь друг друга и только меняют направление на обратное на границе.

Упражнения

5.2.1. Рассмотрим движения бильярдного шара в единичном кубе. Сколько возможных направлений может иметь вектор скорости вдоль орбиты?

5.2.2. Сведите движение бильярдного шара в единичном кубе / к движению свободной частицы на торе и разложите его на сдвиги на торе.

Задачи для углубленного изучения

5.2.3. Опишите редукцию задачи о движении нескольких частиц с неравными массами на отрезке к задаче о бильярде.

5.2.4. Опишите распределение значений функции $\sin n + \cos \sqrt{2}n + \sin \sqrt{3}n$.

КОНСЕРВАТИВНЫЕ СИСТЕМЫ

§ 6.1. Сохранение фазового объема и рекуррентность

Мы увидим, что сохранение фазового объема — довольно естественное свойство, которое наблюдается, например, в динамических системах, возникающих из механики, и что это свойство является непосредственной причиной нетривиальной рекуррентности.

6.1.1. Критерии сохранения объема. До настоящего момента мы интересовались асимптотическим поведением индивидуальных орбит динамической системы. В основных примерах, которые приводились в гл. 4 и 5, наблюдается *рекуррентное* поведение орбит: все орбиты либо являются периодическими, т. е. возвращаются в начальное положение, либо подходят сколь угодно близко к начальному положению, как в случаях иррационального поворота окружности, непериодических параллельных переносов на торе или свободного движения материальной точки на торе. Этот тип поведения отличается от большинства явлений, которые мы наблюдали в гл. 2 и 3, поскольку для них типичным было то, что непериодическая орбита притягивалась к периодическим, а рекуррентность появлялась только для периодических орбит, которые для всех рассматривавшихся нелинейных систем и большинства линейных были немногочисленны.

Ключ к пониманию этого различия заключается в свойстве, которое нельзя заметить, рассматривая лишь индивидуальные орбиты, но которое можно наблюдать при одновременном рассмотрении больших множеств начальных условий. Это свойство *сохранения фазового объема*.

1. Сохранение фазового объема. Это свойство заключается попросту в том, что отображение, определяющее динамическую систему с дискретным временем (или, в случае потоков, отображение в момент времени t), сохраняет объем множеств в фазовом пространстве. Вполне очевидно, почему это свойство не совместимо с некоторыми из простых типов поведения, которые мы наблюдали в предшествующих главах. Например, если x — притягивающая неподвижная (или периодическая) точка отображения f , то достаточно малый шар с центром в точке x отображается под действием f (или итерации отображения f) внутрь еще меньшего шара. Следовательно, объем этого шара *уменьшается*. Это оправдывает другое название свойства сохранения объема — *несжимаемость*.

Пример 6.1.1. Любой параллельный перенос тора есть изометрия; он сохраняет размеры и форму множеств. В частности, образ любого прямоугольника Δ — это прямоугольник со сторонами той же длины, а следовательно, он имеет тот же самый объем, что и Δ . Поскольку каждое измеримое по Риману

множество может быть приближено конечным объединением прямоугольников, очевидно, что объем таких множеств также сохраняется.

Пример 6.1.2. *Линейное закручивание* $T: S^1 \times [0, 1] \rightarrow S^1 \times [0, 1]$ цилиндра (см. п. 2.6.3), $T(x, y) = (x + y, y)$, где значение $x + y$ определено по модулю 1, возникает естественным образом как отображение последования для свободного движения материальной точки на двумерном торе, которое рассматривалось в п. 5.2.2. Это не изометрия, поскольку прямоугольник Δ при этом отображении наклоняется и его образ является параллелограммом с углами $\pi/4$ и $3\pi/4$ (см. рис. 6.1.1). Тем не менее, мы можем понять динамику индивидуальных

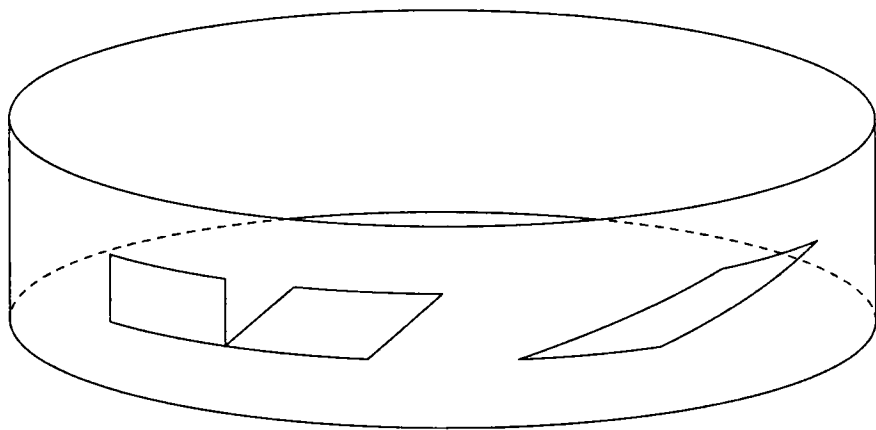


Рис. 6.1.1. Искаженный параллелограмм

орбит в терминах изометрий, если ограничимся рассмотрением произвольной горизонтальной окружности, на которой закручивание действует как поворот. При итерировании отображения параллелограмм $f^n(\Delta)$ становится более «длинным» и более «горизонтальным». Таким образом, форма прямоугольника Δ сильно изменяется. Однако объем (в данном случае площадь) параллелограмма $f(\Delta)$ (и параллелограмма $f^n(\Delta)$ для любого $n \in \mathbb{N}$) совпадает с объемом прямоугольника Δ , потому что основание и высота параллелограмма неизменны. Можно было рассуждать по-другому. Заметим, что площадь параллелограмма $f(\Delta)$ равна $\int_0^1 l(f(\Delta) \cap S^1 \times \{t\}) dt$ и что на любой окружности $C_t = S^1 \times \{t\}$ отображение T действует как поворот, а следовательно, $l(f(\Delta) \cap C_t) = l(\Delta \cap C_t)$, и интегрирование показывает, что площади параллелограммов Δ и $f(\Delta)$ равны.

Это рассуждение справедливо и для более сложных множеств.

2. Линейный случай. Предыдущие рассуждения, конечно, носили очень специальный характер, и нам необходим более систематический метод, позволяющий проверить, сохраняется фазовый объем или нет. Как это обычно делается в анализе, мы сформулируем желаемое условие в *инфинитезимальных* терминах. Поскольку подобные условия основаны на линейной аппроксимации нелинейных

объектов, довольно естественно и полезно посмотреть сначала на линейные отображения. Для линейных отображений мы можем получить ответ, используя средства элементарной линейной алгебры. Если линейное отображение в $\mathbb{R}^n$ представлено в евклидовых координатах матрицей L , то образ стандартного единичного куба $\Delta = \{(x_1, \dots, x_n): 0 \leq x_i \leq 1\}$ — это параллелепипед с объемом $|\det L|$. В более общей ситуации отношение объема образа параллелепипеда P к объему самого параллелепипеда P равно $|\det L|$. Как можно показать с помощью аппроксимации, то же самое справедливо и для множеств более общего вида. Таким образом, линейное отображение сохраняет объем тогда и только тогда, когда оно представляется матрицей с определителем ± 1 . Заметим, что это свойство не зависит от выбора базиса: относительно нового базиса отображение представляется матрицей $L' = C^{-1}LC$, где матрица C обратима и $\det L' = \det C^{-1} \det L \det C = \det L$. Рассуждая иначе, можно сказать, что детерминант равен произведению собственных значений, а последние не изменяются при сопряжениях.

3. Критерий. Теперь рассмотрим (нелинейное) дифференцируемое отображение f , определенное вблизи точки $x_0 \in \mathbb{R}^n$. Мы имеем $f(x) = f(x_0) + Df_{x_0}(x - x_0) + R_{x_0}(x)$, где Df_{x_0} — производная отображения f в точке x_0 , т. е. линейное отображение, которое представляется в стандартных координатах матрицей частных производных отображения f , и $R_{x_0}(x) = o(\|x - x_0\|)$. Таким образом, если мы зафиксируем $\epsilon > 0$ и возьмем достаточно малый параллелепипед Δ с центром в x_0 , то его образ под действием f будет лежать внутри параллелепипеда $f(x_0) + (1 + \epsilon)Df_{x_0}(\Delta - x_0)$ и содержать параллелепипед $f(x_0) + (1 - \epsilon)Df_{x_0}(\Delta - x_0)$. Следовательно, объем малого параллелепипеда приблизительно сохраняется тогда и только тогда, когда $|\det Df_{x_0}| = 1$. Детерминант отображения Df называется *якобианом* отображения f и обозначается Jf .

Теперь легко получить критерий сохранения фазового объема.

Предложение 6.1.3. Пусть $O \subset \mathbb{R}^n$ — открытое множество. Дифференцируемое отображение $f: O \rightarrow \mathbb{R}^n$ сохраняет объем тогда и только тогда, когда $|Jf| = \pm 1$.

Доказательство. Если $|Jf| \neq 1$ в некоторой точке x_0 , то согласно приведенным выше рассуждениям объем любого достаточно малого параллелепипеда изменяется. Если же $|Jf| = 1$, то, аппроксимируя множество A объединением параллелепипедов, мы можем гарантировать, что объем не изменяется больше чем на некоторую долю ϵ . Поскольку ϵ произвольно мало, объем должен сохраняться. $\square$

Конечно же, доказанное предложение является частным случаем известной формулы замены переменных из многомерного анализа. Если рассматривать отображение f как замену переменных и обозначить через χ_A характеристическую функцию множества A , то

$$\text{vol } f(A) = \int_{\mathbb{R}^n} \chi_{f(A)}(x) dx = \left| \int_{\mathbb{R}^n} \chi_A(y) \det \left(\frac{\partial f}{\partial x} \right)^{-1}(y) dy \right|.$$

Таким образом, если $\det \frac{\partial f}{\partial x} \equiv \pm 1$, то $\text{vol}(f(A)) = \text{vol}(A)$.

Определение 6.1.4. Пусть $O \subset \mathbb{R}^n$ — открытое множество. Дифференцируемое отображение $f: O \rightarrow \mathbb{R}^n$ сохраняет ориентацию, если $Jf > 0$.

4. *Дифференциальные уравнения.* Вернемся теперь к дифференциальным уравнениям. Чтобы найти условие несжимаемости решений φ^t системы $\dot{x} = f(x)$ в $\mathbb{R}^n$, посмотрим, как изменяется объем малых параллелепипедов при малых сдвигах $\varphi_{\Delta t}$. Как и прежде, мы имеем $\varphi_{\Delta t}(x) = x + \Delta t f(x) + R(x, \Delta t)$, где $R(x, \Delta t) = o(\Delta t)$. Таким образом, если нас интересуют только члены порядка не выше Δt , мы должны рассмотреть якобиан отображения $\bar{\varphi}\bar{\varphi}(x) = x + \Delta t f(x)$. Но тогда

$$\frac{\partial \bar{\varphi}\bar{\varphi}_i(x)}{\partial x_j} = \delta_i^j + \Delta t \frac{\partial f_i}{\partial x_j},$$

где $(\delta_i^j)_{ij}$ — единичная матрица.

Рис. 6.1.2. Несжимаемость

Для векторного поля $u(x) = (u_1(x_1, \dots, x_m), u_2(x_1, \dots, x_m), \dots, u_m(x_1, \dots, x_m))$ дивергенция определяется соотношением

$$\operatorname{div}(u) = \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \dots + \frac{\partial u_m}{\partial x_m}. \quad (6.1.1)$$

Эта величина появляется, когда мы находим члены порядка Δt :

$$J\bar{\varphi}\bar{\varphi} = 1 + \Delta t \sum_{i=1}^n \frac{\partial f_i}{\partial x_i} + o(\Delta t) = 1 + \Delta t \operatorname{div} f + o(\Delta t).$$

Дифференцируя это равенство по t и полагая $t = 0$, мы получаем

$$\left. \frac{d \operatorname{vol}(\varphi^t(A))}{dt} \right|_{t=0} = \int_A \operatorname{div} f dx,$$

что хорошо известно из векторного анализа. В частности, справедлив такой результат.

Предложение 6.1.5. Если $\dot{x} = f(x)$, где $\operatorname{div} f \equiv 0$, то поток, порожденный векторным полем f , сохраняет фазовый объем.

В следствии 6.2.3 с использованием этого предложения мы получим важный естественный класс систем, сохраняющих объем.

6.1.2. Теорема Пуанкаре о возвращении. Теперь мы покажем, что из сохранения объема динамической системой, фазовое пространство которой имеет конечный объем, следует рекуррентное поведение. Вначале мы докажем частный случай известного результата Пуанкаре¹. Мы сформулируем этот результат для отображений, заданных на *области*, т.е. открытом множестве или замыкании открытого множества.

Теорема 6.1.6. Пусть X является областью конечного объема в $\mathbb{R}^n$ или $\mathbb{T}^n$ и $f: X \rightarrow X$ — обратимое дифференцируемое отображение, сохраняющее объем. Тогда для любой точки $x \in X$ и любого числа $r > 0$ существует такое $n \in \mathbb{N}$, что

$$F^n(B_r(x)) \cap B_r(x) \neq \emptyset. \quad (6.1.2)$$

Доказательство. Предположим что найдутся такие $x \in X$, $r > 0$, что $f^n(B_r(x)) \cap B_r(x) = \emptyset$ для всех $n \in \mathbb{N}$. Тогда $f^{n+k}(B_r(x)) \cap f^k(B_r(x)) = \emptyset$ для всех $n, k \in \mathbb{N}$, так как отображение f^k обратимо. Следовательно, образы всех шаров $B_r(x)$ попарно не пересекаются и

$$\text{vol}(X) \geq \text{vol}\left(\bigcup_{k=0}^{n-1} f^k(B_r(x))\right) = \sum_{k=0}^{n-1} \text{vol}(f^k(B_r(x))) = N \text{vol}(B_r(x))$$

для всех $n \in \mathbb{N}$, в силу того что отображение f^k сохраняет объем. Тогда $\text{vol}(X) = \infty$, поскольку $\text{vol}(B_r(x)) > 0$. Полученное противоречие завершает доказательство. $\square$

Следствие 6.1.7. Пусть $f: X \rightarrow X$ — обратимое дифференцируемое отображение, сохраняющее объем, причем X — область конечного объема в $\mathbb{R}^n$ или $\mathbb{T}^n$. Тогда для любого $x \in X$ существуют такая последовательность точек $y_k \in X$ и такая последовательность $m_k \rightarrow \infty$, что $y_k \rightarrow x$ и $f^{m_k}(y_k) \rightarrow x$ при $k \rightarrow \infty$.

Доказательство. Положим $m_0 = 1$. Будем определять y_k и m_k индуктивно. Применяя теорему 6.1.6 к отображению $f^{-2m_{k-1}}$, мы получим такое число $m \in \mathbb{N}$, для которого существует $y_k \in f^{-2mm_{k-1}}(B_{1/k}(x)) \cap B_{1/k}(x)$. Положим $m_k := 2mm_{k-1}$. Тогда $m_k \rightarrow \infty$, $d(x, y_k) < 1/k$ и $d(x, f^{m_k}(y_k)) < 1/k$. $\square$

Изучим теперь рекуррентное поведение индивидуальных орбит.

Определение 6.1.8. Пусть X — метрическое пространство и $f: X \rightarrow X$ — непрерывное отображение. Точка $x \in X$ называется *положительно рекуррентной* относительно f , если существует такая последовательность $n_k \rightarrow \infty$, что $f^{n_k}(x) \rightarrow x$. Если отображение f обратимо, то точка x называется *отрицательно рекуррентной*, если она положительно рекуррентна для отображения f^{-1} , и соответственно *рекуррентной*, если она является положительно и отрицательно рекуррентной.

¹ Этот результат впервые появился в замечательной статье: *H. Poincaré. Sur le problème des trois corps et les équations de la Dynamique* // Acta Mathematica. 1890. V. 13. P. 1—270. Имеется перевод: *А. Пуанкаре. О проблеме трех тел и об уравнениях динамики. II* // Избранные труды. Т. II. М.: Наука, 1972. С. 357—444.

Рассуждая иначе, мы можем описать рекуррентность, используя понятие ω -предельного множества из определения 4.3.18: точка положительно рекуррентна тогда и только тогда, когда она лежит в собственном ω -предельном множестве. Вообще, не следует ожидать, что для сохраняющего объем отображения все точки рекуррентны, хотя это и было так во всех рассмотренных выше примерах (повороты окружности, сдвиги на торе или линейное закручивание). В п. 6.2.2 мы рассмотрим математический маятник, имеющий нерекуррентные орбиты, а именно те, которые располагаются на гомоклинической петле и сходятся к неустойчивому равновесию. В гл. 7 мы встретимся с сохраняющими объем системами с гораздо более сложной структурой орбит, в которых разные типы поведения, включая и нерекуррентные, встречаются одновременно. В общем случае можно только утверждать, что имеется много рекуррентных орбит в сохраняющих объем системах.

Теорема 6.1.9. Пусть X — замкнутая область конечного объема в $\mathbb{R}^n$ или T^n и $f: X \rightarrow X$ — обратимое сохраняющее объем отображение. Тогда множество рекуррентных точек для f плотно в X .

Доказательство. Для заданных $\epsilon > 0$ и $N \in \mathbb{N}$ точка $x \in X$ называется (ϵ, N) -рекуррентной, если существует такое $n > N$, что $d(f^n(x), x) < \epsilon$. Согласно следствию 6.1.7 множество (ϵ, N) -рекуррентных орбит плотно в X для любых $N \in \mathbb{N}$ и $\epsilon > 0$. Действительно, для данных $x \in X$, $\delta, \epsilon > 0$ и $N \in \mathbb{N}$ мы можем взять такое k , что $d(y_k, x) < \delta$, $n_k > N$ и $d(y_k, x) + d(f^{n_k}(y_k), x) < \epsilon$. Это множество также открыто в силу непрерывности отображения f и его итераций. С другой стороны, точка x положительно рекуррентна тогда и только тогда, когда она $(1/n, k)$ -рекуррентна для всех $n, k \in \mathbb{N}$ (выбирая такие значения $n_k > k$, что $d(f^{n_k}(x), x) < 1/k$, мы получим, что $n_k \rightarrow \infty$ и $f^{n_k}(x) \rightarrow x$). Таким образом, множество всех положительно рекуррентных точек есть пересечение открытых плотных множеств $(1/n, k)$ -рекуррентных точек для $n, k \in \mathbb{N}$. Как следует из теоремы Бэра (лемма А.1.15), это пересечение является плотным. $\square$

Мы доказали плотность множества положительно рекуррентных точек. Точно так же доказывается плотность множества отрицательно рекуррентных точек, если рассмотреть $(2^{-k}, N)$ -рекуррентные точки для f^{-1} . Наконец, мы можем убедиться в плотности множества рекуррентных точек, если рассмотрим точки, которые являются одновременно $(2^{-k}, N)$ -рекуррентными для f и для f^{-1} . Эти множества открыты и плотны, и поэтому, как следует из леммы А.1.15, (счетное) пересечение таких множеств по всем $k, N \in \mathbb{N}$ также плотно.

6.1.3. Равномерная рекуррентность. Рекуррентность, установленная в теореме 6.1.9, нерегулярна сразу в нескольких отношениях. Во-первых, как мы уже заметили, она может не быть однородной в пространстве: некоторые точки могут быть рекуррентны, в то время как другие — нет. Во-вторых, мы не получаем никакой информации относительно множества моментов времени, в которые происходят возвращения. Усилить понятие рекуррентности можно, например, указав, насколько регулярна (или однородна) она относительно *времени*. Пока что мы можем это сделать только для случая периодических точек. Если $x \in X$ — периодическая точка с минимальным периодом n , то орбита $\mathcal{O}(x) = \{x, f(x), \dots, f^{n-1}(x)\}$ точки x — конечное (а следовательно, дискретное) множество, состоящее из n точек, и если r настолько мало, что $d(f^k(x), f^l(x)) > r$ для $k \neq l \in \{0, \dots, n-1\}$, то из неравенства $d(f^n(y), y) < r$ следует, что $f^n(y) = y$ для любой точки $y \in \mathcal{O}(x)$, и, таким образом, множество $\{n \in \mathbb{Z} : d(f^n(y), y) < r\}$ представляет собой арифметическую прогрессию $n\mathbb{Z}$.

Одно из свойств арифметической прогрессии заключается в том, что она имеет непустое пересечение с любым множеством последовательных целых чисел, которое имеет больше элементов, чем разность прогрессии n . Иначе говоря, все пустые отрезки имеют одну и ту же длину. Мы можем ослабить это условие, наложив ограничение на длину таких отрезков.

Определение 6.1.10. Подмножество S в $\mathbb{N}$ или $\mathbb{Z}$ называется *синдетическим*, если имеется такое $N \in \mathbb{N}$, что $\{n+k: 1 \leq k \leq N\} \cap S \neq \emptyset$ для всех $n \in \mathbb{N}$.

Это свойство представляет собой ключ к изучению равномерной рекуррентности для непериодических точек.

Определение 6.1.11. Пусть $f: X \rightarrow X$ — непрерывное отображение метрического пространства X . Точка $x \in X$ называется *равномерно рекуррентной*, если для любого $r > 0$ множество $\{n: d(x, f^n(x)) < r\}$ синдетическое, т. е. существует такое $N = N(r)$, что среди любых N последовательных итераций $f^{n+k}(x)$, $k = 0, \dots, N-1$, имеется по крайней мере одна, для которой $d(x, f^{n+k}(x)) < r$.

Очевидно, что любая периодическая точка равномерно рекуррентна. Доказательство предложения 4.1.1 показывает, что это справедливо для иррационального поворота при $N = [1/r] + 1$. Так как периодические точки равномерно рекуррентны, мы получаем следующий результат.

Предложение 6.1.12. Для поворота окружности все точки равномерно рекуррентны.

Хотя периодический случай показывает, что из равномерной рекуррентности всех точек не следует минимальность, между этими двумя понятиями имеется тесная связь.

Теорема 6.1.13. Предположим, что множество X компактно и $f: X \rightarrow X$ — гомеоморфизм. Точка $x \in X$ равномерно рекуррентна тогда и только тогда, когда замыкание ее орбиты есть компактное минимальное множество (см. определение 4.1.4).

Доказательство. Предположим, что точка x равномерно рекуррентна, и допустим, что U — окрестность точки x с компактным замыканием. Тогда множество $R := \{n \in \mathbb{Z}: f^n(x) \in U\}$ синдетическое, так что мы можем найти такое $N \in \mathbb{N}$, что среди любых N итераций $f^{n+k}(x)$, $k = 0, \dots, N-1$, имеется по крайней мере одна, для которой $f^{n+k}(x) \in U$. Тогда $\mathcal{O}(x) = \{f^n(x): n \in \mathbb{Z}\} = \{f^{n+k}(x): n \in R, 0 \leq k < N\} \subset \bigcup_{k=0}^{N-1} f^k(U) =: U_N$. Так как замыкание множества U_N компактно, компактно и замыкание

орбиты $\mathcal{O}(x)$. Кроме того, из этого следует, что если $y \in \overline{\mathcal{O}(x)}$, то $y \in U_N$, а значит, $\overline{\mathcal{O}(y)} \cap U \neq \emptyset$. В силу произвольности множества U мы заключаем, что $x \in \overline{\mathcal{O}(y)}$, и, таким образом, $\mathcal{O}(y) = \overline{\mathcal{O}(x)}$.

Обратно, предположим, что $\overline{\mathcal{O}(x)}$ — компактное минимальное множество, и рассмотрим окрестность U точки x . Поскольку $\overline{\mathcal{O}(x)} \setminus \{f^n(U): n \in \mathbb{Z}\}$ — замкнутое собственное инвариантное подмножество множества $\overline{\mathcal{O}(x)}$, оно должно быть пустым в силу минимальности. Это показывает, что $\overline{\mathcal{O}(x)} \subset \{f^n(U): n \in \mathbb{Z}\}$, а из определения компактности следует, что имеется конечное подпокрытие, так что $\mathcal{O}(x) \subset \overline{\mathcal{O}(x)} \subset f^m(U_N)$ для некоторых $N \in \mathbb{N}$ и $m \in \mathbb{Z}$.

Как и прежде, положим $R := \{n \in \mathbb{Z}: f^n(x) \in U\}$. Теперь $\mathcal{O}(x) \subset f^m(U_N)$, так что для любого $i \in \mathbb{Z}$ имеются такая точка $y \in U$ и такое неотрицательное число $k < N$, что $f^i(x) = f^m(f^k(y))$. Таким образом, $f^{i-m-k}(x) \in U$ и $i-m-k \in R$. Так как m фиксировано, а $0 \leq k < N$, мы доказали, что множество R синдетическое, как и требуется. $\square$

Из равномерной рекуррентности следует, что замыкания орбит разбивают фазовое пространство.

Предложение 6.1.14. Предположим, что пространство X компактно и $f: X \rightarrow X$ — гомеоморфизм. Замыкания орбит определяют разбиение пространства X на компактные множества тогда и только тогда, когда каждая точка равномерно рекуррентна.

Доказательство. Согласно теореме 6.1.13 и упражнению 4.1.8 из равномерной рекуррентности всех точек следует, что замыкания орбит не пересекаются или совпадают (а также что они компактны). Обратно, из того факта, что замыкания орбит либо не пересекаются, либо совпадают, следует минимальность замыканий орбит, так что с учетом теоремы 6.1.13 и условия компактности мы можем сделать вывод о равномерной рекуррентности всех орбит. $\square$

Линейное закручивание (пример 6.1.2) наглядно поясняет эти результаты.

Упражнения

6.1.1. Докажите, что гомеоморфизм окружности, сохраняющий ориентацию и объем (т. е. сохраняющий длину), есть поворот.

6.1.2. Покажите, что сохраняющее объем отображение не может иметь притягивающей неподвижной точки.

6.1.3. Выясните, сохраняет ли площадь закручивание $T: S^1 \times [0, 1] \rightarrow S^1 \times [0, 1]$, $T(x, y) = (x + f(y), y)$, где сумма определена по модулю 1 и отображение f дифференцируемо.

6.1.4. Выясните, сохраняет ли площадь поток, порожденный дифференциальным уравнением

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} y \\ -x \end{pmatrix}.$$

6.1.5. Выясните, сохраняет ли площадь поток, порожденный дифференциальным уравнением

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} y \\ -\sin x \end{pmatrix}.$$

6.1.6. Выясните, сохраняет ли площадь поток, порожденный дифференциальным уравнением

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} y \\ -y - \sin x \end{pmatrix}.$$

6.1.7. Пусть X — замкнутая область конечного объема в $\mathbb{R}^n$ или $\mathbb{T}^n$ и $f: X \rightarrow X$ — обратимое отображение с притягивающей неподвижной точкой. Покажите, что множество рекуррентных точек отображения f не плотно в X .

6.1.8. Приведите пример метрического пространства и счетного набора открытых плотных множеств в нем, пересечение которых пусто.

6.1.9. Докажите утверждение теоремы 6.1.9 для открытой области с компактным замыканием.

Задачи для углубленного изучения

6.1.10. Пусть X — метрическое пространство и отображение $f: X \rightarrow X$ топологически транзитивно. Покажите, что множество точек, орбиты которых не плотны, есть объединение счетного числа нигде не плотных множеств.

6.1.11. Докажите, что если отрезок представлен в виде счетного объединения замкнутых множеств, то одно из этих множеств содержит отрезок.

§ 6.2. Ньютоновы системы классической механики

Осознание того факта, что механические системы описываются дифференциальными уравнениями (сила определяет вторую производную координаты), и развитие математического анализа произвели глубокую революцию научной мысли и, в частности, создали средства для описания, предсказания и проектирования физических систем, которые исключительно успешно применялись

в последние три столетия. Мы рассмотрим некоторые методы описания и решения подобных систем.

В механической системе набор данных типа координат (или конфигураций; сюда могут включаться, например, углы) и скоростей отдельных частей системы полностью описывает состояние системы в смысле следующего принципа детерминированности: настоящее состояние механической системы однозначно определяет ее будущие состояния. (Или в нашей терминологии: механическая система определяет на пространстве состояний динамическую систему.) Если, например, наша механическая система состоит из единственной материальной точки, тогда состояние задается положением x в евклидовом пространстве и скоростью $v = \dot{x}$, производной функции x по времени t . В частности, если полная эволюция определяется этими данными, т. е. x задается как функция от t , то и вторая производная $\ddot{x}$ функции x также определяется этими данными. Поэтому $\ddot{x}$ выражается через t , x и $\dot{x}$: $\ddot{x} = f(t, x, \dot{x})$. Таким образом, подобные механические системы описываются дифференциальными уравнениями. (Тогда из существования и единственности решений дифференциальных уравнений (теорема 9.4.1) в свою очередь следует принцип детерминированности.) Так как «состояние» механической системы всегда описывается в терминах координат и скоростей, дифференциальные уравнения, которые возникают в механике, всегда имеют второй порядок. Множество положений системы, или ее конфигураций, называется *конфигурационным пространством*, а множество состояний системы называется *пространством состояний* или *фазовым пространством*².

В этом параграфе мы обратимся к механике. Механика является важным предметом, и мы посвятим ей несколько страниц, потому что имеются два аспекта механических систем, которые могут упростить их динамику по сравнению с другими системами дифференциальных уравнений с таким же числом переменных. Так как орбиты консервативной системы остаются на уровнях постоянной энергии, размерность фазового пространства существенно уменьшается, а с другой стороны, рассеяние энергии из-за трения может делать динамику в диссипативных системах асимптотически простой.

Мы начнем с изучения уравнения Ньютона и его основных свойств и введем некоторые механические идеи, используя математический маятник. Мы также обратимся к задаче центральных сил (и второму закону Кеплера), лежащей в основе всей небесной механики, из которой, в свою очередь, динамика получила некоторые из наиболее важных задач. Наконец, мы введем лагранжев формализм в механике, основанный на вариационном принципе: путь механической системы в фазовом пространстве решает некоторую задачу оптимизации.

6.2.1. Уравнение Ньютона. Центральный закон классической механики — закон Ньютона, гласящий, что внешняя сила, действующая на механическую систему (скажем, на материальную точку, твердое тело, планету и т. п.), пропорциональна вызываемому ей изменению скорости:

$$f = ma.$$

²См. также гл. 6 работы: А. Пуанкаре. Наука и гипотеза // О науке. М.: Наука, 1989.

Это уравнение описывает, например, движение материальной точки массы m в $\mathbb{R}^n$ под действием силы f , так как из него можно найти ускорение a . Оно также описывает маятник, в котором присутствуют связи.

1. *Дифференциальные уравнения второго порядка.* Уравнение Ньютона сводится к дифференциальному уравнению второго порядка: если $x \in \mathbb{R}^n$ — положение материальной точки, тогда ускорение определяется следующим образом: $a := \ddot{x} = \frac{d^2x}{dt^2}$. Если сила f является функцией только переменной x (это исключает трение), то получается уравнение

$$m \frac{d^2x}{dt^2} = f(x).$$

Пример 6.2.1. На яблоко, падающее с дерева, действует постоянная сила притяжения mg (где g — константа, приблизительно равная 10 м/с^2), и, следовательно, $\ddot{x} = -g$, где x — высота. (Яблони не настолько высоки, чтобы учитывать сопротивление воздуха в уравнениях.)

Дважды интегрируя, мы получаем $x(t) = -gt^2/2 + v(0)t + x(0)$.

Пример 6.2.2 (гармонический осциллятор). Для точечной массы, прикрепленной к пружине и отклоненной от положения равновесия (покоя) на расстояние x , закон Гука показывает, что сила растяжения пружины равна $-kx$, где k — постоянная характеристика пружины (которая измеряет ее жесткость). Таким образом, мы получаем $\ddot{x} = -kx$.

Решения имеют вид $x(t) = a \sin(\sqrt{k}t) + b \cos(\sqrt{k}t)$, в чем можно убедиться, подобрав два независимых решения и взяв их линейную комбинацию.

2. *Сведение к системе первого порядка.* Чтобы изучать такие системы дифференциальных уравнений второго порядка в достаточно общей ситуации, удобно преобразовать их к системам уравнений первого порядка, определяя скорость как дополнительную независимую переменную, т. е. полагая $v := \dot{x} = \frac{dx}{dt} \in \mathbb{R}^n$. Тогда уравнение $m \frac{d^2x}{dt^2} = f(x)$ преобразуется в систему

$$\begin{aligned} \frac{d}{dt}x &= v, \\ \frac{d}{dt}mv &= f(x). \end{aligned}$$

Это автономная система дифференциальных уравнений первого порядка относительно новых переменных $\begin{pmatrix} x \\ v \end{pmatrix}$. Ее общее решение определяет динамическую систему в пространстве $\mathbb{R}^n \times \mathbb{R}^n$ (или в его подмножестве) в координатах (x, v) (см. п. 9.4.7). Эти уравнения имеют несколько специальных свойств, которые отличают их от автономных систем дифференциальных уравнений общего вида в $\mathbb{R}^{2n}$.

3. *Сохранение объема.* Поскольку все необходимые производные равны нулю, векторное поле в $\mathbb{R}^{2n}$, определенное уравнением Ньютона, является бездивергентным, т. е. имеет нулевую дивергенцию (см. (6.1.1)).

Согласно предложению 6.1.5 из этого следует, что ньютоновы системы сохраняют фазовый объем.

Следствие 6.2.3. *Ньютоновы системы сохраняют объем $dx dv$ в фазовом пространстве.*

4. *Энергия и импульс.* Величина $p := mv$ называется *импульсом*. *Кинетическая энергия* определяется как $\frac{1}{2}m\langle v, v \rangle$.

Если сила f является градиентным векторным полем, т.е. $f = -\nabla V := -\left(\frac{\partial V}{\partial x_1}, \dots, \frac{\partial V}{\partial x_m}\right)$, тогда

$$\frac{d}{dt}mv = -\nabla V. \quad (6.2.1)$$

Функция $V: \mathbb{R}^n \rightarrow \mathbb{R}$ называется *потенциальной энергией*; при этом *полная энергия* $H = \frac{1}{2}m\langle v, v \rangle + V$ сохраняется в системе, потому что

$$\frac{dH}{dt} = \langle v, m\dot{v} \rangle + \frac{dV}{dt} = \langle v, m\ddot{v} \rangle + \langle \dot{x}, \nabla V \rangle = \langle v, m\ddot{v} + \nabla V \rangle = 0$$

и функция с нулевой производной вдоль кривой постоянна на этой кривой. Это полезный простой принцип анализа, который может использоваться при рассмотрении систем с непрерывным временем. Энергия также сохраняется в системах со связями (хотя импульс может и не сохраняться). По причине сохранения энергии эти системы называются *консервативными*.

5. *Геодезический поток.* Предшествующие рассуждения можно провести и для *свободного движения материальной точки*, т.е. движения, при котором сила равна нулю. В евклидовом пространстве мы получим движение с постоянной скоростью вдоль прямой. В пространствах, отличных от евклидовых, получающееся движение имеет постоянную скорость, но понятие прямой должно быть заменено на более общее понятие *геодезической*. Это движение также известно как *геодезический поток*. Примеры — это сфера, геодезический поток на которой представляет собой движение с постоянной скоростью по большим кругам, являющимися геодезическими на сфере, и тор, в случае которого движение происходит по проекциям прямых. Геометрически и физически геодезические можно определять на поверхности в $\mathbb{R}^3$ по определенному направлению, как пересечение поверхности с плоскостью, порожденной нормалью и самим направляющим вектором. Эта геометрическая картина соответствует тому, что при движении по геодезическим с постоянной скоростью ускорение перпендикулярно поверхности, потому что единственная возможная сила — связь, которая удерживает материальную точку на поверхности. Эта связь ортогональна поверхности, поскольку любая касательная составляющая привела бы к силе, действующей непосредственно на материальную точку на поверхности.

Геодезические потоки появляются на страницах этой книги несколько раз — скажем, в п. 5.2.2 и 6.2.8.

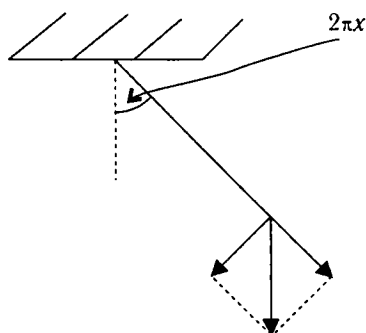


Рис. 6.2.1. Математический маятник

6.2.2. Математический маятник. В качестве примера ньютоновской системы рассмотрим маятник, состоящий из точечной массы на плоскости, прикрепленной стержнем к зафиксированному шарниру, подобно маятнику в старинных часах.

1. *Модель.* Пусть угол отклонения от вертикальной линии составляет $2\pi x$. Маятник находится под действием направленной вниз гравитационной силы mg (где m — масса маятника и g — ускорение свободного падения, приблизительно равное $9,81 \text{ м/с}^2$), угловая компонента которой равна $-mg \sin 2\pi x$. Учитывая, что это

выражение должно равняться произведению массы на ускорение, т. е. $m \cdot 2\pi L \ddot{x}$, мы получаем следующее дифференциальное уравнение, описывающее маятник:

$$2\pi mL \ddot{x} + mg \sin 2\pi x = 0.$$

2. *Переход к безразмерным координатам.* Часто бывает полезно упростить такие дифференциальные уравнения, перейдя к безразмерным координатам, т. е. выбрав масштаб времени таким образом, чтобы коэффициенты уравнения оказались безразмерными и чтобы по возможности их стало меньше. Прежде всего для этого нужно подобрать такое значение времени T (мы определим его ниже), чтобы при замене времени t на безразмерное время $\tau := t/T$ производные функции x по переменной τ были приближенно равны 1. Заметим, что по правилу дифференцирования сложной функции мы имеем $dx/dt = dx/d\tau \cdot d\tau/dt = (1/T)dx/d\tau$ и $d^2x/dt^2 = (1/T^2)d^2x/d\tau^2$. Таким образом, дифференциальное уравнение принимает вид

$$2\pi mL/T^2 \frac{d^2x}{d\tau^2} + mg \sin 2\pi x = 0.$$

Оба слагаемых являются силами и потому станут безразмерными, если мы разделим уравнение на mg :

$$\frac{2\pi L}{gT^2} \frac{d^2x}{d\tau^2} + \sin 2\pi x = 0. \quad (6.2.2)$$

Безразмерный коэффициент $2\pi L/(gT^2)$ должен быть порядка единицы, чтобы вторая производная $d^2x/d\tau^2$ также была порядка единицы, так что мы полагаем $T = \sqrt{2\pi L/g}$ и в итоге приходим к дифференциальному уравнению

$$\ddot{x} + \sin 2\pi x = 0,$$

где точка теперь обозначает дифференцирование по τ . С физической точки зрения выбор величины T естествен, поскольку она непосредственно связана с периодом колебаний гармонического осциллятора, который является результатом линеаризации математического маятника (см. п. 6.2.2.7).

3. *Сведение к системе первого порядка.* Полученное выше дифференциальное уравнение эквивалентно системе дифференциальных уравнений первого порядка

$$\begin{aligned}\dot{x} &= v, \\ \dot{v} &= -\sin 2\pi x,\end{aligned}\tag{6.2.3}$$

$x \in S^1$, $v \in \mathbb{R}$. Этот простой пример особенно интересен тем, что в силу периодичности угловой координаты (но не скорости!) фазовое пространство представляет собой цилиндр $S^1 \times \mathbb{R}$.

Полная энергия определяется выражением $H(x, v) = \frac{1}{2}v^2 - \frac{1}{2\pi} \cos 2\pi x$. Так как это система со связью, полезно в явном виде проверить закон сохранения энергии: $dH/dt = v\dot{v} + (1/2\pi) \sin 2\pi x \dot{x} = 0$. Таким образом, орбиты находятся на кривых уровня $H = \text{const}$.

4. *Орбиты.* Если $-1/(2\pi) < H < 1/(2\pi)$, то каждый энергетический уровень состоит из единственной замкнутой кривой, соответствующей колебаниям вокруг устойчивого положения равновесия $(x, v) = (0, 0)$. Такие орбиты отделены от орбит с более высокой энергией, соответствующих полным и повторяющимся вращениям вокруг шарнира, *гомоклинической петлей* (см. определение 2.3.4), для которой $H = 1/(2\pi)$, содержащей неустойчивое положение равновесия $(x, v) = (1/2, 0)$ (см. рис. 6.2.2). Мы не рекомендуем проверять характер этих орбит на маятнике в старинных часах. Прежде всего, орбиты настолько

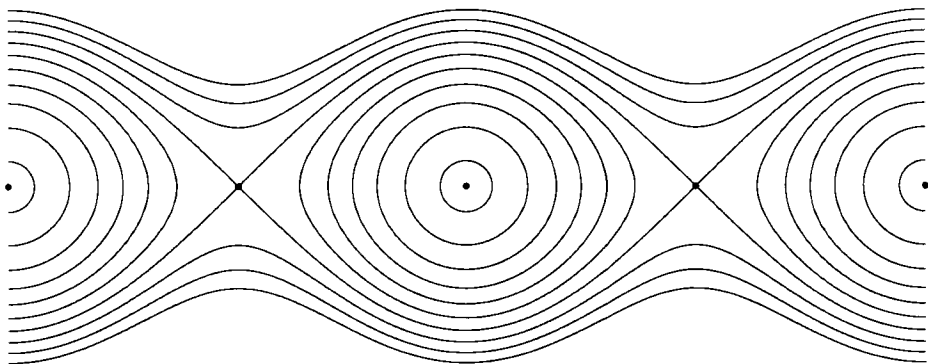


Рис. 6.2.2. Фазовый портрет математического маятника

неустойчивы, что (из-за трения, хотя и малого) мы фактически не наблюдаем их, а во-вторых, вы попросту сломаете часы, если все же попытаете. Для $H > 1/(2\pi)$ каждый энергетический уровень состоит из двух орбит, соответствующих вращениям в противоположных направлениях. Таким образом, почти все орбиты этой системы периодические, но тем не менее имеется два различных семейства орбит. Орбиты с более низкой энергией располагаются вокруг положения равновесия $(0, 0)$. Любые две из них могут быть непрерывно деформированы друг в друга. Орбиты с более высокими энергиями располагаются вокруг цилиндра (подобно

резинке на свернутом в рулон листе бумаги), и они также могут быть продеформированы друг в друга. Но ни одна из орбит с низкой энергией не может быть непрерывно деформирована в какую-либо орбиту с высокой энергией. В соответствии с этим два семейства орбит отделены друг от друга *сингулярными* орбитами, гомоклиническими к неустойчивому положению равновесия $(1/2, 0)$.

5. *Глобальная картина.* Качественно эта ситуация соответствует топологическому цилиндру, имеющему форму изогнутой трубы (подобно сливной трубе), для которого H является функцией высоты (см. рис. 6.2.3). Тогда кривые уровня — это горизонтальные сечения. Устойчивое положение равновесия $(0, 0)$

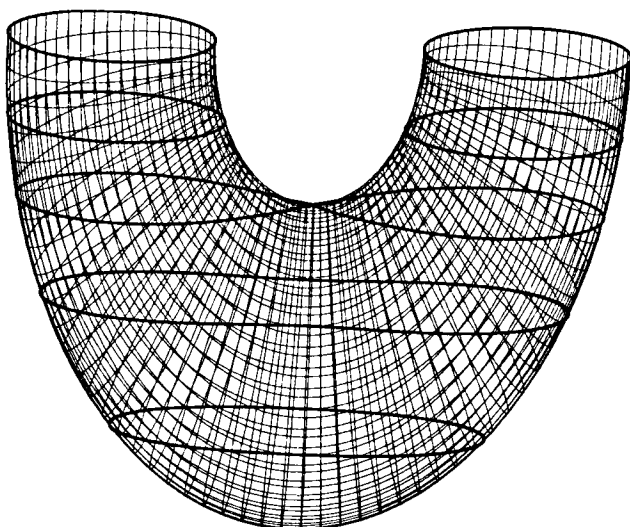


Рис. 6.2.3. Энергия как функция высоты

соответствует самой нижней точке на этой трубе, неустойчивое положение равновесия — седловой точке трубы. Уровни энергии выше $1/(2\pi)$ состоят из пар замкнутых кривых выше седловой точки, которые можно свободно передвигать вверх и вниз, не опуская при этом ниже седла. Более низкие уровни энергии располагаются ниже седловой точки, и они могут быть полностью сжаты к самой нижней точке цилиндра. Кривая в виде восьмерки на уровне седловой точки вообще не может никуда перемещаться. Итак, фазовое пространство в целом распадается в объединение регулярных кривых и одной сингулярной — восьмерки. Преобразование этой системы за время t действует на естественный параметр в каждой из этих инвариантных (деформированных) окружностей как перенос. Таким образом, мы выяснили, что динамика распадается на повороты (деформированных) окружностей, в частности, повороты окружностей встречаются естественным образом в простых динамических системах.

6. *Интегрируемость, инвариантный элемент длины.* Разложение фазового пространства на кривые уровня не только полезно для хорошего каче-

ственного и интуитивного понимания, но также и представляет собой хороший способ записи точных решений с использованием функций, которые получаются из элементарных при помощи интегрирования, взятие обратной функции и алгебраических операций. Например, в нашем случае явные решения выражаются через эллиптические интегралы, которые нельзя непосредственно выразить через элементарные функции.

Предложение 6.2.4 из следующего параграфа предоставляет механизм для получения решений. Чтобы кратко описать здесь этот механизм, мы запишем стандартный элемент площади в фазовом пространстве как $dH dl$, где дифференциал dH инвариантен относительно потока, поскольку инвариантна функция H , а dl — элемент длины на кривых $H = \text{const}$, разделенный на $\|\nabla H\|$. Как следует из системы (6.2.3), $\|\nabla H\|$ есть скорость движения вдоль кривых $H = \text{const}$, так что дифференциал dl также инвариантен относительно потока и, следовательно, поток сохраняет площадь. Как мы уже отметили в § 6.1, последнее утверждение эквивалентно тому, что дивергенция равна нулю, и, следовательно, применимо ко всем ньютоновым системам.

7. Линеаризация. Сейчас мы имеем возможность увидеть на примере, как линеаризация действует на фазовый портрет и насколько хорошо линеаризованная система отражает настоящую динамику системы. Поскольку линеаризация должна представлять хорошее локальное приближение нелинейных систем, наиболее интересно рассматривать ее вблизи положения равновесия. Мы рассмотрим движение вблизи устойчивого положения равновесия, т. е. *малые колебания*.

Мы хотим сравнить нашу картину с явными решениями системы дифференциальных уравнений

$$\begin{aligned}\dot{x} &= v, \\ \dot{v} &= -2\pi x,\end{aligned}$$

которая получается из (6.2.3), если заменить правую часть ее линейным приближением в точке $(0, 0)$. Это не что иное, как *гармонический осциллятор* (см. пример 6.2.2).

Полная энергия гармонического осциллятора равна $H(x, v) = v^2/2 + \pi x^2$. Заметим, что разложение гамильтониана

$$H(x, v) = \frac{1}{2}v^2 - \frac{1}{2\pi} \cos 2\pi x$$

в ряд Тейлора до членов второго порядка есть

$$\frac{1}{2}v^2 - \frac{1}{2\pi} \left(1 - \frac{1}{2}(2\pi x)^2\right) = \frac{1}{2}v^2 - \frac{1}{2\pi} + \pi x^2,$$

т. е. оно совпадает с соответствующим разложением для линеаризованной системы (с точностью до аддитивной постоянной $-1/(2\pi)$, которая не меняет ни множества уровней, ни производные). Кривые уровня для гармонического осциллятора — это линии $v^2/2 + \pi x^2 = \text{const}$, т. е. эллипсы с центрами в начале

координат. Из явного вида решений

$$A \begin{pmatrix} \sin(\sqrt{2\pi}(t+c))/\sqrt{\pi} \\ \cos(\sqrt{2\pi}(t+c))/\sqrt{2} \end{pmatrix}$$

мы видим, что все они имеют один и тот же период.

Качественно этот простой фазовый портрет очень напоминает портрет для случая математического маятника вблизи начала координат и соответственно линеаризованная система дает в данном случае разумную аппроксимацию. Заметим, однако, что имеются два различия. Вдали от начала координат фазовые портреты становятся существенно различными из-за присутствия второго положения равновесия и гомоклинической петли в случае математического маятника. Кроме того, периодические решения вокруг точки $(0, 0)$ имеют более длинные периоды для большей энергии, потому что скорость около неустойчивого равновесия мала. В частности, решения нелинейной системы не имеют постоянных периодов вблизи точки $(0, 0)$, т. е. период зависит от амплитуды колебаний. Это один из факторов, которые нужно учитывать при проектировании часов с маятниками. Амплитуда должна оставаться постоянной (или маятник следует снабдить циклоидальными ограничителями, чтобы исключить зависимость от амплитуды).

Заметим, что фазовое пространство математического маятника и гармонического осциллятора разлагается в инвариантные (деформированные) окружности и отображение в момент времени $t = 1$ просто сдвигает параметр на каждой окружности на постоянную величину, зависящую от окружности. Таким образом, мы можем абсолютно естественно получить детальное качественное описание этого отображения, рассматривая каждую из этих окружностей по отдельности, а следовательно, просто изучая повороты окружности (см. п. 4.1.1).

6.2.3. Инвариантный объем на энергетических уровнях. Как мы только что отметили, ньютоновы системы сохраняют не только фазовый объем, но и множества уровня функции Гамильтона H . Из п. 6.2.2.6 мы знаем, что в случае математического маятника поток также сохраняет параметр длины на кривых уровня энергии. Мы пришли к этому выводу непосредственно из рассмотрения уравнений движения (6.2.3) и использовали его, чтобы доказать сохранение площади. Теперь мы покажем, что любое отображение, которое сохраняет объем и множества уровня функции, также сохраняет и объем на этих множествах уровня. Чтобы лучше понять идею, изучим сначала внимательно двумерный случай, который, в частности, пояснит наши результаты относительно маятника.

Предложение 6.2.4. *Предположим, что отображение $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ сохраняет площадь и H — инвариантная функция, т. е. $H(f(p)) = H(p)$ для всех $p \in \mathbb{R}^2$. Тогда каждое множество уровня, которое не содержит никаких критических точек, может быть разложено на кривые $c_z := \{p: H(p) = z\}$, каждая из которых параметризуется как $c_z(t)$, так что $\dot{f}(c_z(t)) = c_z(t + s(z))$ для некоторой функции s , зависящей только от множества уровня. Другими словами, относительно этого параметра отображение f действует на каждой кривой как параллельный перенос.*

Доказательство. Параметризуем кривую c_z так, чтобы для всех t выполнялось равенство $\|\dot{c}_z(t)\| \|DH_{c_z(t)}\| = 1$. Рассмотрим единичный касательный вектор v и единичный нормальный вектор $w = DH/\|DH\|$ к c_z в точке $p = c_z(t)$. Тогда площадь прямоугольника P , порожденного векторами ev и ew , равна e^2 , и, как показано на рис. 6.2.4, площадь его образа $f(P)$ под действием f равна (с точностью до малой относительной ошибки, которая стремится к нулю при $\epsilon \rightarrow 0$) площади параллелограмма в точке $f(p) = c_z(\hat{t})$, порожденного образами ev' и ew' векторов ev и ew под действием Df . Если считать вектор ev' основанием параллелограмма, то его высота равна δ , где δ — длина проекции вектора ew' на нормаль к c_z в точке $f(p)$. Линейной аппроксимацией мы добиваемся выполнения



Рис. 6.2.4. Движение на множествах уровня

приближенных равенств

$$\begin{aligned} H(p + \epsilon w) &\approx H(p) + \epsilon \|DH_p\|, \\ H(f(p) + \epsilon w') &\approx H(f(p)) + \delta \|DH_{f(p)}\|. \end{aligned}$$

Поскольку отображение f сохраняет функцию H , левые части равны, так же как и первые члены в правых частях. Поэтому $\delta \approx \epsilon \frac{\|DH_p\|}{\|DH_{f(p)}\|}$ и $f(P)$ имеет площадь $\delta \epsilon \|v'\| \approx \frac{\|v'\| \epsilon^2 \|DH_p\|}{\|DH_{f(p)}\|}$, которая в силу сохранения площади должна равняться ϵ^2 . Из этого следует, что $\|v'\| \|DH_{f(p)}\| = \|DH_p\|$ и, таким образом, $v' \|DH_{f(p)}\| = v \|DH_p\|$. (Заметим, что в эти последние уравнения не входит ϵ , а поэтому они являются точными.) Если мы возьмем единичный вектор $v = c'_z(t) \|DH_p\|$, тогда $v' = Df(c'_z(t)) \|DH_p\|$ и, таким образом,

$$c'_z(\bar{t}) \frac{d\bar{t}}{dt} = \frac{d}{dt} c_z(\bar{t}) = Df(c'_z(t)) = \frac{v'}{\|DH_{f(p)}\|} = \frac{v}{\|DH_{f(p)}\|} = \frac{c'_z(t) \|DH_p\|}{\|DH_{f(p)}\|} = c'_z(\bar{t}),$$

поскольку всюду $c'_z \|DH\| = 1$. Из этого очевидным образом следует, что $\frac{d\bar{t}}{dt} = 1$, а значит, $\bar{t} = t + s(z)$ для некоторой функции s , зависящей только от z , т.е. только от инвариантной кривой. $\square$

Заметим снова, что это описание в точности соответствует ситуации, с которой мы столкнулись при рассмотрении математического маятника. Еще одна общая черта состоит в том, что и здесь мы проводим лишь локальные вычисления. Это показывает, что совсем не обязательно рассматривать отображение, определенное на всем пространстве $\mathbb{R}^2$, или же функцию H , определенную на всем фазовом пространстве. С некоторыми уточнениями наш результат справедлив для отображения открытого подмножества в $\mathbb{R}^2$, которое сохраняет площадь, и для функции H на этом открытом множестве. Аналогично мы можем применить этот результат и к отображениям цилиндра $S^1 \times \mathbb{R}$, поскольку все наши локальные вычисления точно так же применимы и в этом случае. Основную причину этого можно объяснить геометрически: рассмотрим *локальную трубку тока*, т.е. малую площадь, которая загибается локальной трансверсалью за малый промежуток времени. Если следовать по этой трубке тока далее к области, где множества уровня близки друг к другу (большой градиент функции H), то эти множества уровня «сжимают» трубку в поперечном направлении, а поэтому должны растягивать вдоль в силу сохранения площади. Это растяжение соответствует увеличению скорости в областях большого градиента функции H . То же самое рассуждение работает и в более высоких размерностях, но тогда растяжение соответствует увеличению объема на множествах уровня.

Теорема 6.2.5. Если отображение $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ сохраняет объем, а H — инвариантная функция, тогда на каждом множестве уровня можно определить функцию объема Ivol , которая инвариантна относительно f . Иными словами, если для любого открытого подмножества O не критического множества уровня мы определяем $\text{Ivol}(O) := \int_O \frac{1}{\|\text{grad } H\|}$, тогда $\text{Ivol}(O) = \text{Ivol}(f(O))$ для любого такого O .

Доказательство. Как и прежде, выберем точку p (не критическую для H) на множестве уровня $A_h := \{p: H(p) = h\}$ функции H и возьмем ортонормированное множество $v_1, \dots, v_{n-1}$, состоящее из касательных векторов к A_h в точке p . Пусть v_n — единичный нормальный вектор к A_h в точке p (направленный в ту же сторону, что и $\text{grad } H$). Рассмотрим параллелепипед P , порожденный векторами $\epsilon v_1, \dots, \epsilon v_n$, объем которого равен ϵ^n , и обозначим параллелепипед, порожденный векторами

$\epsilon v_1, \dots, \epsilon v_{n-1}$, через Q . Образ $f(P)$ — это по существу параллелепипед, порожденный векторами $\epsilon Df v_1, \dots, \epsilon Df v_n$. Объем последнего не изменится, если мы заменим вектор $\epsilon Df v_n$ на его проекцию на нормаль v'_n к A_h в точке $f(p)$ (поскольку объем равен площади основания, умноженной на высоту). Обозначим длину этой проекции через δ . Линейная аппроксимация приводит к приближенным соотношениям

$$\begin{aligned} H(p + \epsilon v_n) &\approx H(p) + \epsilon \|\operatorname{grad} H_p\|, \\ H(f(p) + v'_n) &\approx H(f(p)) + \delta \|\operatorname{grad} H_{f(p)}\|. \end{aligned}$$

Левые части и первые члены в правых частях совпадают, так что (с точностью до малой погрешности) мы имеем $\delta = \epsilon \|\operatorname{grad} H_p\| / \|\operatorname{grad} H_{f(p)}\|$. Поскольку объем $\delta \operatorname{vol}(f(Q))$ параллелепипеда $f(P)$ равен объему $\epsilon \operatorname{vol}(Q)$ для P , мы получаем, что

$$\operatorname{Ivol}(Q) = \frac{\operatorname{vol}(Q)}{\|\operatorname{grad} H_p\|} = \frac{\operatorname{vol}(f(Q))}{\|\operatorname{grad} H_{f(p)}\|} = \operatorname{Ivol}(f(Q)). \quad \square$$

6.2.4. Первые интегралы. Ключевым наблюдением при нашем изучении математического маятника был тот факт, что полная энергия сохраняется, т. е. является *первым интегралом*. Следовательно, двумерное фазовое пространство системы распадается в инвариантные одномерные кривые уровня постоянной энергии. На каждой такой кривой имеются только несколько простых возможностей для поведения решений, так как по существу мы имеем дело с автономными дифференциальными уравнениями первого порядка. Для *регулярных* уровней энергии, т. е. не критических значений полной энергии, векторное поле не обращается в нуль. Следовательно, если решение ограничено, оно должно быть периодическим, а если решение не ограничено, то оно стремится к бесконечности по одной из не критических кривых уровня энергии. Как мы видели в случае маятника, если время стремится к $+\infty$ и $-\infty$, непостоянные решения на критическом уровне энергии асимптотически притягиваются к (возможно различным) постоянным решениям (ср. с § 2.3). Хотя в принципе возможна и патологическая ситуация, когда критическое решение не сходится к неподвижной точке, а вместо этого блуждает около целой кривой неподвижных решений, этого все же не происходит в естественных моделях. Таким образом, для ньютоновых систем с одной степенью свободы приведенное выше простое описание позволяет провести достаточно полный качественный анализ поведения орбит.

6.2.5. Центральные силы. Небесная механика, т. е. наука о движении планет вокруг Солнца, спутников вокруг планет и т. д., является одним из главных разделов классической механики. Самая простая модель рассматривает два тела, двигающихся свободно, но подчиняющихся закону взаимного притяжения. Можно либо перейти к координатам с началом в центре тяжести системы, либо предположить, что одно из тел (Солнце) намного тяжелее другого, а следовательно, по существу является стационарным (или двигающимся, но с постоянной скоростью). В любом случае положение второго тела (планеты) можно считать точкой $x \in \mathbb{R}^3 \setminus \{0\}$, а его скорость — вектором $v \in \mathbb{R}^3$. Потенциальная энергия гравитационного поля равна $V(x) = -1/\|x\|$, и уравнение Ньютона принимает вид

$$\ddot{x} = \nabla \frac{1}{\|x\|} = -\frac{x}{\|x\|^3} \quad \text{или} \quad \begin{aligned} \dot{x} &= v, \\ \dot{v} &= -\frac{x}{\|x\|^3}. \end{aligned}$$

Кинетическая энергия, как обычно, равна $\langle v, v \rangle / 2$. Таким образом, полная энергия равна $E(x, v) = \langle v, v \rangle / 2 - 1/\|x\|$. Она сохраняется, так как наши уравнения имеют вид (6.2.1). Имеются и другие константы движения, а именно компоненты вектора *углового момента* $x \times v = (x_2 v_3 - x_3 v_2, x_3 v_1 - x_1 v_3, x_1 v_2 - x_2 v_1)$. Чтобы проверить это, заметим, например, что

$$\frac{d}{dt}(x_1 v_2 - x_2 v_1) = \dot{x}_1 v_2 + x_1 \dot{v}_2 - \dot{x}_2 v_1 - x_2 \dot{v}_1 = v_1 v_2 - \frac{x_1 x_2}{\|x\|^3} - v_2 v_1 + \frac{x_2 x_1}{\|x\|^3} = 0. \quad (6.2.4)$$

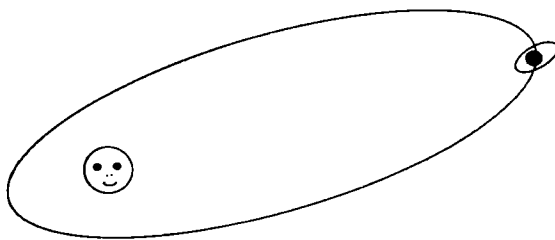


Рис. 6.2.5. Центральная сила

(См. также лемму 6.2.6). Мы будем описывать динамику, решая уравнения движения в явном виде. Поскольку $v \perp x \times v$, движение происходит в плоскости, перпендикулярной векторному произведению $x \times v$. Таким образом, для любого данного направления $x \times v$ задача сводится к задаче в $\mathbb{R}^2 \setminus \{0\}$, т.е. мы можем считать, что $x_3 = v_3 = 0$ (после подходящей замены координат).

Попутно заметим, что $x_1 v_2 - x_2 v_1$ есть удвоенная площадь треугольника с вершинами 0 , x и $x + v$. Таким образом, $x_1 v_2 - x_2 v_1$ — это удвоенная производная площади, заштрихованной вектором x . Тот факт, что эта производная остается постоянной, известен как второй закон Кеплера: луч от Солнца до планеты заметает равные площади за равные интервалы времени. Если $A := x_1 v_2 - x_2 v_1 \neq 0$, то можно показать, что орбиты находятся на конических сечениях. Вспомним из аналитической геометрии, что в полярных координатах конические сечения задаются уравнением $r = ed/(1 + e \cos(\theta - \theta_0))$ с эксцентриситетом $e \in (0, 1)$ для эллипсов, $e = 1$ для парабол и $e > 1$ для гипербол. Положив $r = \|x\|$, мы получим

$$\frac{d}{dt} \left(\frac{x_1}{r} \right) = \frac{v_1 r^2 - x_1(x, v)}{r^3} = -(x_1 v_2 - x_2 v_1) \frac{x_2}{r^3} = A v_2,$$

так что $A v_2 = x_1/r + C$ для некоторого $C \in \mathbb{R}$. Аналогично $A v_1 = -x_2/r - D$. Тогда $C x_1 + D x_2 + r = A x_1 v_2 - x_1^2/r - A x_2 v_1 - x_2^2/r + r = (x_1 v_2 - x_2 v_1) = A^2$, и в полярных координатах $x_1 = r \cos \alpha$, $x_2 = r \sin \alpha$ мы получаем

$$r(\alpha) = \frac{r A^2}{r + C x_1 + D x_2} = \frac{A^2}{1 + C \cos \alpha + D \sin \alpha} = \frac{A^2}{1 + \sqrt{C^2 + D^2} \cos(\alpha - \beta)}, \quad (6.2.5)$$

где $\cos \beta = C/\sqrt{C^2 + D^2}$, $\sin \beta = D/\sqrt{C^2 + D^2}$, т.е. угол β таков, что расстояние $r(\beta)$ является минимальным (*угол перигелия*). Уравнение (6.2.5) является уравнением конических сечений с эксцентриситетом $e = \sqrt{C^2 + D^2}$, который выражается через E и A , т.е. через значения энергии и углового момента:

$$\begin{aligned} e^2 = C^2 + D^2 &= \left(\frac{A v_2 - x_1}{r} \right)^2 + \left(\frac{A v_1 + x_2}{r} \right)^2 = \\ &= \frac{x_1^2 + x_2^2}{r^2} + 2A^2 \frac{v_1^2 + v_2^2}{2} - 2A \frac{x_1 v_2 - x_2 v_1}{r} = 1 + 2EA^2. \end{aligned} \quad (6.2.6)$$

Рис. 6.2.6. Второй закон Кеплера

Таким образом, орбита является эллипсом, если $E < 0$, гиперболой, если $E > 0$, и параболой, если $E = 0$. Следует подчеркнуть два главных качественных свойства решений задачи о центральных силах. Все ограниченные орбиты являются периодическими (эллиптические орбиты). Все неограниченные орбиты стремятся к бесконечности как в положительном, так и в отрицательном времени (гиперболические и параболические орбиты). Эта простая дихотомия — определенное свойство гравитационного потенциала, т. е. выбор существенно зависит от показателя степени величины $r = \|x\|$, появляющегося в потенциале V . Для других степеней величины r ограниченные орбиты, как правило, непериодические. Рассмотрим в качестве примера такое следствие из общей теории относительности. Планета Меркурий имеет несколько иной показатель степени величины r в потенциале. Соответственно и угол перигелия медленно изменяется с течением времени, т. е. орбита представляет собой почти эллипс, но она не замкнута. Это медленное изменение угла перигелия называется прецессией. На самом деле взаимодействие Меркурия с Венерой также приводит к некоторому изменению угла, довольно хорошо согласующемуся с теорией Ньютона. Однако благодаря повышению точности наблюдений в XIX в. удалось обнаружить разность между наблюдаемой прецессией ($5,70''$ в год) и прецессией, которая должна происходить из-за взаимодействия планет согласно закону притяжения Ньютона ($5,27''$ в год). (Меркурий — удобный объект для наблюдений, поскольку его орбита более вытянутая, чем орбиты других планет Солнечной системы: эксцентриситет орбиты Меркурия составляет 0,2056, перигелий и афелий соответственно равны $4,59 \cdot 10^7$ км и $6,97 \cdot 10^7$ км.) Применяя общую теорию относительности, мы получаем необходимую поправку³.

6.2.6. Гармонический осциллятор. Простая задача центральных сил задается потенциалом $V(x) = \|x\|^2$ на плоскости. Уравнение Ньютона приобретает вид

$$\ddot{x} = -\nabla \|x\|^2 = -x$$

и покомпонентно раскладывается в гармонические осцилляторы. Поэтому решения — это независимые колебания с одинаковыми частотами в каждой координате. Плоские орбиты соответственно проходят по эллипсам с центрами в начале координат.

Периодические решения получаются только для потенциалов $\|x\|^2$ и $\|x\|^{-2}$. Это послужило одной из причин, благодаря которым Ньютон сделал вывод о том, что сила тяжести (по крайней мере локально) задается потенциалом обратного квадрата.

6.2.7. Сферический маятник. Рассмотрим несложную на вид систему с центральной силой, определяемую сферическим маятником, т. е. точечной массой, прикрепленной к точке стержнем и находящейся под действием тяжести. Легко написать уравнения движения, если заметить, что мы можем использовать такую же потенциальную энергию V , как в п. 6.2.1.4. Потенциальная энергия определяется высотой точки над положением равновесия, т. е. $U(x) = 1 - \sqrt{1 - x_1^2 - x_2^2}$. В этом случае второй первый интеграл движения, независимый от энергии, — это угловой момент относительно вертикальной оси, т. е. третья координата углового момента. (Это связано с естественной осевой симметрией системы.) Чтобы описать движение для фиксированных

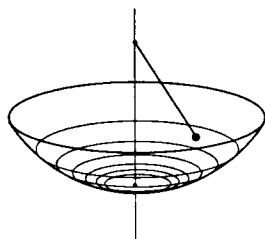


Рис. 6.2.7. Сферический маятник

³A. Einstein. Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie // Sitzungsberichte der königlich preußischen Akademie der Wissenschaften. 1915. Bd. XLVII. S. 831—839. Имеется перевод: А. Эйнштейн. Объяснение движения перигелия Меркурия в общей теории относительности // Собрание научных трудов. Т. I. М.: Наука, 1965. С. 439—447.

значений обоих интегралов, мы воспользуемся полярными координатами. Они хорошо приспособлены к осевой симметрии и к тому факту, что сила $-\nabla U$ радиальна, т. е. направлена к началу координат. Запишем $x = (x_1, x_2) = (r \cos \theta, r \sin \theta)$.

Лемма 6.2.6. *В центральном поле сил угловой импульс сохраняется.*

Доказательство. Угловой импульс определяется как векторное произведение $M := x \times \dot{x}$. По правилу дифференцирования произведения $\dot{M} = \dot{x} \times \dot{x} + x \times \ddot{x} = x \times \ddot{x} = 0$, поскольку в центральном поле сил векторы x и $\ddot{x}$ коллинеарны. $\square$

Чтобы выразить угловой импульс в полярных координатах, мы выберем базис, состоящий из радиального единичного вектора v_r и углового единичного вектора v_θ , перпендикулярного вектору v_r и направленного в сторону увеличения θ , где оба базисных вектора зависят от времени. Тогда $\dot{v}_r = \dot{\theta} v_\theta$ и $\dot{v}_\theta = -\dot{\theta} v_r$, так что $\dot{x} = (d/dt)(\|x\|v_r) = d\|x\|/dt v_r + \|x\|\dot{v}_r = \dot{r}v_r + r\dot{\theta}v_\theta$, и, следовательно, угловой импульс имеет вид

$$M = X \times \dot{x} = x \times \dot{r}v_r + x \times r\dot{\theta}v_\theta = r\dot{\theta}x \times v_\theta = r^2\dot{\theta}v_r \times v_\theta.$$

В силу леммы 6.2.6 величина $r^2\dot{\theta}$ остается постоянной.

Мы можем использовать это обстоятельство для упрощения задачи, найдя уравнение движения для r , которое не зависит от θ .

Дифференцируя равенство $\dot{x} = \dot{r}v_r + r\dot{\theta}v_\theta$ и используя соотношения $\dot{v}_r = \dot{\theta}v_\theta$ и $\dot{v}_\theta = -\dot{\theta}v_r$, мы получим

$$-\frac{\partial U}{\partial r}v_r = -\nabla U = \ddot{x} = (\ddot{r} - r\dot{\theta}^2)v_r + (2\dot{r}\dot{\theta} + r\ddot{\theta})v_\theta,$$

поэтому $\ddot{r} - r\dot{\theta}^2 = -\partial U/\partial r$ и $2\dot{r}\dot{\theta} + r\ddot{\theta} = 0$. Подставляя $\dot{\theta} = \|M\|/r^2$ (угловой импульс), мы получаем $\ddot{r} = -\partial U/\partial r + \|M\|/r^3$ — искомое уравнение, не содержащее θ . Поскольку $U = 1 - \sqrt{1 - r^2}$, мы имеем $\partial U/\partial r = r/\sqrt{1 - r^2}$, и, следовательно, $\ddot{r} = (\|M\|/r^3) - (r/\sqrt{1 - r^2})$, где $\|M\|$ определяется из начальных условий.

Это и есть обещанное уравнение движения, которое включает только r .

6.2.8. Уравнение Лагранжа и вариационный подход. Используя функцию

$$L(x, v) = \frac{1}{2}m\langle v, v \rangle - V(x), \quad (6.2.7)$$

мы можем переписать уравнение Ньютона (6.2.1) в виде

$$\frac{d}{dt} \frac{\partial L}{\partial v} = \frac{\partial L}{\partial x}. \quad (6.2.8)$$

Это уравнение называется *уравнением Лагранжа* или *Эйлера—Лагранжа*. Лагранж ввел этот формализм, в частности, по той причине, что если, как показано выше, использовать равенство $\dot{f} = ta$, мы столкнемся с трудностями при рассмотрении систем со связями. Например, трехмерный математический маятник состоит из точечной массы, которая прикреплена стержнем к неподвижной точке и благодаря этому удерживается на сфере (см. п. 6.2.7). Для работы с этой системой необходимо ввести понятие связей и сил, которые всегда должны быть именно такими, чтобы связи сохранялись. В данном случае формализм Лагранжа существенно упрощает задачу, поскольку он не зависит от выбора системы координат.

Теорема 6.2.7. Пусть L — гладкая функция координат $(x, v) \in \mathbb{R}^n \times \mathbb{R}^n$, $x, y \in \mathbb{R}^n$ и $T > 0$. Определим функционал действия Лагранжа

$$F(c) := \int_0^T L(c(t), \dot{c}(t)) dt \quad (6.2.9)$$

на таких параметризованных гладких кривых $c: [0, T] \rightarrow \mathbb{R}^n$, что $c(0) = x$, $c(T) = y$. Кривая c является критической точкой для F тогда и только тогда, когда она удовлетворяет уравнению (6.2.8).

Доказательство. Пусть L — гладкая функция координат $(x, v) \in \mathbb{R}^n \times \mathbb{R}^n$, $x, y \in \mathbb{R}^n$ и $T > 0$. Рассмотрим такие гладкие кривые $c: [0, T] \rightarrow \mathbb{R}^n$, что $c(0) = x$, $c(T) = y$. Тогда функционал действия Лагранжа (6.2.9) определен корректно. Чтобы найти такую кривую c , что значение $F(c)$ минимально, рассмотрим кривые $c_s: [0, T] \rightarrow \mathbb{R}^n$, гладко зависящие от $s \in (-\epsilon, \epsilon)$ и удовлетворяющие условиям $c_0 = c$ и $c_s(0) = x$, $c_s(T) = y$. Тогда $F(c_s)$ — вещественнозначная функция переменной s , и если значение $F(c_0)$ минимально, то c — критическая точка функционала F , потому что для любых таких кривых c_s интегрирование по частям приводит к соотношению

$$\begin{aligned} 0 = \frac{d}{ds} F(c_s) \Big|_{s=0} &= \frac{d}{ds} \Big|_{s=0} \int_0^T L(c_s(t), \dot{c}_s(t)) dt = \int_0^T \left(\frac{\partial L}{\partial x} \frac{dc_s}{ds} \Big|_{s=0} + \frac{\partial L}{\partial v} \frac{d}{ds} \Big|_{s=0} \dot{c}_s(t) \right) dt = \\ &= \left[\frac{\partial L}{\partial v} \frac{dc_s}{ds} \Big|_{s=0} \right]_0^T - \int_0^T \left(\frac{d}{dt} \frac{\partial L}{\partial v} - \frac{\partial L}{\partial x} \right) \frac{dc_s}{ds} \Big|_{s=0} dt = - \int_0^T \left(\frac{d}{dt} \frac{\partial L}{\partial v} - \frac{\partial L}{\partial x} \right) \frac{dc_s}{ds} \Big|_{s=0} dt \end{aligned}$$

(мы используем то обстоятельство, что $\frac{dc_s}{ds} \Big|_{s=0} = 0$ для $t = 0, T$). Последний интеграл равен нулю независимо от значений $\frac{dc_s}{ds}$ вдоль c_0 . Тогда $\frac{d}{dt} \frac{\partial L}{\partial v} - \frac{\partial L}{\partial x} = 0$, потому что иначе существовал бы такой момент времени $t \in (0, T)$, что это выражение было бы отлично от нуля в точке $c_0(t)$; выбирая такую кривую c_s , что $\frac{dc_s}{ds}(t) = \frac{d}{dt} \frac{\partial L}{\partial v}(c_0(t)) - \frac{\partial L}{\partial x}(c_0(t))$ и $\frac{dc_s}{ds} = 0$ вне малой окрестности точки t , мы получаем, что интеграл отличен от нуля, а это противоречит нашему выбору c_0 в качестве критической точки.

Таким образом, мы приходим к уравнению Лагранжа (6.2.8) при минимизации интегралов по кривым, и если функция L выбрана в соответствии с равенством (6.2.7), то критические точки — это в точности решения уравнения Ньютона. $\square$

Итак, решение уравнения Лагранжа (6.2.8) — а следовательно, и описание динамики Ньютона — сводится к решению вариационной задачи, т.е. нахождению критических точек некоторого функционала. Это соответствует эвристическому принципу, согласно которому многие естественные процессы определенным образом оптимизированы. Естественный функционал действия, который приводит к уравнению (6.2.8), определен на бесконечномерном пространстве. Это приводит к значительным техническим осложнениям. Поэтому мы не будем использовать здесь этот подход. Однако в ситуации бильярда с дискретным временем, введенного в § 6.3, мы рассмотрим функцию действия с конечным числом переменных, и вариационный подход позволит получить полезные результаты. Подробнее об этом говорится в гл. 14.

Рассмотрим свободно движущуюся материальную точку, для которой функция Лагранжа не имеет слагаемого, соответствующего потенциальной энергии. Из уравнения Лагранжа (6.2.8) следует, что орбиты минимизируют действие, которое связано с энергией. Это означает, что орбиты также минимизируют расстояние между любыми из точек (если только они расположены не слишком далеко друг от друга). Поэтому орбиты являются *геодезическими*, т.е. кривыми, которые локально минимизируют длину. Интуитивно это соответствует тому факту, что свободно перемещающаяся материальная точка движется по линии, являющейся обобщением прямой, — геодезической, т.е. «прямой линии» искривленного пространства. Соответствующий поток называется *геодезическим потоком*.

Упражнения

6.2.1. Камень, брошенный в колодец, достигает дна за одну секунду. Какова глубина колодца?

6.2.2. Рассмотрим частицу, находящуюся под действием лишь силы тяжести. Напишите уравнение Ньютона, рассматривая координаты частицы (x, y, z) как функции времени (где z — высота), и решите получающееся дифференциальное уравнение.

6.2.3. Футбольный мяч подброшен с начальной вертикальной скоростью 30 м/с. Насколько высоко он взлетит?

6.2.4. Докажите лемму 6.2.6, используя вычисления в координатах, подобные (6.2.4).

6.2.5. Используя элементарную сферическую геометрию, опишите динамику геодезического потока на сфере (а именно, на единичной сфере в $\mathbb{R}^3$).

Задачи для углубленного изучения

6.2.6. Рассмотрим систему из n материальных точек в $\mathbb{R}^3$, считая, что их попарное взаимодействие зависит только от расстояний между ними, т. е. $V(x) = \sum V_{ij}(\|x_i - x_j\|)$. Покажите, что координаты скорости центра тяжести и углового момента являются первыми интегралами.

6.2.7. (Задача двух тел на плоскости) Покажите, что для системы двух материальных точек на плоскости, взаимодействие между которыми описывается так же, как и в предыдущем упражнении, четыре интеграла (энергия, угловой момент и координаты скорости центра тяжести) независимы. Опишите движение относительно центра тяжести.

6.2.8. Найдите решения задачи о математическом маятнике, используя метод, описанный в п. 6.2.2.6.

6.2.9. Докажите, что орбиты замкнуты лишь для тех случаев, когда центральная сила линейно зависит от расстояния или обратно пропорциональна его квадрату.

§ 6.3. Биллиард: определение и примеры

В п. 4.2.5 мы изучали класс динамических систем, которые могут рассматриваться как механические или оптические. Механическая модель — модель, в которой материальная точка перемещается в ограниченной области, упруго отражаясь от стенок. Такие системы называются биллиардными потоками. В п. 4.2.5 эти системы были получены при рассмотрении простой системы с двумя частицами на отрезке. На самом деле модель биллиарда возникает и во многих других ситуациях, но независимо от того, получены ли биллиарды из конкретной модели, их изучение очень показательно по следующей причине.

В таких задачах мы практически не сталкиваемся с формальными вопросами, которые, как правило, представляют большую трудность в динамике, и можем рассматривать только интересные качественные вопросы.⁴

Таким образом, биллиарды легче поддаются детальному изучению, хотя они и относятся к системам со сложной динамикой. Хотя в п. 4.2.5 биллиард возник при описании системы двух точек на отрезке прямой, в п. 5.2.3 мы видели, что эта модель на самом деле применима для любого числа точек на отрезке.

⁴*G. D. Birkhoff*. Dynamical systems. Providence, RI: AMS, 1927. (American Mathematical Society Colloquium Publications, v. 9). P. 170. Имеется перевод: *Дж. Биркгоф*. Динамические системы. М.—Л., Гостехиздат, 1941.

В этом и следующем параграфах мы сосредоточим внимание на изучении бильярдов, с которыми мы еще не сталкивались в п. 4.2.5 или даже 5.2.3, а именно выпуклых бильярдов, т. е. бильярдов, в которых стол имеет гладкую выпуклую границу, подобно окружности или эллипсу (рис. 2.2.2). Однако не все результаты здесь зависят от выпуклости.

6.3.1. Бильярдный поток. Рассмотрим движение материальной точки (или светового луча) в ограниченной области D на плоскости с границей B . В случае традиционного бильярда (бильярдного стола в клубе) областью является прямоугольник, а в случае примера из п. 4.2.5 — треугольник. Орбитами движения являются последовательности отрезков прямых в D , где любые два последовательных отрезка имеют общую точку на границе, и в этой точке эти два отрезка составляют один и тот же угол с касательной к границе. Поэтому угол падения равен углу отражения, так же как в случае с зеркалом (см. рис. 6.3.1). Если

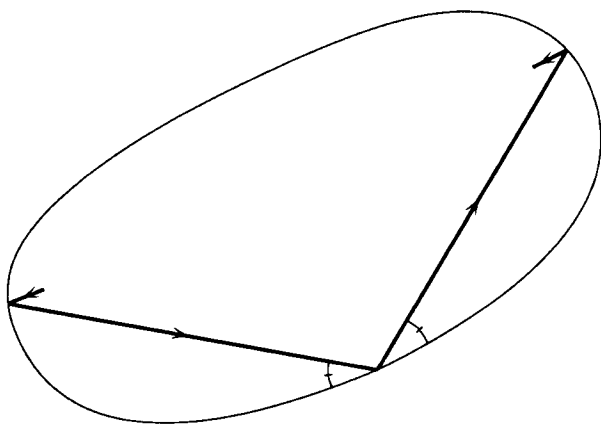


Рис. 6.3.1. Бильярд

орбита попадает в угол на границе, то движение заканчивается (потому что отражение в такой точке не определено). Можно представлять себе область, в которой происходит движение, как стол с лузами по углам. Скорость движения считается постоянной (трение отсутствует). Каждая орбита полностью определена, если заданы начальное положение и начальное направление движения, иными словами, фазовое пространство этой системы — это множество всех касательных векторов фиксированной длины (например, единичных), которые лежат внутри D , а также векторов с началами в точках границы, направленных внутрь. Мы можем описать такие точки евклидовыми координатами (x_1, x_2) начальной точки и циклической угловой координатой α вектора направления.

6.3.2. Бильярдное отображение. Бильярдный поток является системой с непрерывным параметром времени, но в те моменты времени, в которые происходит отражение, изменение направления является разрывным. С учетом этого обстоятельства в п. 4.2.5.3 была введена конструкция развертки для бильярда

в многоугольнике. Для выпуклого биллиарда построить развертку не удастся, и лучше описывать систему другим способом, игнорируя время между столкновениями с границей и используя дискретное время, т. е. строя *отображение сечения*, которое по данной конфигурации столкновения (граничная точка вместе с вектором, направленным внутрь) определяет следующую конфигурацию столкновения. Мы не теряем никакой информации, потому что два последовательных столкновения вполне определяют отрезок между ними. Поэтому мы рассматриваем только множество C граничных точек с приложенными к ним векторами, направленными внутрь, и определяем отображение ϕ на этом множестве C , приписывая каждому из этих начальных условий место следующего столкновения и направление отражения. Мы можем рассматривать это отображение, даже если имеются углы на границе, оно просто не определено в таких точках.

Отображение $\phi: C \rightarrow C$ обычно называется *отображением биллиардного шара* или просто *биллиардным отображением* и более детально описывается следующим образом: вектор $v \in C$ с началом в точке $p \in B$ определяет ориентированную прямую l , которая пересекает B в двух точках p и p' . Тогда $\phi(v)$ — это вектор с началом в точке p' , направленный внутрь по прямой, полученной отражением l от касательной к B в точке p' . Естественные координаты в фазовом пространстве C — это циклический параметр длины $s \in [0, L)$ на B , где L есть полная длина границы B (заметим, что согласно п. 2.6.2 отождествление L с 0 превращает этот полуинтервал в окружность), и угол $\theta \in (0, \pi)$ с положительно ориентированным касательным направлением. Таким образом, фазовое пространство — это цилиндр (см. п. 2.6.3). Заметим, что если точка p фиксирована, то увеличение угла θ приводит к монотонному увеличению значения p' (см. рис. 6.3.2). Это означает, что отображение, определенное таким образом на

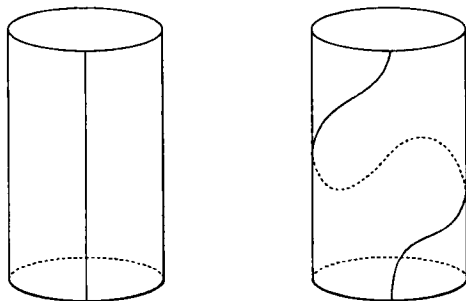


Рис. 6.3.2. Фазовое пространство биллиарда

цилиндре, является закручивающим (см. определение 14.2.1). С частным случаем такой ситуации мы встречались в примере 6.1.2.

6.3.3. Модели биллиарда. Биллиардное отображение на цилиндре не описывает полностью биллиардный поток, потому что оно не позволяет определить промежутки времени между столкновениями. Однако эти промежутки можно вычислить как длины отрезков между точками столкновений. Мы впервые встретились с биллиардом в п. 4.2.5, где эта система явилась результатом моделирования движения двух частиц на отрезке прямой. Выпуклые биллиарды, о которых идет

речь в этой главе, также появляются как соответствующие модели для других задач. Слова Биркгофа, процитированные в начале § 6.3, были сказаны в связи со следующей аналогией. Рассмотрим материальную точку, свободно движущуюся по выпуклой поверхности без внешних сил. Иными словами, материальная точка связана только с поверхностью и перемещается исключительно под действием сил инерции. В качестве довольно дорогой физической реализации этой системы можно было бы взять поверхности в виде полости, которая стационарна при отсутствии силы тяжести. Капля ртути перемещается по внутренней поверхности этой полости описанным способом (будучи связанной с поверхностью центробежной силой). Иной путь описания этой комбинации связей и свободного движения состоит в том, чтобы считать, что ускорение всегда направлено по нормали к поверхности (потому что единственная сила, действующая на материальную точку, — это связь).

Если рассматриваемая поверхность — это трехосный эллипсоид и мы сжимаем его, делая одну ось короче, тогда предельная динамика при уменьшении самой короткой оси до нуля будет такой же, как и в случае бильярда с эллиптическим столом. Хотя бильярд и не является точной моделью свободного движения на любом эллипсоиде, наше описание динамики эллиптического бильярда во многом сходно с описанием свободного движения материальной точки на эллипсоиде, но описывается легче. Подобные аналогии имеются и между другими формами бильярдных столов и свободными движениями материальной точки на соответствующих поверхностях. Бильярдные модели помогли открыть свойства движения, аналоги которых были позднее установлены для соответствующих поверхностей.

6.3.4. Окружность. Самый простой выпуклый бильярд — окружность. Пусть D — единичный диск с границей $B = \{(x, y) : x^2 + y^2 = 1\}$. Бильярдное отображение для этого бильярда может быть описано явно с использованием

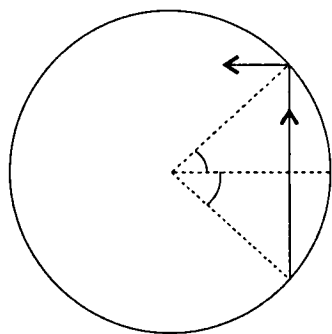


Рис. 6.3.3. Бильярдное отображение в круге

циклического параметра длины s на окружности и угла $\theta \in (0, \pi)$ с положительным касательным направлением. Таким образом, фазовое пространство отображения — цилиндр $C = S^1 \times (0, \pi)$, где s — угловая координата на C .

1. *Бильярдное отображение.* Отображение ϕ для бильярда имеет вид $(s', \theta') = (s + 2\theta, \theta)$, так что угол θ — константа движения (т. е. является постоянным вдоль каждой орбиты). Заметим, что это по существу линейное закручивание из примера 6.1.2, которое показано на рис. 6.1.1, и что оно также является результатом свободного движения материальной точки на торе (см. п. 5.2.2). Это означает, что цилиндр C распадается на ϕ -инвариантные окружности $\theta = \theta_0$. Динамика на инвариантной окружности — поворот на угол $2\theta_0$, и для любой орбиты бильярда последовательные столкновения с границей B лежат на орбите поворота границы на угол $2\theta_0$. Следовательно,

с границей B лежат на орбите поворота границы на угол $2\theta_0$. Следовательно,

если угол θ_0 соизмерим с 2π (т. е. $\theta_0/\pi \in \mathbb{Q}$, градусная величина угла θ_0 рациональна), тогда бильярдное отображение на окружности периодическое и орбиты представляют собой вписанные звездчатые многоугольники. Если угол θ_0 несоизмерим с 2π , тогда согласно предложению 4.1.1 все орбиты плотны на окружности (см. рис. 6.3.4).

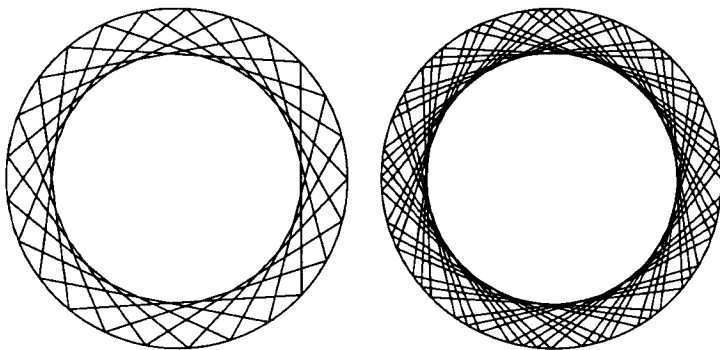


Рис. 6.3.4. Рациональная орбита и часть иррациональной орбиты кругового бильярда

2. Каустики. Инвариантная окружность $\theta = \theta_0$ соответствует всем тем лучам, которые составляют угол θ_0 с границей $B = \{(x, y): x^2 + y^2 = 1\}$. Объединение этих лучей представляет собой круговое кольцо $\cos^2 \theta \leq x^2 + y^2 \leq 1$. Его внутренняя граница $x^2 + y^2 = \cos^2 \theta$ называется *каустикой*, связанной с инвариантной окружностью. Дополнение к этому круговому кольцу есть пересечение всех левых полуплоскостей относительно этих лучей. Каустика, в частности, является огибающей лучей, определяющих ее, т. е. гладкой касательной к каждому лучу из совокупности, или, как в нашем случае, обладает тем свойством, что если луч касается ее, то ее касается и отражение этого луча от границы бильярдного стола.

Рассмотрим крайний случай, когда точка находится в центре круглого стола. Тогда любой луч, проходящий через центр, при отражении переходит в себя, и если считать, что в этом случае вообще существует каустика, то каустикой будет сама эта точка, т. е. эта точка является фокусом. Это наводит на мысль, что каустики являются естественными обобщениями фокусов в случае неполного фокусирования. Соответствующая инвариантная окружность для бильярдного отображения на цилиндре описывается уравнением $\theta = \pi/2$ и состоит из орбит периода 2. Это вполне очевидно как из геометрических соображений, так и из формулы для бильярдного отображения.

3. Вариационный подход. Укажем теперь альтернативный способ описания связи между последовательными точками столкновений траектории бильярда с границей. Если заданы две точки столкновения и известно примерное положение точки столкновения, которое произошло между двумя данными столкновениями, то можно найти промежуточную точку точно, используя закон отражения

(если мы не знаем приблизительное положение промежуточной точки, то получим две противоположные точки). Равенство углов в промежуточной точке связано с тем, что выбор промежуточной точки минимизирует сумму длин двух получающихся лучей. Действительно, если углы не равны, сумму длин можно уменьшить, перемещаясь по направлению к меньшему углу. Заметим, что это наблюдение вообще не зависит от того, находимся ли мы в круге. В самом деле, способ нахождения орбиты, основанный на минимизации какого-либо функционала с данными концевыми точками, уже знаком нам из лагранжевой механики, где мы минимизировали действие. Это обстоятельство не является простым совпадением, а связано с механической природой бильярда. При рассмотрении системы бильярда как оптической установки мы можем также описать этот вариационный подход как принцип Ферма «спешащего светового луча», который достигает цели, следуя по самому короткому маршруту.

6.3.5. Эллипс. Рассмотрим теперь эллиптическую область D с границей $B = \left\{ (x, y): \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \right\}$.

1. *Периодические точки.* В отличие от случая окружности, эллиптический бильярд не имеет инвариантной окружности с орбитами периода 2, проходящими через центр, однако имеются две особые орбиты периода 2 на осях симметрии эллипса. Оси симметрии являются единственными прямыми, которые пересекают эллипс под прямыми углами. Соответственно концевые точки длинной оси симметрии могут быть охарактеризованы как единственная пара граничных точек с максимальным расстоянием между ними. Аналогичным образом, концевые точки короткой оси симметрии могут быть охарактеризованы как седловые точки для функции расстояния между концами. Длина длинной оси равна *диаметру* эллипса, т. е. максимальному расстоянию между двумя его точками. Длина короткой оси равна *ширине*, которая определяется как минимальная ширина полосы (между двумя параллельными прямыми), содержащей эллипс, т. е. ширина самого узкого коридора, через который можно пронести этот эллиптический стол.

2. *Производящая функция.* Мы еще вернемся к экстремальному свойству этих специальных орбит, а сейчас в связи с ними дадим следующее определение. Параметризуем границу B параметром длины дуги s и рассмотрим две точки p и p' с соответствующими координатами s и s' на B . Пусть $H(s, s')$ — евклидово расстояние между точками p и p' , взятое со знаком минус. Функция H называется *производящей функцией* для бильярда (и будет рассматриваться в общем случае в п. 6.4.2). Таким образом, длинная орбита периода 2 соответствует минимуму функции H , а короткая орбита соответствует седловой точке функции H . Мы увидим, что любой выпуклый бильярд имеет по крайней мере две орбиты периода два, которые могут быть описаны подобным образом в терминах диаметра и ширины. Здесь следует вновь отметить аналогию с вариационным подходом в лагранжевой механике.

Для окружности производящая функция равна $H(s, s') = -2 \sin \frac{1}{2}(s' - s)$. Как и следует ожидать, эта функция имеет большое число критических точек,

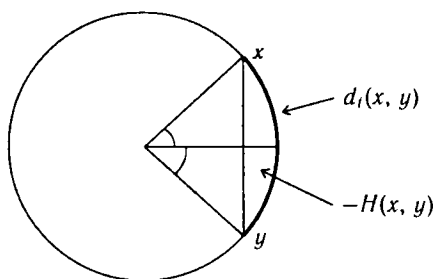


Рис. 6.3.5. Производящая функция окружности

а именно любая такая пара (s, s') , что $s' - s = \pi$, будет критической точкой, и множество критических точек соответствует всем диаметрам.

3. *Каустики*. Биллиард с эллиптическим столом имеет множество каустик.

Предложение 6.3.1. *Всякий меньший софокусный эллипс (т.е. эллипс с теми же фокусами) является каустикой.*

Доказательство. Для иллюстрации доказательства рассмотрим рис. 6.3.6. На нем изображен биллиард с эллиптическим столом с фокусами f_1 и f_2 , луч p_0p_1 , отличный от отрезка, соединяющего фокусы, и его образ p_1p_2 при биллиардном отображении. Как мы видим, эти два луча образуют одинаковые углы

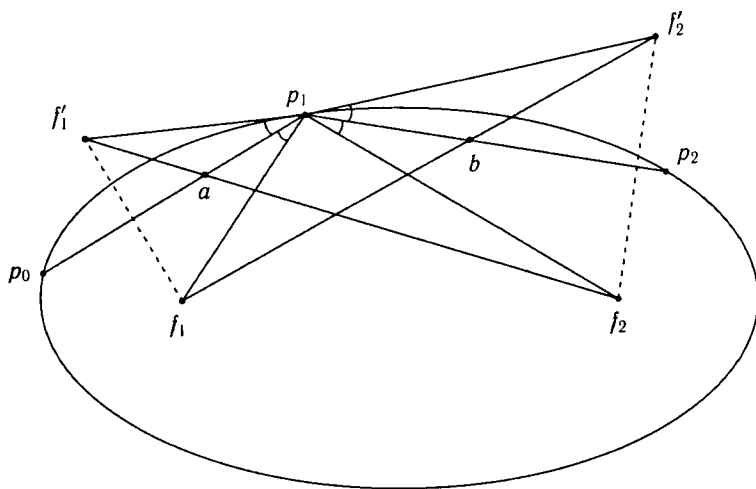


Рис. 6.3.6. Софокусные эллипсы являются каустиками

с касательной в точке p_1 . То же самое справедливо для лучей f_1p_1 и p_1f_2 , являющихся также частью орбиты. Поэтому углы $p_0p_1f_1$ и $f_2p_1p_2$ равны, как показано на рисунке. Теперь отразим f_1p_1 относительно p_0p_1 и f_2p_1 относительно p_2p_1 и обозначим полученные образы через f'_1p_1 и f'_2p_2 соответственно. По построению два новых угла равны двум углам, рассмотренным выше. Поэтому треугольник

$s_1 p_1 f'_2$ получается из треугольника $s'_1 p_1 f_2$ вращением вокруг p_1 , и, следовательно, $l(f_1 f'_2) = l(f'_1 f_2) =: L$.

Заметим теперь, что луч $p_0 p_1$ касается софокусного эллипса в точке a , потому что отражение луча $a f'_1$ от $p_0 p_1$ есть луч $a f_1$, а это может произойти только для отражения от касательной к софокусному эллипсу, определенному уравнением $l(f_2 x) + l(x f_1) = l(f_2 f'_1) = L$, на котором лежит a . Аналогично b является точкой касания с тем же самым эллипсом, определенным уравнением $l(f_1 x) + l(x f_2) = l(f_1 f'_2) = L$. $\square$

Поэтому имеется семейство инвариантных окружностей в фазовом пространстве S этого билиарда, которые соответствуют семействам касательных лучей к данному софокусному эллипсу. Эти окружности могут быть параметризованы, например, (положительным) эксцентриситетом соответствующей эллиптической каустики.

Но это только часть всей картины.

Предложение 6.3.2. *Имеется каустика, соответствующая любому лучу, который проходит между фокусами. Она состоит из (двух ветвей) гиперболы с теми же самыми фокусами.*

Доказательство. Доказательство почти такое же, как для предложения 6.3.1, а соответствующие построения показаны на рис. 6.3.7. Заметим, что мы

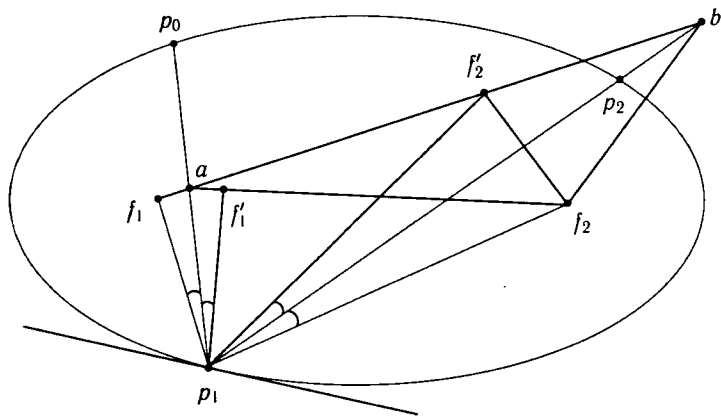


Рис. 6.3.7. Софокусиные гиперболы — каустики

вновь можем добиться равенства $l(f_1 f'_2) = l(f'_1 f_2) =: \Delta$, осуществляя вращение соответствующих треугольников вокруг p_1 , и что a и b — точки касания с гиперболой $l(f_1 x) - l(f_2 x) = \pm \Delta$ (a и b соответствуют здесь противоположным знакам). $\square$

Последовательные касания всегда происходят с противоположными ветвями гиперболы (см. рис. 6.3.8). Таким образом, каждая из этих каустик порождает пару замкнутых инвариантных дуг в S , параметризованных (отрицательным) эксцентриситетом соответствующей гиперболы, на каждую из которых попе-

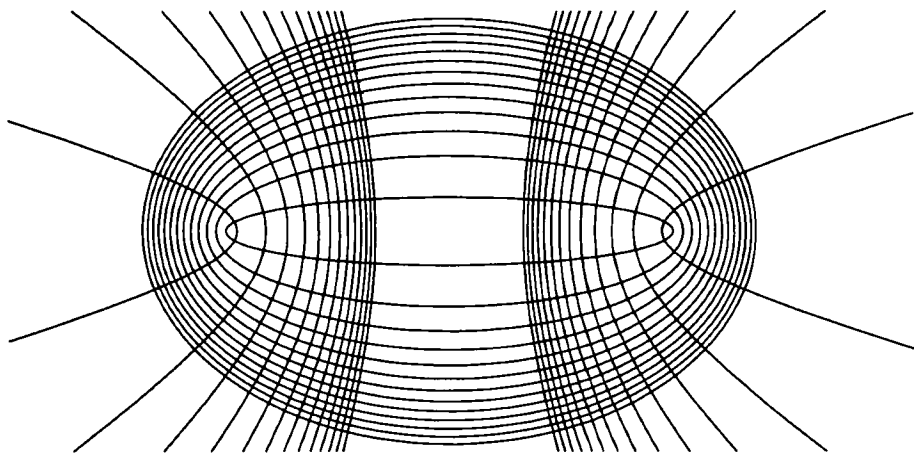


Рис. 6.3.8. Эллиптический биллиард с софокусными эллипсами и гиперболами

ременно попадает биллиардное отображение. Эта совокупность инвариантных множеств отделена от множеств, соответствующих положительному эксцентриситету, кривой, которая соответствует семейству лучей, проходящих через фокусы (см. рис. 6.3.9). Единственная орбита, которая остается вне этой классификации, — это орбита периода 2, соответствующая малой оси эллипса.

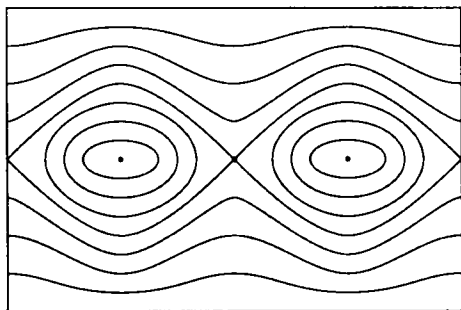


Рис. 6.3.9. Фазовый портрет эллиптического биллиарда

4. Инвариантные окружности. Для изучения движения на инвариантной окружности, соответствующей эллиптической каустике, мы воспользуемся тем, что отображение для биллиарда сохраняет площадь, если взять в качестве второй координаты — $\cos \theta$ вместо θ (см. предложение 6.4.2). Поскольку инвариантные окружности входят в семейство кривых, соответствующих эллиптическим каустикам, в данном случае эксцентриситет является инвариантной функцией на этой части фазового пространства. Как показано после предложения 6.2.4, сохранение площади и наличие инвариантной функции позволяют нам параметризовать эти инвариантные кривые таким способом, чтобы движение (при биллиардном отображении) вдоль каждой из них являлось поворотом окружности. Поэтому мы можем полностью описать динамику биллиардного отображения на этом

множестве: это просто открытое множество попарно различных инвариантных окружностей, каждая из которых вращается в соответствии с подходящей параметризацией.

Замечание 6.3.3. Число вращения может изменяться (в этом легко убедиться, рассмотрев крайние случаи и воспользовавшись непрерывностью).

Упражнения

6.3.1. Опишите бильярдное отображение для прямоугольного бильярда.

6.3.2. Опишите бильярдное отображение в равностороннем треугольнике.

6.3.3. Опишите движение бильярдного шара в круговом кольце между двумя концентрическими окружностями.

6.3.4. Опишите движение бильярдного шара в четверти круга $\{(x, y) \in \mathbb{R}^2 : x > 0, y > 0, x^2 + y^2 \leq 1\}$.

6.3.5. Покажите, что орбита бильярда, проходящая через фокусы эллипса, накапливается на большой оси. (Поскольку то же самое справедливо для обратной орбиты, эти орбиты составляют гетероклиническую петлю, или сепаратрису, для орбиты периода 2, проходящей через оба фокуса.)

Задачи для углубленного изучения

6.3.6. Найдите первый интеграл для бильярдного потока в эллипсе, который является квадратичной функцией каждой из координат и скоростей.

§ 6.4. Выпуклые бильярды

При изучении кругового и эллиптического бильярдов мы встретились с некоторыми особенностями, на которые следует обращать внимание при изучении других бильярдных столов, а именно с наличием периодических точек и каустик. Кроме того, мы определили некоторые понятия, помогающие упорядочить наши исследования. Например, изучая структуру орбит, мы будем рассматривать преимущественно качественные свойства бильярдного отображения на фазовом цилиндре, в противоположность рассуждениям с использованием наглядной геометрии. Это продиктовано прежде всего отсутствием явного описания бильярдного отображения для эллипса, для которого мы тем не менее описали важные качественные особенности отображения, а именно разложение в инвариантные множества, которые легко изучаются индивидуально.

6.4.1. Гладкая выпуклость. Сейчас мы приступим к изучению бильярдов, граница которых задается гладкой замкнутой кривой B , выпуклой в более сильном смысле, чем в определении 2.2.13. Мы потребуем, чтобы граница имела кривизну, отличную от нуля. Это условие эквивалентно тому, что при параметризации кривой B длиной дуги вторая производная никогда не обращается в нуль.

Из этого следует (строгая) выпуклость (см. определение 2.2.13), т. е. отсутствие точек возврата и «выпуклости внутрь», и мы получаем определяющее

свойство, состоящее в том, что каждая прямая, входящая внутрь стола, входит и выходит трансверсально под углом из интервала $(0, \pi)$ и пересекает границу строго в двух точках. Часто бывает достаточно, чтобы выполнялось это последнее геометрическое условие, и при этом допустимо наличие изолированных нулей второй производной.

Имеются ситуации, где условие на производную необходимо, и мы называем билиард, удовлетворяющий этому условию, *строго дифференциально выпуклым*. Это более сильное понятие выпуклости, чем «строгая выпуклость», введенная в определении 2.2.13.

Таким образом, как и в случае круга и эллипса, фазовое пространство представляет собой цилиндр S , параметризованный параметром s на границе (обычно длиной дуги) и углом $\theta \in (0, \pi)$.

6.4.2. Производящая функция. Как и в случае круга и эллипса, определим функцию H , взяв две точки p и p' на границе B с соответствующими значениями параметра длины s и s' и полагая значение $H(s, s')$ равным евклидову расстоянию между p и p' , взятому с противоположным знаком. Функция H называется *производящей функцией* для билиарда. Хотя мы обычно не можем найти явной формулы для H , как в случае билиарда в круге, эта функция все же поддается аналитическому исследованию.

Лемма 6.4.1. Если θ' — угол между отрезком, соединяющим точки p и p' , и отрицательно направленной касательной в точке p' , а θ — угол между отрезком, соединяющим точки p и p' , и положительно направленной касательной в точке p , тогда

$$\begin{aligned}\frac{\partial}{\partial s'} H(s, s') &= -\cos \theta', \\ \frac{\partial}{\partial s} H(s, s') &= \cos \theta.\end{aligned}\tag{6.4.1}$$

(См. рис. 6.4.1.)



Рис. 6.4.1. Производные производящей функции

Доказательство. Воспользуемся соотношениями

$$\begin{aligned}\frac{\partial}{\partial s'} H(s, s') &= -\frac{d}{dt} d(p, c(t)) = -\frac{d}{dt} \sqrt{\langle c(t) - p, c(t) - p \rangle} = \\ &= -\frac{1}{2\sqrt{\langle c(t) - p, c(t) - p \rangle}} 2\langle c'(t), c(t) - p \rangle = -\frac{\langle c(t) - p, c(t) - p \rangle}{\|c(t) - p\|}.\end{aligned}$$

Для $t = s'$ последнее выражение равно в точности $-\cos \theta'$, потому что c' — единичный вектор. Второе равенство доказывается аналогично. $\square$

Производящая функция помогает выяснить, когда последовательность точек на границе лежит на орбите. Любые две точки, конечно же, принадлежат некоторой орбите, но, в отличие от пары точек, три точки не всегда находятся на части одной орбиты. Те тройки, которые лежат на одной орбите, могут быть описаны как критические точки некоторого функционала. Рассмотрим три точки p_{-1} , p_0 и p_1 с соответствующими координатами s_{-1} , s_0 и s_1 на B . Если они являются частью орбиты билиарда, тогда по определению отрезки, соединяющие p_{-1} с p_0 и p_0 с p_1 , образуют одинаковые углы с касательной в точке p_0 . Следовательно, по лемме 6.4.1 мы получаем

$$\frac{d}{ds} H(s_{-1}, s) + \frac{d}{ds} H(s, s_1) = 0 \quad \text{в точке } s = s_0, \quad (6.4.2)$$

т.е. p_0 является *критической точкой* функционала $s \mapsto H(s_{-1}, s) + H(s, s_1)$, определенного на тройках граничных точек. Как и в случае лагранжева формализма, отрезок орбиты динамической системы можно описать как критическую точку функционала, определенного на пространстве «потенциальных» отрезков орбит динамической системы. Повторяя процедуру, мы получим отрезки орбит как критические точки функционалов, зависящих от нескольких переменных.

6.4.3. Сохранение площади. Явные формулы (6.4.1) для производных производящей функции, в свою очередь, полезны для изучения билиардных отображений. Например, из них следует, что билиардное отображение сохраняет площадь на фазовом цилиндре C , если вместо θ использовать координату $r = -\cos \theta$.

Предложение 6.4.2. В координатах (s, r) билиардное отображение $\phi(s, r) = (S(s, r), R(s, r))$ является сохраняющим площадь и ориентацию (см. п. 6.1.1.3).

Доказательство. Уравнения (6.4.1) упрощаются и принимают вид

$$\begin{aligned}\frac{\partial}{\partial s'} H(s, s') &= r', \\ \frac{\partial}{\partial s} H(s, s') &= -r,\end{aligned} \quad (6.4.3)$$

где $r' = -\cos \theta'$. Определим $\tilde{H}(s, r) := H(s, S(s, r))$. Тогда

$$\frac{\partial \tilde{H}}{\partial s} = \frac{\partial H}{\partial s} + \frac{\partial H}{\partial s'} \frac{\partial S}{\partial s} = -r + R \frac{\partial S}{\partial s}$$

и

$$\frac{\partial \tilde{H}}{\partial r} = \frac{\partial H}{\partial s'} \frac{\partial S}{\partial r} = R \frac{\partial S}{\partial r},$$

так что, вычисляя $\frac{\partial^2 \tilde{H}}{\partial s \partial R}$ двумя разными способами, мы получаем

$$-1 + \frac{\partial R}{\partial r} \frac{\partial S}{\partial s} + R \frac{\partial^2 S}{\partial s \partial r} = \frac{\partial^2 \tilde{H}}{\partial s \partial r} = \frac{\partial^2 \tilde{H}}{\partial r \partial s} = \frac{\partial R}{\partial s} \frac{\partial S}{\partial r} + R \frac{\partial^2 S}{\partial r \partial s}$$

и $\frac{\partial R}{\partial r} \frac{\partial S}{\partial s} - \frac{\partial R}{\partial s} \frac{\partial S}{\partial r} = 1$. Это означает, что якобиан отображения ϕ равен 1, а следовательно, ϕ сохраняет площадь и ориентацию (см. предложение 6.1.3 и определение 6.1.4). $\square$

6.4.4. Гладкость билиардного отображения. Уравнения (6.4.3) полезны не только для доказательства свойства сохранения площади. Они также задают динамику, потому что с их помощью мы можем локально определить функции S и R . Во-первых, можно доказать гладкость билиардного отображения.

Предложение 6.4.3. *Предположим, что B — кривая класса C^k , т. е. евклидовы координаты — функции параметра длины класса C^k . Тогда S и R есть функции класса C^{k-1} для $0 < r < 1$.*

Доказательство. Применим теорему о неявной функции (теорема 9.2.3) к отображению

$$0 = F(s, s', r, r') := \begin{pmatrix} \frac{\partial}{\partial s'} H(s, s') - r' \\ \frac{\partial}{\partial s} H(s, s') + r \end{pmatrix}.$$

Условия теоремы о неявной функции выполнены, поскольку матрица полной производной

$$\begin{pmatrix} \frac{\partial^2}{\partial s'^2} H(s, s') & -1 \\ \frac{\partial^2}{\partial s \partial s'} H(s, s') & 0 \end{pmatrix}$$

отображения F относительно (s', r') обратима. Действительно, ее детерминант, равный $\frac{\partial^2}{\partial s \partial s'} H(s, s') = \frac{\partial r'}{\partial s}$, отличен от нуля по геометрическим соображениям: если увеличивать s при фиксированном s' , то θ' и, следовательно, r' уменьшаются (см. рис. 6.4.2), так что

$$\frac{\partial^2}{\partial s \partial s'} H(s, s') = \frac{\partial r'}{\partial s} < 0. \quad (6.4.4)$$

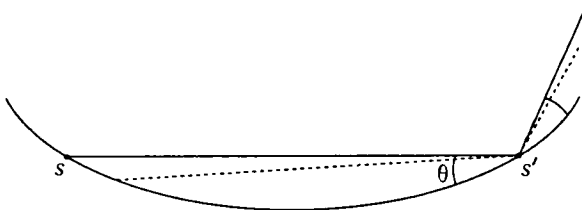


Рис. 6.4.2. Увеличение s при постоянном s'

Если B — кривая класса C^k , то производящая функция также является функцией класса C^k и по теореме о неявной функции (теорема 9.2.3) S и R являются функциями класса C^{k-1} для $0 < r < 1$. $\square$

6.4.5. Специальные орбиты периода два в выпуклом бильярде. Теперь мы обобщим наше наблюдение, сделанное ранее, согласно которому в эллиптическом бильярде можно получить геометрическое описание двух орбит периода два, используя понятия диаметра и ширины области. Мы будем использовать лишь некоторые свойства производящей функции. Интуитивно следующий результат вполне очевиден.

Предложение 6.4.4. Пусть D — выпуклая ограниченная область, граница которой B — кривая класса C^2 с кривизной, отличной от нуля. Тогда соответствующее бильярдное отображение имеет по крайней мере две различные орбиты с периодом 2, которые описываются следующим образом: для одной из них расстояние между соответствующими граничными точками равно диаметру области D , а для другой это расстояние равно ширине области D .

Доказательство. Производящая функция $H(s, s')$ определена и непрерывна на торе $B \times B$ и дифференцируема везде, кроме диагонали. Так как она равна нулю на диагонали и отрицательна в других точках, она достигает минимума d вне диагонали. Пусть точка (s, s') такова, что $H(s, s') = d$. Поскольку это критическая точка, из уравнений (6.4.1) следует, что $\theta = \theta' = \pi/2$, так что мы получаем первую из орбит с периодом два. (В предыдущем рассуждении используется только выпуклость, и его можно легко модифицировать для кривых класса C^1 .) Теперь рассмотрим кривую $(s, g(s))$ на торе, где $s' = g(s)$ — координата граничной точки, отличной от s , на прямой, проходящей через s , для которой $\theta = -\theta'$. (Эта прямая соединяет две точки с параллельными касательными, так что минимальная из длин таких отрезков и есть ширина области; см. рис. 6.4.3.) На этой

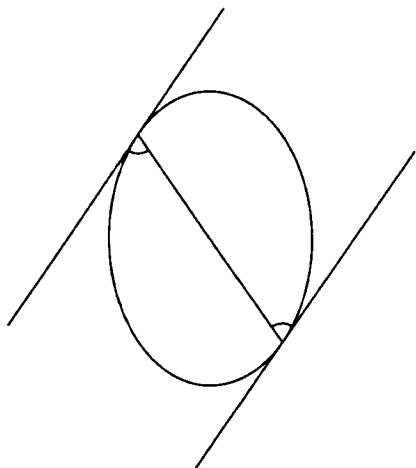


Рис. 6.4.3. Нахождение ширины

кривой функция H ограничена сверху отрицательным числом и, таким образом, достигает отрицательного максимума w . Параметризуем s и s' углом α , который образует соединяющая прямая с некоторым фиксированным направлением. Тогда из равенства $\theta = -\theta'$ следует, что

$$\frac{\partial H(s(\alpha), s'(\alpha))}{\partial \alpha} = \frac{\partial H}{\partial s} \frac{ds}{d\alpha} + \frac{\partial H}{\partial s'} \frac{ds'}{d\alpha} = \cos \theta \left(\frac{ds}{d\alpha} + \frac{ds'}{ds'} \right).$$

Если s — параметр длины дуги, тогда производная $\frac{ds}{d\alpha}$ равна кривизне границы B в точке, соответствующей точке s , а следовательно, она отлична от нуля (и конечна). Таким образом, производная $\frac{ds}{d\alpha}$ положительна. То же самое справедливо для $\frac{ds'}{d\alpha}$, и поэтому в критической точке функции $H(s(\alpha), s'(\alpha))$ мы имеем $\cos \theta = 0$. Следовательно, имеется прямая требуемого вида, для которой $\theta = \pi/2$, так что мы получили вторую точку периода два. $\square$

Мы можем построить аналог первой орбиты из предложения 6.4.4, имеющий период 3, если рассмотрим вписанный треугольник с самым большим периметром. Подобное построение работает и для орбит периода четыре. Для более высоких периодов уже имеются различные типы орбит, например, вписанные выпуклые пятиугольники и звездчатые пентаграммы.

Построение этих орбит в более общей ситуации сохраняющих площадь закручивающих отображений приведено в § 14.1. Имеются также аналогии и для второго типа орбит.

В случае эллипса (и, конечно же, круга) все орбиты периода выше чем два входят в непрерывные семейства, соответствующие их инвариантной окружности, но это свойство оказывается исключительным.

6.4.6. Уравнение зеркала из геометрической оптики. Теперь займемся поиском каустик в выпуклом билиарде. Напомним, что они являются огибающими семейств орбит, для которых отраженное семейство лучей имеет ту же самую огибающую. Мы сейчас определим их несколько более аккуратно, но совершенно ясно, что для изучения каустик важно понимать, как связаны огибающая совокупности лучей и огибающая отраженной совокупности. Другой подход состоит в том, чтобы для данной гладкой дуги на билиардном столе и совокупности касательных лучей рассмотреть все дуги, получающиеся отражением касательных лучей от границы билиардного стола. Какова их огибающая? Чтобы получить каустику, необходимо, конечно же, чтобы новая огибающая была другой частью той же самой кривой.

Для решения этой задачи мы привлечем элементарную дифференциальную геометрию. Для этого необходимо определить огибающую для данной параметризации семейства лучей. Чтобы параметризовать семейство лучей в плоскости, выберем кривую c , параметризованную переменной $s \in (-\epsilon, \epsilon)$, и совокупность $v(\cdot)$ единичных векторов, параметризованных той же переменной. Обозначим параметр вдоль луча через t и получим параметризацию семейства лучей $r(s, t) = c(s) + tv(s)$. Огибающая этого семейства — кривая, которая пересекает каждый луч только один раз (и касается лучей; см. рис. 6.4.4), так что она может быть параметризована как $r(s, f(s))$ для некоторой функции f . Свойство касания с этими лучами означает, что кривая

$$\frac{d}{ds} r(s, f(s)) = c'(s) + f'(s)v(s) + f(s)v'(s)$$

параллельна v , т. е. не имеет компоненты, нормальной к v . Мы можем выразить это условие, выбрав v' в качестве нормального вектора к v (это возможно, потому что v — векторы единичной длины). Таким образом, если $v' \neq 0$ (чтобы получить огибающую, мы предполагаем, что лучи не параллельны),

условие касания принимает вид

$$0 = \left\langle \frac{d}{ds} r(s, f(s)), v'(s) \right\rangle = \langle c'(s) + f'(s)v(s) + f(s)v'(s), v'(s) \rangle = \\ = \langle c'(s) + f(s)v'(s), v'(s) \rangle = \langle c'(s), v'(s) \rangle + f(s)\langle v'(s), v'(s) \rangle. \quad (6.4.5)$$

Отсюда f определяется однозначно:

$$f(s) = -\frac{\langle c'(s), v'(s) \rangle}{\langle v'(s), v'(s) \rangle}.$$

Заметим, в частности, что если кривая c вырождается в точку, тогда c является фокусом семейства лучей. И действительно, согласно приведенной выше формуле соотношение $f \equiv 0$ также параметризует фокус: $r(s, f(s)) = r(s, 0) = c(s)$. Чтобы связать огибающую с огибающей семейства отраженных от границы (точнее, от ее части) лучей, удобно параметризовать совокупность лучей этим способом и взять в качестве кривой c участок границы, который необходимо изучить. Преимущество состоит в том, что совокупность отраженных лучей может быть параметризована с использованием той же самой кривой c .

Таким образом, мы параметризуем часть границы бильярдного стола с помощью кривой c , используя параметр длины дуги s , т.е. такой параметр, что $T := c'$ является единичным вектором. Если мы выберем нормальный вектор N к c , направленный внутрь бильярдного стола, тогда кривизну κ кривой c можно определить соотношением $T'(s) = \kappa(s)N(s)$. Из этого, в частности, следует, что $N'(s) = -\kappa(s)T(s)$. Тогда мы получаем, что $\kappa \geq 0$ для выпуклого бильярда. Если $c''(s) \neq 0$, то $\kappa(s) > 0$. В частности, строго дифференцируемые выпуклые бильярдные столы имеют всюду ненулевую граничную кривизну.

Параметризуем семейство лучей с помощью равенства

$$r(s, t) = c(s) + tv(s),$$

где вектор v направлен внутрь и образует угол $\alpha(s) \in [0, \pi]$ с вектором $T(s)$. Тогда отраженное семейство может быть параметризовано равенством

$$\bar{r}(s, t) = c(s) + t\bar{v}(s),$$

где вектор $\bar{v}$ обращен внутрь и образует угол $\pi - \alpha(s)$ с $T(s)$. Результат отражения от границы на огибающей кратко описывается уравнением зеркала из геометрической оптики (см. рис. 6.4.5).

Теорема 6.4.5. Если f и $\bar{f}$ — огибающие этих двух семейств соответственно, то

$$\frac{1}{f} + \frac{1}{\bar{f}} = \frac{2\kappa}{\sin \alpha}.$$

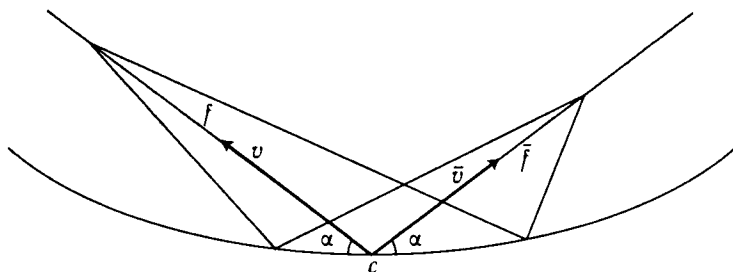


Рис. 6.4.5. Уравнение зеркала

Доказательство. Как и в формулировке, мы опускаем переменную s . Мы можем записать

$$v = \cos \alpha T + \sin \alpha N,$$

$$\bar{v} = -\cos \alpha T + \sin \alpha N$$

и, дифференцируя, получим

$$v' = -\sin \alpha \cdot \alpha' T + \cos \alpha T' + \cos \alpha \cdot \alpha' N + \sin \alpha N' = -(\alpha' + x) \sin \alpha T + (\alpha' + x) \cos \alpha N,$$

$$\bar{v}' = \sin \alpha \cdot \alpha' T - \cos \alpha T' + \cos \alpha \cdot \alpha' N + \sin \alpha N' = (\alpha' - x) \sin \alpha T + (\alpha' - x) \cos \alpha N.$$

Тогда

$$f = -\frac{\langle T, v' \rangle}{\langle v', v' \rangle} = -\frac{-(\alpha' + x) \sin \alpha}{(\alpha' + x)^2} = \frac{\sin \alpha}{\alpha' + x}$$

и аналогично $\bar{f} = \frac{\sin \alpha}{\alpha' - x}$, а следовательно, $\frac{1}{f} + \frac{1}{\bar{f}} = \frac{(\alpha' + x) - (\alpha' - x)}{\sin \alpha} = \frac{2x}{\sin \alpha}$. $\square$

Этот результат справедлив и в экстремальном случае, когда огибающая вырождается в одну точку, как, например, для круглого билиардного стола с совокупностью лучей, которые проходят через центр, при отражении переходят в себя и возвращаются к центру. В этом случае $f = \bar{f} = \rho$, где ρ — радиус, и $\sin \alpha = 1$, так что $x = 1/\rho$. Другой экстремальный случай — случай параллельного пучка прямых, сталкивающихся с границей круглого билиарда. В этом случае мы можем взять в уравнении зеркала $1/f = 0$ (выбирая точку p , в которой луч, проходящий через центр, сталкивается с границей) и получим, что $\bar{f} = \rho/2$, т. е. пучок фокусируется (примерно) посередине между центром и точкой p .

6.4.7. Каустики. Теперь мы можем изучить каустики, используя уравнение зеркала. Для круга и эллипса мы видели, что иногда каустики были связаны с инвариантными окружностями, а иногда — нет. Обобщим это разделение на более общий случай. Для этого определим каустики более строго, а также уточним понятие инвариантной окружности.

Определение 6.4.6. *Инвариантная окружность* Γ для билиардного отображения ϕ — это ϕ -инвариантное множество в C , которое является графиком непрерывной функции (отличной от 0 или π) из B в $[0, \pi]$. *Каустика* — это кусочно гладкая кривая γ , все касательные к которой являются частью орбиты билиарда, причем если луч орбиты билиарда определяет касательную к γ , то же верно и для его образа при отображении ϕ . Говорят, что каустика возникла из инвариантной окружности, если семейство определяющих ее лучей образует инвариантную окружность относительно ϕ в C . Каустика называется выпуклой, если она является выпуклой кривой.

Каустика не обязательно должна лежать внутри билиардного стола (гиперболы для эллиптического билиарда не лежат внутри стола), но это, конечно же, так для выпуклых каустик (иначе они имели бы касательные, которые не пересекали бы билиардный стол).

Выпуклая каустика возникает из инвариантной окружности, определенной касательными к ней, и альтернативный способ ее описания таков: пересечение всех левых полуплоскостей относительно этих лучей или пересечение правых полуплоскостей непусто, и граница этой области является каустикой.

Имеются и невыпуклые каустики, которые могут содержаться в билиардном столе. Пример показан на рис. 6.4.6. Этот пример получен возмущением кругового билиарда, и при его построении используется то обстоятельство, что центр кругового билиарда — вырожденная каустика. Оказывается, наличие выпуклых каустик накладывает ограничения на геометрию билиардного стола.

Теорема 6.4.7. *Выпуклый C^2 -билиард с точкой нулевой кривизны не имеет выпуклых каустик.*

Доказательство. Чтобы убедиться в этом, предположим, что на рис. 6.4.7 кривая γ — выпуклая каустика, и рассмотрим два таких касательных луча к ней с общей точкой $p \in B$, что один из них является образом другого при билиардном отображении. Каустика есть огибающая семейства лучей, определенных инвариантной окружностью, так же как и семейства их образов (в силу инвариантности). Таким образом, если f_p и $\bar{f}_p$ обозначают расстояния от точки p до точек касания, мы можем применить уравнение зеркала и получим, что

$$\frac{1}{f} + \frac{1}{\bar{f}} = \frac{2x}{\sin \alpha},$$

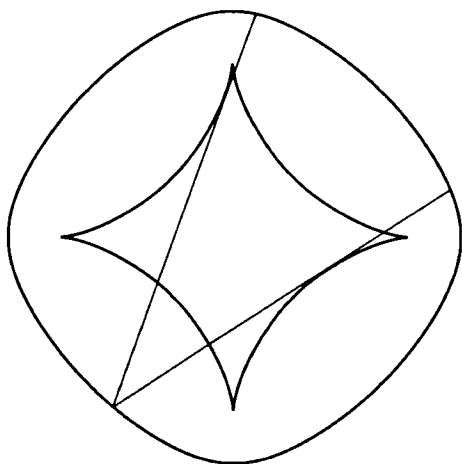


Рис. 6.4.6. Невыпуклая каустика

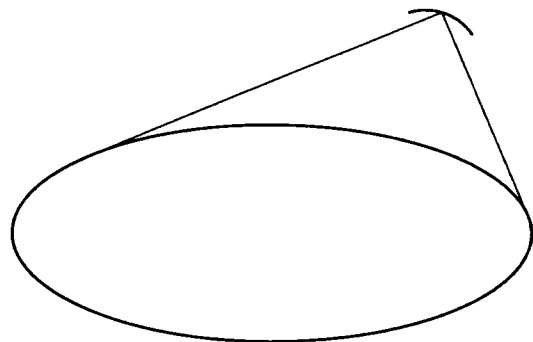


Рис. 6.4.7. Выпуклая каустика

где κ — кривизна границы B в точке p и α — угол, который образуют два луча с касательной к B в точке p . Левая часть этого последнего уравнения положительна, так что $\kappa \neq 0$. $\square$

Это показывает, что билиарды с граничной точкой нулевой кривизны (которые являются выпуклыми, но не гладко выпуклыми) совсем не похожи на интегрируемые системы, такие как эллипс, для которых фазовое пространство распадается в инвариантные окружности и динамика на этих окружностях является легкой для понимания. Такие билиарды должны быть динамически гораздо более сложными.

Весьма замечательно, что это единственный способ избежать выпуклых каустик. Строго дифференцируемый выпуклый билиард всегда имеет бесконечно много каустик, на самом деле их множество даже имеет ненулевую меру⁵.

6.4.8. Конструкция шнура. С другой стороны, те же самые соображения позволяют построить множество различных билиардов с выпуклыми каустиками. Мы можем заранее задать выпуклую кривую и строить семейство билиардов, которые имеют эту каустическую. Для этого обозначим длину каустики между точками касания (n на стороне вне точки p) на рис. 6.4.7 через ℓ_p . Тогда справедлив следующий результат.

Предложение 6.4.8. Величина $S(\gamma) := \int \kappa + \bar{\ell}_p + \ell_p$ не зависит от p .

⁵Это связано с КАМ-теорией. См.: В. Ф. Лазуткин. Существование континуума закрытых инвариантных кривых для выпуклого билиарда // УМН. 1972. Т. 27. С. 201—202.

Доказательство. Продифференцируем правую часть по параметру длины s на B , который параметризует точку p . Обозначая параметр длины на γ через t и его значения в точках касания через t_p и $\bar{t}_p$, мы получаем

$$\frac{d}{ds} \bar{t}_p = \cos \alpha - \frac{d}{ds} t_p, \quad \frac{d}{ds} \bar{t}_p = -\cos \alpha + \frac{d}{ds} \bar{t}_p, \quad \frac{d}{ds} \ell_p = \frac{d}{ds} t_p - \frac{d}{ds} \bar{t}_p.$$

Сумма этих производных, очевидно, равна нулю. $\square$

Число $L(\gamma) := S(\gamma) - l(\gamma)$ называется *параметром Лазуткина* для каустики γ .

Предыдущее предложение позволяет строить билиарды, имеющие данную выпуклую кривую γ в качестве каустики. Эта так называемая *конструкция шнура* состоит в следующем. Мы берем замкнутый шнурок некоторой длины $S > l(\gamma)$, натянутый вокруг кривой γ , и оттягиваем контур концом карандаша (на рис. 6.4.7 конец карандаша находится в верхней точке). Перемещая карандаш вдоль кривой γ таким образом, что контур остается натянутым, мы получаем билиардный стол, для которого γ является каустикой (причем $S(\gamma) = S$). При различных значениях S получаются различные билиарды с каустикой γ (и параметр Лазуткина измеряет, насколько длина шнура больше длины каустики). Хорошо известен частный случай этой конструкции — построение эллипса. В этом случае шнурок натягивается на отрезок прямой (с концами в фокусах эллипса). Конечно же, это не гладко выпуклая каустика. Используя шнуры разной длины, мы получаем софокусные эллипсы.

Эта конструкция шнура дает возможность найти множество билиардов с данной каустикой, но она не позволяет предъявить билиардный стол, для которого существует много каустик. На самом деле имеется давняя нерешенная задача, сформулированная Биркгофом: предположим, что билиардный стол имеет открытое множество каустик (а не только изолированные). Обязательно ли это эллиптический стол?

Упражнения

6.4.1. Покажите, что если выпуклый билиардный стол имеет две перпендикулярные оси симметрии, тогда билиардное отображение имеет орбиту периода 4.

6.4.2. Обобщите формулировку предыдущего упражнения для случая двух осей симметрии, образующих угол $2\pi/n$.

6.4.3. Для равностороннего треугольника, квадрата и правильного пятиугольника опишите билиардные столы, полученные с помощью конструкции шнура.

6.4.4. Найдите функционал, зависящий от нескольких переменных, критические точки которого порождают периодические орбиты билиарда.

6.4.5. Приведите пример выпуклого билиардного стола, отличного от круга, который имеет непрерывное семейство орбит периода 2.

Задачи для углубленного изучения

6.4.6. Приведите пример билиардного стола, отличного от круга, который имеет орбиту периода 2 по любому направлению.

6.4.7. Покажите, что семейство орбит из предыдущего упражнения определяет невыпуклую каустикой в качестве огибающей.

6.4.8. Используя рассуждения из п. 6.2.3, докажите, что билиардное отображение сохраняет площадь. Для этого рассмотрите *поток* билиарда через сечение, определенное границей. Поток через поверхность — это интеграл от нормальной компоненты скорости.

6.4.9. Постройте гладкую кривую, для которой, как на рис. 6.4.6, астроида является невыпуклой каустикой.

ПРОСТЫЕ СИСТЕМЫ СО СЛОЖНОЙ СТРУКТУРОЙ ОРБИТ

В этой главе рассматривается обширный набор свойств для целого ряда примеров. Наличие этих свойств вытекает из общей теории, о которой пойдет речь в гл. 10. Во всех изучаемых здесь примерах (за исключением квадратичного отображения f_4) мы рассматриваем гиперболические динамические системы (или символические динамические системы), и свойства, которые мы обсуждаем, являются общими для таких систем.

§ 7.1. Рост числа периодических точек

Периодические орбиты представляют собой наиболее характерный класс орбит. До настоящей главы мы главным образом рассматривали отображения с небольшим количеством периодических орбит или, как в случае рационального поворота, только лишь с периодическими орбитами. В этих простых примерах не встречались орбиты различных периодов для одного и того же отображения. Даже в наиболее сложных примерах, которые мы рассматривали, периодические орбиты возникали в качестве регулярных семейств для каждого периода, как, например, инвариантные кривые поворотов плоскости, линейных закручиваний, отображений для математического маятника при $t = 1$ и билиардов. В этих примерах мы уделяли больше времени согласованности, чем сложности. Теперь мы встретимся с первыми примерами, где периодические орбиты ведут себя совершенно иначе. Если в этих случаях имеются периодические точки с различными периодами, то мы хотим знать, сколько их.

Определение 7.1.1. Для отображения $f: X \rightarrow X$ обозначим через $P_n(f)$ число периодических точек f с (не обязательно минимальным) периодом n , т. е. число неподвижных точек для f^n .

В этом параграфе вводится много новых примеров динамических систем. Здесь они представлены с точки зрения структуры периодических орбит, но в дальнейшем выяснятся и многие другие замечательные особенности их структуры орбит.

7.1.1. Линейные растягивающие отображения. Рассмотрим необратимое отображение окружности E_2 , задаваемое в мультипликативной записи соотношением

$$E_2(z) = z^2, \quad |z| = 1,$$

а в аддитивной — соотношением

$$E_2(x) = 2x \pmod{1}. \quad (7.1.1)$$

Предложение 7.1.2. *Справедливо равенство $P_n(E_2) = 2^n - 1$, причем периодические точки для E_2 образуют плотное множество в S^1 .*

Доказательство. Если $E_2^n(z) = z$, то $z^{2^n} = z$, и $z^{2^n-1} = 1$. Таким образом, каждый корень из единицы степени $2^n - 1$ является периодической точкой для E_2 с периодом n . Имеется в точности $2^n - 1$ таких корней, и они равномерно распределены на окружности, т. е. интервалы между ними равны. В частности, при увеличении n эти интервалы уменьшаются (см. рис. 7.1.1). $\square$

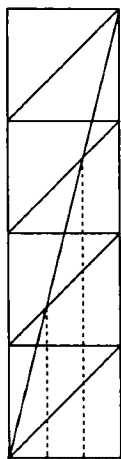


Рис. 7.1.1. Периодические точки для растягивающего отображения

Как следует из предложения 7.1.2, в качестве естественной меры асимптотического роста числа периодических точек можно взять экспоненциальную скорость роста $p(f)$ для последовательности $P_n(f)$:

$$p(f) = \overline{\lim}_{n \rightarrow \infty} \frac{\ln_+ P_n(f)}{n}, \quad (7.1.2)$$

где

$$\ln_+(x) = \begin{cases} \ln(x), & \text{если } x \geq 1, \\ 0 & \text{в противном случае.} \end{cases}$$

В частности, предложение 7.1.2 показывает, что

$$p(E_2) = \overline{\lim}_{n \rightarrow \infty} (\ln 2^n + \ln(1 - 2^{-n}))/n = \ln 2.$$

Отображения

$$E_m: x \mapsto mx \pmod{1},$$

где m — целое число, по модулю большее единицы, представляют собой непосредственное обобщение отображения E_2 . Неудивительно, что множества периодических орбит этих отображений также плотны. Доказательство дословно повторяет доказательство предложения 7.1.2, если заменить 2 на m .

Предложение 7.1.3. *Справедливо равенство $P_n(E_m) = |m^n - 1|$, причем периодические точки отображения E_m образуют плотное множество.*

Доказательство. Уравнение $z = E_m^n(z) = z^{m^n}$ имеет $|m^n - 1|$ решений, которые равномерно распределены. $\square$

См. также п. 7.1.3.

Другое свойство отображений E_m , которое стоит упомянуть, — это сохранение меры, определяемой длиной. Это свойство аналогично свойству сохранения фазового объема, о котором говорилось в § 6.1. Конечно же, длина образа любой дуги возрастает; однако, если рассмотреть *полный прообраз* дуги Δ при отображении E_m , видно, что он состоит из $|m|$ дуг, каждая длины $l(\Delta)/|m|$, расположенных на окружности на равных расстояниях друг от друга. Анализ, проведенный в п. 6.1.2, может быть расширен на необратимые отображения, сохраняющие объем, так что множество рекуррентных точек в этой ситуации также плотно.

7.1.2. Квадратичные и подобные им отображения. Для $\lambda \in \mathbb{R}$ определим функцию $f_\lambda: \mathbb{R} \rightarrow \mathbb{R}$, $f_\lambda(x) := \lambda x(1 - x)$. Для $0 \leq \lambda \leq 4$ функция f_λ отображает единичный отрезок $I = [0, 1]$ в себя. Семейство функций f_λ называется *квадратичным*. Для $\lambda \leq 3$ это семейство функций подробно рассмотрено в § 2.5, и асимптотическое поведение для любого такого λ довольно просто и изменяется в зависимости от λ лишь несколько раз. Как оказывается, для значений параметра из остающейся части отрезка в квадратичном семействе можно наблюдать изумительный набор весьма разнообразных сложных типов поведения, которые сменяются с калейдоскопической скоростью (см. рис. 7.1.2 и гл. 11). Заметим,



Рис. 7.1.2. Бифуркационная диаграмма

что $P_n(f_\lambda) \leq 2^n$, потому что n -я итерация отображения f_λ является многочленом степени 2^n , а следовательно, уравнение $(f_\lambda)^n(x) = x$ имеет максимум 2^n решений. В то время как следует ожидать, что в комплексной плоскости это уравнение действительно должно иметь в точности 2^n решений для большинства значений параметра λ , это, конечно, не так для действительных решений.

Здесь мы рассматриваем поведение квадратичного семейства для больших значений параметра, а именно $\lambda \geq 4$. В то время как для $\lambda > 4$ отрезок $[0, 1]$ не сохраняется, множество точек, которые остаются в этом отрезке, все же весьма интересно.

Анализируя поведение квадратичного семейства на единичном отрезке для $0 \leq \lambda \leq 3$ (см. § 2.5), мы получали только простое периодическое поведение: имеются лишь точки с периодом 1 или 2, и их число невелико. Этот анализ

не очень трудно расширить до $\lambda = 1 + \sqrt{6}$ (см. предложение 11.2.1). С другой стороны, справедлив следующий результат.

Предложение 7.1.4. Для $\lambda \geq 4$ выполнено равенство $P_n(f_\lambda) = 2^n$.

Доказательство. Поскольку $P_n(f_\lambda) \leq 2^n$, достаточно доказать обратное неравенство. Для этого мы используем следующее наблюдение: если отображение $f: \mathbb{R} \rightarrow \mathbb{R}$ непрерывно и $\Delta \subset [0, 1]$ — такой отрезок, что один его конец отображается в 0, а другой — в 1, тогда по теореме о промежуточном значении имеется неподвижная точка f в Δ .

Теперь отметим, что, поскольку $[0, 1] \subset [f_\lambda(0), f_\lambda(1/2)]$ и $[0, 1] \subset [f_\lambda(1/2), f_\lambda(1)]$, имеются отрезки $\Delta_0 \subset [0, 1/2]$ и $\Delta_1 \subset [1/2, 1]$, образы которых при отображении f_λ в точности совпадают с отрезком $[0, 1]$. Это позволяет нам получить две неподвижные точки для отображения f . Ненулевая неподвижная точка действительно находится внутри отрезка Δ_1 , потому что правый конец отрезка Δ_1 — это 1 и, следовательно, он отображается в 0, а левый конец отображается в 1, и поэтому ни один из них не может быть неподвижной точкой.

Кроме того, каждый из прообразов отрезков Δ_0 и Δ_1 при отображении f состоит из двух отрезков, так что имеется четыре отрезка, образы которых при отображении f^2 в точности совпадают с отрезком $[0, 1]$. Каждый из них содержит неподвижную точку отображения f_λ^2 , и снова каждая такая точка (кроме 0) находится внутри соответствующего отрезка, так что никакие две из этих неподвижных точек не совпадают.

Повторяя эти рассуждения последовательно, для дальнейших итераций отображения f_λ , мы получаем 2^n отрезков, образы которых при отображении f_λ^n совпадают с отрезком $[0, 1]$ и каждый из которых поэтому содержит по крайней мере одну неподвижную точку, что и доказывает существование 2^n различных орбит периода n для f_λ . $\square$

Эти рассуждения, показывающие, что $P_n(f_\lambda) \geq 2^n$, применяются на самом деле к любому такому непрерывному отображению $f: [0, 1] \rightarrow \mathbb{R}$, что $f(0) = f(1) = 0$ и $f(c) \geq 1$ для некоторого $c \in [0, 1]$. В этом более общем случае, однако, удобнее

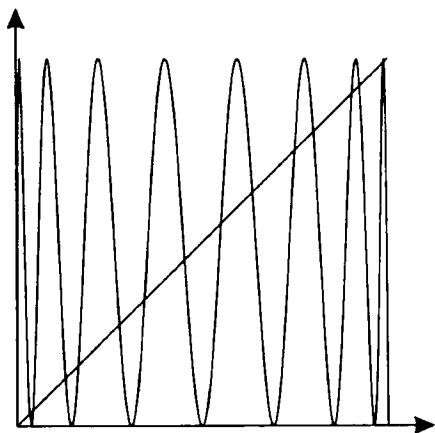


Рис. 7.1.3. Периодические точки отображения f_4

говорить об отрезках, образы которых при итерациях отображения f^n содержат отрезок $[0, 1]$, чем об отрезках, образы которых в точности равны $[0, 1]$.

В случае квадратичных отображений для $\lambda > 4$ можно несколько усовершенствовать предыдущее рассуждение и показать, что имеется в точности 2^n периодических точек (не используя того, что степень многочлена, задающего отображение f^n равна 2^n). Это также справедливо и для некоторых непрерывных отображений f более общего вида, которые являются монотонными на $[0, c]$, так же как и на $[c, 1]$. Непрерывное отображение, определенное на отрезке, возрастающее слева от внутренней точки и убывающее справа, называется *уни-модальным*. В итоге мы получаем следующий результат.

Предложение 7.1.5. *Если отображение $f: [0, 1] \rightarrow \mathbb{R}$ непрерывно, $f(0) = f(1) = 0$ и существует такое число $c \in [0, 1]$, что $f(c) > 1$, то $P_n(f) \geq 2^n$. Если, кроме того, отображение f унимодально и является растягивающим (т. е. $|f(x) - f(y)| > |x - y|$) на каждом отрезке из $f^{-1}((0, 1))$, тогда $P_n(f) = 2^n$.*

Доказательство основывается на следующей лемме.

Лемма 7.1.6. *Обозначим через $\mathcal{M}_k$ множество таких непрерывных отображений $f: [0, 1] \rightarrow \mathbb{R}$, что $f^{-1}((0, 1)) = \bigcup_{i=1}^k I_i$, где $I_i \subset [0, 1]$ — интервалы, отображение f монотонно на I_i и $f(I_i) = (0, 1)$. Тогда $f \circ g \in \mathcal{M}_k$, если $f \in \mathcal{M}_k$ и $g \in \mathcal{M}_l$.*

Доказательство. Если $f \in \mathcal{M}_k$ и $g \in \mathcal{M}_l$, то $f^{-1}((0, 1)) = \bigcup_{i=1}^k I_i$ и $g^{-1}(I_i) = \bigcup_{j=1}^l J_{ij}$, где множества $\{J_{ij}: 1 \leq i \leq k, 1 \leq j \leq l\}$ попарно не пересекаются, и $(f \circ g)^{-1}((0, 1)) = \bigcup_{ij} J_{ij}$. Композиция $f \circ g$ монотонна на J_{ij} , и $f \circ g(J_{ij}) = (0, 1)$. $\square$

Доказательство предложения 7.1.5. Лемма 7.1.6 показывает, что $P_n(f) \geq k^n$ для $f \in \mathcal{M}_k$, поскольку $f^n \in \mathcal{M}_{k^n}$. Если отображение f растягивающее на каждом отрезке в $f^{-1}((0, 1))$, тогда то же самое справедливо для итераций отображения f . Это показывает, что на каждом из этих отрезков имеется не более одного решения уравнения $f^n(x) = x$. Поэтому $P_n(f) \leq k^n$, что доказывает равенство. $\square$

7.1.3. Растягивающие отображения и степень. Теперь мы рассмотрим нелинейное обобщение растягивающих отображений E_m . Мы используем аддитивную форму записи для отображений окружности. В этих обозначениях производные отображений могут быть выражены как функции с действительными значениями.

Определение 7.1.7. Непрерывно дифференцируемое отображение $f: S^1 \rightarrow S^1$ называется *растягивающим* отображением, если $|f'(x)| > 1$ для всех $x \in S^1$.

Поскольку функция f' непрерывная и периодическая, $|f'|$ достигает минимума, который, следовательно, больше чем 1.

Из предложения 4.3.1 следует существование отображения $F: \mathbb{R} \rightarrow \mathbb{R}$, которое удовлетворяет условиям $[F(x)] = f([x])$ и $F(s+1) = F(s) + \deg(f)$, где $\deg(f)$ — степень отображения f . Это число имеет следующее простое свойство.

Лемма 7.1.8. *Если отображения $f, g: S^1 \rightarrow S^1$ непрерывны, то $\deg(g \circ f) = \deg(f) \deg(g)$, и, в частности, $\deg(f^n) = \deg(f)^n$.*

Доказательство. Если F, G — поднятия отображений f и g соответственно, то $G(s+k) = G(s+k-1) + \deg(g) = \dots = G(s) + k \deg(g)$ и $G(F(s+1)) = G(F(s) + \deg(f)) = \dots = G(F(s)) + \deg(g) \deg(f)$. $\square$

Это свойство полезно при подсчетах числа периодических точек.

Предложение 7.1.9. *Если $f: S^1 \rightarrow S^1$ — растягивающее отображение, тогда $|\deg(f)| > 1$ и $P_n(f) = |\deg(f)^n - 1|$.*

Доказательство. Из неравенства $|f'| > 1$ следует, что $|F'| > 1$ для любого поднятия, так что по теореме о конечных приращениях (теорема А.2.3) мы получаем $|\deg(f)| = |F(x+1) - F(x)| > 1$. Используя правило дифференцирования сложной функции, легко видеть, что любая итерация растягивающего отображения сама является растягивающим отображением, поэтому согласно лемме 7.1.8 достаточно рассмотреть случай $n = 1$. Возьмем поднятие F отображения f и рассмотрим его на отрезке $[0, 1]$. Неподвижные точки отображения f — проекции точек x , для которых $F(x) - x \in \mathbb{Z}$. Функция $g(x) := F(x) - x$ удовлетворяет условию $g(1) = g(0) + \deg(f) - 1$, так что по теореме о промежуточном значении найдется по крайней мере $|\deg(f) - 1|$ точек x , в которых $g(x) \in \mathbb{Z}$. Если $g(0) \in \mathbb{Z}$, тогда имеется $|\deg(f) - 1| + 1$ таких точек, но 0 и 1 проектируются в одну и ту же точку на S^1 . Далее, $g'(x) \neq 0$, так что функция g строго монотонна и, следовательно, принимает каждое значение не более одного раза. Таким образом, имеется в точности $|\deg(f) - 1|$ неподвижных точек на S^1 . $\square$

Рассуждение, доказывающее, что $P_n(f) \geq |\deg(f)^n - 1|$, работает для любого непрерывного отображения. Оно тривиально для отображений степени 1. Действительно, иррациональные повороты вообще не имеют ни неподвижных, ни периодических точек. Для отображений степени 0 то же рассуждение попросту гарантирует наличие неподвижной точки. Для таких отображений f , что $|\deg(f)| > 1$, из этого результата следует экспоненциальный рост числа периодических точек: $\rho(f) \geq \ln_+(|\deg(f)|)$.

7.1.4. Гиперболическое линейное отображение тора. Все предыдущие примеры были одномерными, но рост и распределение периодических точек, которые мы наблюдали в них, также появляются и в более высоких размерностях.

Это удобно продемонстрировать с помощью модели, которая строится из следующего линейного отображения в $\mathbb{R}^2$:

$$L(x, y) = (2x + y, x + y) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Если два вектора (x, y) и (x', y') представляют один и тот же элемент тора $\mathbb{T}^2$, т. е. если $(x - x', y - y') \in \mathbb{Z}^2$, то $L(x, y) - L(x', y') \in \mathbb{Z}^2$, так что $L(x, y)$ и $L(x', y')$

также представляют один и тот же элемент тора $\mathbb{T}^2$. Таким образом, L определяет отображение $F_L: \mathbb{T}^2 \rightarrow \mathbb{T}^2$:

$$F_L(x, y) = (2x + y, x + y) \pmod{1}.$$

Это отображение на самом деле является автоморфизмом тора, рассматриваемого как аддитивная группа. Отображение F_L обратимо, потому что определитель матрицы $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ равен 1, так что L^{-1} также имеет целочисленные элементы (на самом деле $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$), и, следовательно, по тем же соображениям обратная матрица определяет отображение $F_{L^{-1}} = F_L^{-1}$ на $\mathbb{T}^2$. Собственные значения отображения L равны

$$\lambda_1 = \frac{3 + \sqrt{5}}{2} > 1 \quad \text{и} \quad \lambda_1^{-1} = \lambda_2 = \frac{3 - \sqrt{5}}{2} < 1. \quad (7.1.3)$$

На рис. 7.1.4 показано, как F_L действует на фундаментальном квадрате $I = \{(x, y): 0 \leq x < 1, 0 \leq y \leq 1\}$. Прямые со стрелками — направления собственных векторов. Для любой матрицы L с определителем ± 1 отображение F_L сохраняет площади множеств на торе.

Рис. 7.1.4. Гиперболическое отображение тора

Предложение 7.1.10. *Периодические точки отображения F_L образуют плотное множество, и $P_n(F_L) = \lambda_1^n + \lambda_1^{-n} - 2$.*

Доказательство. Чтобы доказать свойство плотности, мы покажем, что точки с рациональными координатами являются периодическими. Пусть $x, y \in \mathbb{Q}$. Приводя дроби к общему знаменателю, запишем $x = s/q, y = t/q$, где $s, t, q \in \mathbb{Z}$. Тогда $F_L\left(\frac{s}{q}, \frac{t}{q}\right) = \left(\frac{2s+t}{q}, \frac{s+t}{q}\right)$ — рациональная точка, координаты которой

также имеют знаменатель q . Но имеется только q^2 различных точек на $\mathbb{T}^2$, координаты которых могут быть представлены как рациональные числа со знаменателем q , и все итерации $F_L^n(s/q, t/q)$, $n = 0, 1, 2, \dots$, принадлежат этому конечному множеству. Таким образом, они должны повторяться, т. е. $F_L^n(s/q, t/q) = F_L^m(s/q, t/q)$ для некоторых $n, m \in \mathbb{Z}$. Но поскольку отображение F_L обратимо, $F_L^{n-m}(s/q, t/q) = (s/q, t/q)$ и $(s/q, t/q)$ — периодическая точка, как и требуется. Из этого следует свойство плотности. Отметим, что, напротив, не все рациональные точки являются периодическими для E_m (см. упражнение 7.1.1).

Покажем теперь, что только точки с рациональными координатами являются периодическими точками для F_L . Запишем $F_L^n(x, y) = (ax + by, cx + dy) \pmod{1}$, где $a, b, c, d \in \mathbb{Z}$. Если $F_L^n(x, y) = (x, y)$, тогда

$$ax + by = x + k,$$

$$cx + dy = y + l$$

для $k, l \in \mathbb{Z}$. Поскольку 1 не является собственным значением для L^n , мы можем решить эту систему относительно (x, y) :

$$x = \frac{(d-1)k - bl}{(a-1)(d-1) - cb}, \quad y = \frac{(a-1)l - ck}{(a-1)(d-1) - cb}.$$

Таким образом, $x, y \in \mathbb{Q}$.

Теперь вычислим $P_n(F_L)$. Отображение

$$G = F_L^n - \text{Id}: (x, y) \mapsto ((a-1)x + by, cx + (d-1)y) \pmod{1}$$

является корректно заданным необратимым отображением тора на себя. Как и прежде, если $F_L^n(x, y) = (x, y)$, тогда $(a-1)x + by$ и $cx + (d-1)y$ — целые числа, а следовательно, $G(x, y) = 0 \pmod{1}$, т. е. неподвижные точки отображения F_L^n — это в точности прообразы точки $(0, 0)$ при отображении G . По модулю 1 это в точности точки из $\mathbb{Z}^2$ в $(L^n - \text{Id})([0, 1) \times [0, 1))$. Мы сейчас докажем, что их число определяется как площадь фигуры $(L^n - \text{Id})([0, 1) \times [0, 1))$, которая равна $|\det(L^n - \text{Id})| = |(\lambda_1^n - 1)(\lambda_1^{-n} - 1)| = \lambda_1^n + \lambda_1^{-n} - 2$. $\square$

Лемма 7.1.11. *Площадь параллелограмма с целочисленными вершинами равна числу точек решетки, которые он содержит, если считать две точки на ребрах как одну точку и все вершины как одну точку.*

Доказательство. Обозначим площадь параллелограмма через A . Если посчитать количество точек решетки, которые он содержит, указанным выше способом, мы получим целое число N , которое не меняется при любых параллельных переносах параллелограмма.

Теперь рассмотрим каноническое разбиение плоскости копиями этого параллелограмма, полученными при переносе на целые кратные ребер. Обозначим через l самую длинную диагональ. Площадь элементов этого разбиения может быть определена с помощью обратного рассуждения, а именно мы можем посчитать число элементов, находящихся в квадрате $[0, n) \times [0, n)$, для $n > 2l$. Те элементы, которые лежат внутри, покрывают меньший квадрат $[l, n-l) \times [l, n-l)$ полностью, так что есть по меньшей мере $\frac{(n-2l)^2}{A} \geq \frac{n^2}{A} \left(1 - \frac{4l}{n}\right)$ элементов. Так как

любой элемент, который пересекает квадрат, содержится в $[l, n-l) \times [l, n-l)$, имеется самое большее $\frac{(n+2l)^2}{A} = \frac{n^2}{A} \left(1 + \frac{4l}{n} \left(1 + \frac{l}{n}\right)\right) < \frac{n^2}{A} \left(1 + \frac{6l}{n}\right)$ элементов разбиения.

Число точек в квадрате с целыми координатами, равное n^2 , не меньше числа точек в элементах разбиения в квадрате и не больше числа точек в элементах, которые пересекаются с квадратом. Поэтому $N \cdot \frac{n^2}{A} \left(1 - \frac{4l}{n}\right) \leq n^2 \leq N \cdot \frac{n^2}{A} \left(1 + \frac{6l}{n}\right)$ и $1 - \frac{4l}{n} \leq \frac{A}{N} \leq 1 + \frac{6l}{n}$ для всех $n > 2l$. Это показывает, что $N = A$. $\square$

7.1.5. Обратные пределы. Самый близкий обратимый аналог отображения E_2 , который нам известен к настоящему моменту, — это автоморфизм тора, индуцированный отображением $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. Сейчас мы немного отклонимся от основной темы и опишем общую конструкцию, которая «делает отображение обратимым», т. е. позволяет получить обратимое отображение из необратимого некоторым стандартным способом. Способ преодолеть необратимость состоит в том, чтобы заменить точки данного пространства такими последовательностями $(x_n)_{n \in \mathbb{Z}}$, что $f(x_n) = x_{n+1}$. При этом исчезает неопределенность при взятии прообраза. Действительно, отображение $F((x_n)_{n \in \mathbb{Z}}) := (x_{n+1})_{n \in \mathbb{Z}}$, очевидным образом, обратимо.

Определение 7.1.12. Если X — метрическое пространство и отображение $f: X \rightarrow X$ непрерывно, то *обратный предел* определяется на пространстве

$$X' := \{(x_n)_{n \in \mathbb{Z}} : x_n \in X \text{ и } f(x_n) = x_{n+1} \text{ для всех } n \in \mathbb{Z}\}$$

соотношением $F((x_n)_{n \in \mathbb{Z}}) := (x_{n+1})_{n \in \mathbb{Z}}$.

Рассмотрим отображение $f = E_2$ на S^1 . Тогда обратный предел — это пространство

$$\mathbb{S} := \{(x_n)_{n \in \mathbb{Z}} : x_n \in X \text{ и } f(x_n) = x_{n+1} \text{ для всех } n \in \mathbb{Z}\}$$

с отображением $F((x_n)_{n \in \mathbb{Z}}) := (x_{n+1})_{n \in \mathbb{Z}} = (2x_n)_{n \in \mathbb{Z}} \pmod{1}$. Это пространство называется *соленоидом*.

Указать точку из $\mathbb{S}$ можно более экономно, не перечисляя всю последовательность. Если мы определим, например, элемент x_0 , все следующие элементы будут определены однозначно (орбитой элемента $x_0 \in S^1$ под действием E_2). Чтобы точно определить все предыдущие члены последовательности, необходимо только выбрать один из двух прообразов на каждом шаге. Для любого заданного элемента x_0 этот выбор можно закодировать односторонней последовательностью из нулей и единиц. Так как пространство Ω_2 таких последовательностей есть канторово множество (см. п. 7.3.5), соленоид $\mathbb{S}$ локально является произведением интервала (множества точек на S^1 вблизи x_0) и канторова множества.

Имеется красивый способ сделать конструкцию обратного предела наглядной. Начиная с окружности «начальных условий» x_0 число возможных прообразов x_{-1} увеличивается вдвое, так что окружность следует удвоить подобно резинке, обмотанной вокруг газеты. Но имеется вдвое больше прообразов второго

порядка и т. д., так что необходимо удваивать окружность до бесконечности. Это аналогично построению тринчного канторова множества, где мы сначала выбираем два интервала, затем четыре, и т. д.

Аккуратная геометрическая реализация этой конструкции изложена в § 13.2 и представлена на рис. 13.2.1. Этот рисунок демонстрирует богатство динамических идей и может служить эмблемой хаотической динамики. Наряду с подковой и линейными автоморфизмами тора, растягивающее отображение E_2 и солениод входят в число гиперболических динамических систем, лучше всего поддающихся анализу, и именно на основе их изучения были разработаны основные понятия и методы, позволяющие изучать и описывать любую хаотическую динамическую систему. Об этих методах будет идти речь в настоящей и следующей главах, а также в гл. 10.

Упражнения

7.1.1. Докажите, что для растягивающего отображения E_m ($|m| \geq 2$) рациональные точки являются прообразами периодических точек.

7.1.2. Найдите необходимое и достаточное условие для того, чтобы рациональная точка была периодической под действием E_m .

7.1.3. Докажите предложение 7.1.3 для $m < -1$.

7.1.4. Докажите, что для любого $n \in \mathbb{N}$ и любого $\lambda \geq 4$ квадратичное отображение f_λ имеет периодическую точку, *минимальный* период которой (см. определение 2.2.6) равен n .

7.1.5. Приведите пример непрерывного отображения $f: [0, 1] \rightarrow \mathbb{R}$, для которого существует такое $c \in [0, 1]$, что $f(c) > 1$, $f(0) = f(1) = 0$ и $P_n(f) > 2^n$.

7.1.6. Приведите пример такого гладкого унимодального отображения f , что $P_n(f) < 2^n$.

7.1.7. Покажите, что непрерывное отображение f на S^1 может быть деформировано к $E_{\deg(f)}$, т. е. что имеется такое непрерывное отображение $F: [0, 1] \times S^1 \rightarrow S^1$, что $F(0, \cdot) = E_{\deg(f)}$ и $F(1, \cdot) = f$.

7.1.8. Покажите, что отображения с различными степенями не могут быть деформированы друг в друга, т. е. не существует такого непрерывного отображения $F: [0, 1] \times S^1 \rightarrow S^1$, что $\deg(F(0, \cdot)) \neq \deg(F(1, \cdot))$.

7.1.9. Предположим, что отображение $f: S^1 \rightarrow S^1$ имеет степень 2 и 0 — притягивающая неподвижная точка. Покажите, что $P_n(f) > 2^n$.

7.1.10. Рассмотрим последовательность Фибоначчи (см. п. 1.2.2, пример 2.2.9 и п. 3.1.9). Покажите, что последовательность последних разрядов всех чисел Фибоначчи периодическая.

7.1.11. Примените конструкцию обратного предела к гомеоморфизму и докажите, что результат естественным образом эквивалентен первоначальной системе.

Задачи для углубленного изучения

7.1.12. Докажите, что соленоид из п. 7.1.5 является связным, но не линейно связным множеством.

§ 7.2. Топологическая транзитивность и хаос

Мы покажем, что некоторые из примеров, рассмотренных в предыдущем параграфе, топологически транзитивны в смысле определения 4.1.3, т. е. имеют плотные орбиты. Однако в то же самое время имеется бесконечно много периодических точек, что отличает эти примеры от иррациональных поворотов и других топологически транзитивных примеров из гл. 4 и 5. Для растягивающих отображений и гиперболических линейных отображений тора мы даже доказали, что объединение периодических точек плотно, что означает, что плотные и периодические орбиты неразрывно переплетены.

Таким образом, глобальная структура орбит в этих примерах гораздо сложнее. Переплетение плотных и периодических орбит — неотъемлемая особенность сложности структуры орбиты. Из этого следует неустойчивость любой орбиты по начальным условиям (см. определение 7.2.11 и теорему 7.2.12), что считается существенной особенностью *хаоса*.

Определение 7.2.1. Непрерывное отображение $f: X \rightarrow X$ метрического пространства называется *хаотическим*, если оно топологически транзитивно и его периодические точки плотны¹.

Повороты окружности показывают, что ни одно условие в отдельности не приводит к большой сложности.

Вскоре мы покажем, что растягивающие и гиперболические отображения являются хаотическими. На самом деле мы докажем даже более сильное свойство топологического перемешивания (см. определение 7.2.5), которое отсутствует в примерах из гл. 4 и 5. Перед тем как определить свойство перемешивания, мы приведем альтернативное определение топологической транзитивности.

7.2.1. Критерий топологической транзитивности. Мы определили топологическую транзитивность как существование плотной орбиты. Однако полезно иметь еще одну характеристику в терминах подмножеств фазового пространства. Чтобы включить сюда и необратимые отображения, мы скажем, что последовательность $(x_i)_{i \in \mathbb{Z}}$ является орбитой отображения f , если $f(x_i) = x_{i+1}$ для всех $i \in \mathbb{Z}$. Однако в дальнейшем мы будем просто писать $f^i(x)$ для $i \in \mathbb{Z}$, чтобы сохранить знакомые обозначения.

Предложение 7.2.2. Пусть X — полное сепарабельное метрическое пространство (т. е. в нем имеется счетное плотное подмножество) без изо-

¹ Универсально принятого определения хаоса нет, но это определение эквивалентно одному из наиболее распространенных, которое можно найти в литературе, и которое было предложено Робертом Девени (Devaney).

лированных точек. Если $f: X \rightarrow X$ — непрерывное отображение, тогда следующие четыре условия эквивалентны:

- 1) отображение f топологически транзитивно, т. е. имеет плотную орбиту;
- 2) отображение f имеет плотную положительную полуорбиту;
- 3) если U и V — непустые открытые подмножества пространства X , тогда существует такое $N \in \mathbb{Z}$, что $f^N(U) \cap V \neq \emptyset$;
- 4) если U и V — непустые открытые подмножества пространства X , тогда существует такое $N \in \mathbb{N}$, что $f^N(U) \cap V \neq \emptyset$.

Разумеется, импликации $(4) \Rightarrow (3)$ и $(2) \Rightarrow (1)$ очевидны. Оставшиеся импликации вытекают из следующего утверждения:

Лемма 7.2.3. Пусть X — метрическое пространство и $f: X \rightarrow X$ — непрерывное отображение. Тогда из условия (1) следует (3). Если пространство X не имеет изолированных точек, тогда из условия (1) следует (4). Если пространство X сепарабельно, тогда из условия (3) следует (1) и из условия (4) следует (2).

Доказательство. Пусть отображение f топологически транзитивно, и пусть орбита точки $x \in X$ плотна. Тогда существует такое $n \in \mathbb{Z}$, что $f^n(x) \in U$, и имеется такое $m \in \mathbb{Z}$, что $f^m(x) \in V$, а значит, $f^{m-n}(U) \cap V \neq \emptyset$. Из этого следует условие (3).

Если мы можем выбрать $m > n$, тогда, положив $N := m - n$, мы получим условие (4). В противном случае воспользуемся предположением, что пространство X не имеет изолированных точек. Тогда $f^m(x)$ — не изолированная точка, а поэтому имеется такое $n_k \in \mathbb{Z}$, что $|n_k| \rightarrow \infty$, $f^{n_k}(x) \in V$ и $f^{n_k}(x) \rightarrow f^m(x)$ при $k \rightarrow \infty$. Действительно, $n_k \rightarrow -\infty$, так как $n_k \leq n$ в соответствии с предположением (иначе мы возвращаемся к первому случаю), а следовательно, мы можем выбрать среди чисел n_k такое $m' < 2m - n$, что $f^{m'}(x) \in f^{m-n}(U)$. Тогда $x' := f^{n-m}(f^{m'}(x)) \in U$ и $f^{2m-n-m'}(x') = f^m(x) \in V$, а значит, $f^N(U) \cap V \neq \emptyset$, где $N := 2m - n - m' \in \mathbb{N}$. Таким образом, если пространство X не имеет изолированных точек, то $(1) \Rightarrow (4)$.

Теперь допустим, что имеет место сепарабельность и выполнено одно из условий пересечения (3) или (4). Одно и то же рассуждение позволит доказать, что из условия (3) следует (1) и что из условия (4) следует (2). Для счетного плотного подмножества $S \subset X$ обозначим через $U_1, U_2, \dots$ счетное множество шаров с центрами в точках из множества S с рациональным радиусом. Нам необходимо построить орбиту или полуорбиту, которая пересекается с каждым из множеств U_n . Согласно условию (3) существует такое $N_1 \in \mathbb{Z}$, что $f^{N_1}(U_1) \cap U_2 \neq \emptyset$. Если выполнено условие (4), мы можем взять $N_1 \in \mathbb{N}$. Пусть V_1 — такой открытый шар радиуса не больше $1/2$, что $\bar{V}_1 \subset U_1 \cap f^{-N_1}(U_2)$ (см. рис. 7.2.1). Существует такое $N_2 \in \mathbb{Z}$, что множество $f^{N_2}(V_1) \cap U_3$ является непустым, и, если условие (4) выполнено, мы можем взять $N_2 \in \mathbb{N}$. Теперь возьмем такой открытый шар V_2 радиуса не больше $1/4$, что $\bar{V}_2 \subset V_1 \cap f^{-N_2}(U_3)$. По индукции построим последовательность таких вложенных открытых шаров V_n радиуса не больше 2^{-n} ,

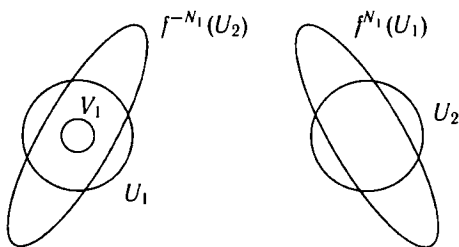


Рис. 7.2.1. Конструкция доказательства

что $\bar{V}_{n+1} \subset V_n \cap f^{-N_{n+1}}(U_{n+2})$. Центры этих шаров образуют последовательность Коши, предел которой x — единственная точка в пересечении $V = \bigcap_{n=1}^{\infty} \bar{V}_n = \bigcap_{n=1}^{\infty} V_n$.

Тогда $f^{N_{n-1}}(x) \in U_n$ для каждого $n \in \mathbb{N}$ и для всех $N_n \in \mathbb{N}$, если выполнено условие (4).

Если отображение f необратимо, на последнем шаге мы можем встретиться с выбором отрицательных значений N_n : возьмем такое i_k , что $N_{i_k} < 0$ для всех k и $N_{i_{k+1}} < N_{i_k}$. Выберем $x_0 = x$ и $x_{N_{i_k}} \in U_{i_{k+1}}$. Вместе с соотношением $f(x_k) = x_{k+1}$ это определяет орбиту точки x . $\square$

Следствие 7.2.4. *Непрерывное открытое (см. определение A.1.16) отображение f полного сепарабельного метрического пространства без изолированных точек топологически транзитивно тогда и только тогда, когда не существует двух непересекающихся открытых непустых f -инвариантных множеств.*

Доказательство. Если $U, V \subset X$ — открытые множества, тогда инвариантные множества $\bar{U} := \bigcup_{n \in \mathbb{Z}} f^n(U)$ и $\bar{V} := \bigcup_{n \in \mathbb{Z}} f^n(V)$ открыты, потому что f — открытое отображение, и поэтому имеют общие точки по предположению, так что $f^n(U) \cap f^m(V) \neq \emptyset$ для некоторых $n, m \in \mathbb{Z}$. Тогда $f^{n-m}(U) \cap V \neq \emptyset$ и отображение f топологически транзитивно в соответствии с предложением 7.2.2. Обратное утверждение очевидно: плотная орбита пересекается с каждым открытым множеством. $\square$

7.2.2. Топологическое перемешивание. Имеется свойство динамической системы, из которого немедленно следует топологическая транзитивность согласно только что доказанному критерию, но в действительности оно намного сильнее.

Определение 7.2.5. Непрерывное отображение $f: X \rightarrow X$ называется *топологически перемешивающим*, если для любых двух непустых открытых множеств $U, V \subset X$ имеется такое $N \in \mathbb{N}$, что $f^n(U) \cap V \neq \emptyset$ для каждого $n > N$.

Согласно предложению 7.2.2 всякое топологически перемешивающее отображение является топологически транзитивным. С другой стороны, отображения из наших простых примеров не являются перемешивающими. Никакой сдвиг T_γ , в частности никакой поворот окружности, не является топологически перемешивающим. Это следует из сохранения при параллельных переносах естественной

метрики на торе, индуцированной стандартной евклидовой метрикой в $\mathbb{R}^n$, и из следующего общего утверждения.

Лемма 7.2.6. *Изометрии не являются топологически перемешивающими.*

Доказательство. Пусть $f: X \rightarrow X$ — изометрия (т.е. отображение, которое сохраняет метрику на X). Возьмем различные точки $x, y, z \in X$ и положим $\delta := \min(d(x, y), d(y, z), d(z, x))/4$. Пусть U, V_1, V_2 — δ -шары с центрами в x, y, z соответственно. Поскольку f сохраняет диаметр любого множества, диаметр множества $f^n(U)$ не больше чем 2δ , в то время как расстояние между любыми точками $p \in V_1$ и $q \in V_2$ больше чем 2δ . Таким образом, для каждого n либо $f^n(U) \cap V_1 = \emptyset$, либо $f^n(U) \cap V_2 = \emptyset$. $\square$

7.2.3. Растягивающие отображения. Для растягивающих отображений мы получим свойство топологического перемешивания, доказав более сильное утверждение, согласно которому для любого открытого множества его образ при некоторой итерации отображения покрывает всю окружность. Для линейных растягивающих отображений E_m это очевидно: каждое открытое множество содержит отрезок вида $[l/|m|^k, (l+1)/|m|^k]$ для некоторых целых чисел k и $l \leq |m|^k$. Образ этого отрезка при отображении E_m^k есть S^1 .

Предложение 7.2.7. *Растягивающие отображения окружности S^1 топологически перемешивают.*

Доказательство. Пусть $f: S^1 \rightarrow S^1$ — такое отображение, что $|f'(x)| \geq \lambda > 1$ для всех x . Рассмотрим поднятие F отображения f на $\mathbb{R}$. Тогда $|F'(x)| \geq \lambda$ для $x \in \mathbb{R}$. Если $[a, b] \subset \mathbb{R}$ — отрезок, тогда по теореме о конечных приращениях А.2.3 существует такое $c \in (a, b)$, что $|F(b) - F(a)| = |F'(c)(b - a)| \geq \lambda(b - a)$, и поэтому длина любого интервала увеличивается на множитель λ^n (по меньшей мере) при действии F^n . Следовательно, для каждого отрезка I существует такое $n \in \mathbb{N}$, что длина отрезка $F(I)$ превышает 1. Поэтому образ проекции отрезка I на S^1 при действии отображения f^n покрывает S^1 . Так как каждое открытое множество из S^1 содержит отрезок, это доказывает, что каждое открытое множество имеет образ при действии какой-то итерации отображения f , который покрывает S^1 . $\square$

Следствие 7.2.8. *Линейные растягивающие отображения окружности S^1 являются хаотическими.*

Доказательство. Транзитивность следует из предложения 7.2.7, а плотность множества периодических точек — из предложения 7.1.3. $\square$

Для нелинейных растягивающих отображений этот результат также справедлив, что доказывается с помощью теоремы 7.4.3 (которая сформулирована ниже лишь для отображений степени два, но справедлива для любого растягивающего отображения).

7.2.4. Гиперболическое линейное отображение на торе. Гиперболическое линейное отображение F_L тора, индуцированное линейным отображением L с матрицей $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, было введено в п. 7.1.4. Собственные векторы, соответ-

ствующие первому собственному значению, принадлежат прямой $y = \frac{\sqrt{5}-1}{2}x$. Совокупность прямых, параллельных этой прямой, инвариантна относительно L ,

и L равномерно растягивает расстояния на этих прямых в λ_1 раз. Аналогично имеется инвариантное семейство сжимающихся прямых $y = \frac{-\sqrt{5}-1}{2}x + \text{const}$.

Предложение 7.2.9. Автоморфизм F_L является топологически перемешивающим.

Доказательство. Зафиксируем открытые множества $U, V \subset \mathbb{T}^2$ и отметим, что L -инвариантное семейство прямых

$$y = \frac{\sqrt{5}-1}{2}x + \text{const} \quad (7.2.1)$$

проектируется на $\mathbb{T}^2$ как F_L -инвариантное семейство орбит линейного потока T'_ω с иррациональным угловым коэффициентом $\omega = (1, (\sqrt{5}-1)/2)$. Согласно предложению 5.1.3 этот поток минимален. Таким образом, проекция каждой прямой всюду плотна на торе, следовательно, U содержит часть J растягивающейся прямой, и, кроме того, для любого $\epsilon > 0$ существуют число $T = T(\epsilon)$ и отрезок растягивающейся прямой длины T , который пересекается с любым ϵ -шаром на торе. Так как все отрезки данной длины есть параллельные переносы друг друга, это свойство справедливо для всех отрезков. Теперь возьмем такое ϵ , что V содержит ϵ -шар, и такое $N \in \mathbb{N}$, что $f^N(J)$ имеет длину по крайней мере T . Тогда $f^n(J) \cap V \neq \emptyset$ для $n \geq N$ и, таким образом, $f^n(U) \cap V \neq \emptyset$ для $n \geq N$. $\square$

Следствие 7.2.10. Автоморфизм F_L является хаотическим.

Доказательство. Для доказательства достаточно применить предложения 7.2.9 и 7.1.10. $\square$

7.2.5. Хаос. В начале этого параграфа мы мотивировали переход к изучению хаотических отображений тем, что для таких отображений характерна неустойчивость по начальным условиям. Дадим теперь формальное определение и проверим это утверждение.

Определение 7.2.11. Отображение $f: X \rightarrow X$ метрического пространства является *неустойчивым* по начальным условиям, если имеется такое число $\Delta > 0$, называемое *постоянной неустойчивости*, что для каждого $x \in X$ и для каждого $\epsilon > 0$ существует точка $y \in X$, удовлетворяющая неравенствам $d(x, y) < \epsilon$ и $d(f^N(x), f^N(y)) \geq \Delta$ для некоторого $N \in \mathbb{N}$.

Это означает, что малейшая погрешность (ϵ) в любом начальном условии (x) может привести к макроскопическому изменению (Δ) в эволюции динамики. Соответственно Δ сообщает нам, в каком масштабе эти погрешности появляются. Предположим, что мы начинаем изучение динамической системы в состоянии x , позволяем ей развиваться в течение некоторого времени, а затем пробуем по-

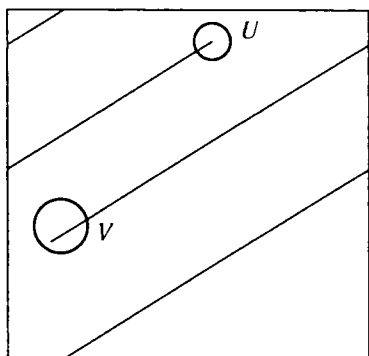


Рис. 7.2.2. Топологическое перемешивание

вторить тот же эксперимент. Даже если воспроизвести x с точностью до одной миллиардной дюйма, начальная крохотная погрешность может перерасти в большую разницу в поведении за конечное (и часто относительно короткое) время, т. е. вторая попытка провести тот же самый эксперимент приведет к совершенно другому результату по сравнению с первой. Об этом и говорил Пуанкаре в своем комментарии, приведенном в п. 1.1.1.

Для линейных растягивающих отображений это свойство вполне очевидно: любая начальная погрешность орбиты для E_m растет экспоненциально (с коэффициентом $|m|$ на каждой итерации), до тех пор пока не вырастет больше чем в $1/2|m|$ раз. В частности, $\delta = 1/2|m|$ — постоянная неустойчивости. С другой стороны, это свойство, очевидно, не выполняется для изометрий, потому что точки не удаляются друг от друга при итерациях.

Важной особенностью определения является то обстоятельство, что Δ не зависит ни от x , ни от ϵ , а только от рассматриваемой системы. Таким образом, *самая маленькая погрешность в любой момент* может привести к расхождению порядка Δ в конечном результате².

Теорема 7.2.12. *Хаотические отображения являются неустойчивыми по начальным условиям, за исключением случая, когда все пространство состоит из единственной периодической орбиты.*

Доказательство. Если все пространство не состоит из единственной периодической орбиты, то из плотности множества периодических точек следует наличие двух различных периодических орбит. Так как они не имеют общих точек, найдутся такие периодические точки p, q , что $\Delta := \min\{d(f^n(p), f^m(q)) : n, m \in \mathbb{Z}\}/8 > 0$. Покажем теперь, что Δ — постоянная неустойчивости.

Если $x \in X$, то орбита одной из этих двух точек, например точки q , находится на расстоянии по крайней мере 4Δ от x . Действительно, если бы обе орбиты находились на расстоянии меньше чем 4Δ от x , тогда расстояние между ними самими было бы меньше 8Δ .

Возьмем произвольное $\epsilon \in (0, \Delta)$. В силу плотности множества периодических точек имеется периодическая точка $p \in B(x, \epsilon)$, период которой мы обозначим через n . Тогда множество

$$V := \bigcap_{i=0}^n f^{-i}(B(f^i(q), \Delta))$$

точек, первые n итераций которых находятся от соответствующих итераций точки q на расстоянии не больше чем Δ , — открытая окрестность точки q . Согласно предложению 7.2.2 (точнее, той его части, которая не требует полноты) существует такое $k \in \mathbb{N}$, что $f^k(B(x, \epsilon)) \cap V \neq \emptyset$, т. е. существует такая точка $y \in B(x, \epsilon)$, что $f^k(y) \in V$. Если $j := \lfloor k/n \rfloor + 1$, тогда $k/n < j \leq (k/n) + 1$ и

$$k = n \cdot \frac{k}{n} < nj \leq n\left(\frac{k}{n} + 1\right) = k + n.$$

²Метеоролог Эдвард Лоренц описал это как «эффект бабочки»: похоже, что погода описывается хаотической динамической системой, так что взмах крыльев бабочки в Рио-де-Жанейро может через несколько дней вызвать тайфун в Токно.

Если мы возьмем $N := nj$, тогда это показывает, что $0 < N - k \leq n$. Поскольку $f^N(p) = p$, из неравенства треугольника следует, что

$$\begin{aligned} d(f^N(p), f^N(y)) &= d(p, f^N(y)) \geq \\ &\geq d(x, f^{N-k}(q)) - d(f^{N-k}(q), f^N(y)) - d(p, x) \geq \\ &\geq 4\Delta - \Delta - \Delta = 2\Delta, \end{aligned} \quad (7.2.2)$$

так как $p \in B(x, \epsilon) \subset B(x, \Delta)$ и

$$f^N(y) = f^{N-k}(f^k(y)) \in f^{N-k}(V) \subset B(f^{N-k}(q), \Delta)$$

по определению множества V . Обе точки p и y лежат в $B(x, \epsilon)$, и согласно неравенствам (7.2.2) либо $d(f^N(p), f^N(x)) \geq \Delta$, либо $d(f^N(y), f^N(x)) \geq \Delta$. $\square$

Замечание 7.2.13. Имеются отображения, неустойчивые по начальным условиям, которые не являются хаотическими, как, например, линейное закручивание из п. 6.1.1. При этом отображении для любой точки x существуют произвольно близко лежащие точки (на вертикальном отрезке, проходящем через x), которые перемещаются на значительное расстояние после достаточно большого числа итераций. Множество периодических точек состоит из точек, вторая координата которых рациональна, а следовательно, оно плотно. С другой стороны, ясно, что это отображение не топологически транзитивно.

Неустойчивость по начальным условиям следует из топологического перемешивания, без условия на периодические точки.

Предложение 7.2.14. *Топологически перемешивающее отображение (на пространстве с более чем одной точкой) является неустойчивым по начальным условиям.*

Доказательство. Возьмем такое $\Delta > 0$, что найдутся точки x_1, x_2 , удовлетворяющие неравенству $d(x_1, x_2) > 4\Delta$. Мы покажем, что выбранное значение Δ и есть постоянная неустойчивости.

Пусть $V_i = B_\Delta(x_i)$ для $i = 1, 2$. Предположим, что $x \in X$ и U — окрестность точки x . В силу топологического перемешивания найдутся такие числа $N_1, N_2 \in \mathbb{N}$, что $f^n(U) \cap V_1 \neq \emptyset$ для $n \geq N_1$ и $f^n(U) \cap V_2 \neq \emptyset$ для $n \geq N_2$. Для $n \geq N := \max(N_1, N_2)$ найдутся такие точки $y_1, y_2 \in U$, что $f^n(y_1) \in V_1$ и $f^n(y_2) \in V_2$, а следовательно, $d(f^n(y_1), f^n(y_2)) \geq 2\Delta$. Применяя неравенство треугольника, получаем $d(f^n(y_1), f^n(x)) \geq \Delta$ или $d(f^n(y_2), f^n(x)) \geq \Delta$. $\square$

Упражнения

7.2.1. Найдите максимальную постоянную неустойчивости для отображения E_2 .

7.2.2. Найдите точную верхнюю грань постоянных неустойчивости для отображения F_L из п. 7.2.4.

7.2.3. Докажите, что для топологически перемешивающего отображения любое число, меньшее чем $\text{диаметр} \sup\{d(x, y) : x, y \in X\}$ пространства X , является постоянной неустойчивости.

7.2.4. Рассмотрим линейное закручивание $T: S^1 \times [0, 1] \rightarrow S^1 \times [0, 1]$, $T(x, y) = (x + y, y)$, из п. 6.1.1, о котором также говорилось в замечании 7.2.13. Докажите, что оно имеет следующее свойство *частичного топологического перемешивания*: если $U, V \subset S^1$ — непустые открытые множества, то существует такое $N(U, V) \in \mathbb{N}$, что $T^n(U \times [0, 1]) \cap (V \times [0, 1]) \neq \emptyset$ для любого $n \geq N$.

7.2.5. Покажите, что для отображения компактного пространства неустойчивость по начальным условиям — топологический инвариант (см. п. 7.3.6).

7.2.6. Докажите, что для любых двух периодических точек отображения F_L множество гетероклинических точек (см. определение 2.3.4) плотно.

7.2.7. Рассмотрим (2×2) -матрицу целых чисел L без собственных значений, по модулю равных 1, и такую, что $|\det L| > 1$. Докажите, что индуцированное необратимое гиперболическое линейное отображение $F_L: \mathbb{T}^2 \rightarrow \mathbb{T}^2$ топологически перемешивает.

§ 7.3. Кодирование

Одна из наиболее важных идей, используемых при изучении сложной динамики, на первый взгляд довольно необычна. Она заключается в том, что мы отказываемся от некоторой информации и следим за поведением орбит лишь приблизительно. Точнее говоря, мы можем разбить фазовое пространство на конечное число частей и следить за орбитой, указывая только, в какой части она находится в данное время. Это напоминает замотанного туриста, который путешествует по Европе и думает, что раз сейчас вторник, то он, должно быть, в Бельгии. Более техническая аналогия получится, если проследить, как чей-либо мобильный телефон регистрируется в ячейках сотовой сети при перемещениях абонента.

В этих примерах мы действительно теряем информацию, потому что последовательность европейских стран или локальных сотовых станций не точно определяет местонахождение путешественника в произвольный заданный момент. Однако движение по орбитам динамической системы происходит не произвольно, и детерминированная природа динамики приводит к тому, что полный маршрут такого вида может (и часто именно так и есть) дать всю информацию относительно движения точки. Процесс получения такой информации называется *кодированием* динамической системы.

7.3.1. Линейные растягивающие отображения. Линейные растягивающие отображения

$$E_m: S^1 \rightarrow S^1, \quad E_m(x) = mx \pmod{1},$$

из п. 7.1.1 хаотичны (см. следствие 7.2.8), т. е. для них существуют как плотные орбиты (предложение 7.2.7), так и счетное плотное множество периодических орбит (предложение 7.1.3). Таким образом, структура орбит сложна и в то же время сильно неоднородна. Сейчас мы посмотрим на эти отображения с другой точки зрения, которая в свою очередь позволяет более глубоко оценить сложность структуры орбит. Для простоты обозначений предположим, как и прежде, что $m = 2$.

Рассмотрим двоичные отрезки

$$\Delta_n^k := \left[\frac{k}{2^n}, \frac{k+1}{2^n} \right] \quad \text{для } n = 1, \dots \text{ и } k = 0, 1, \dots, 2^n - 1$$

(для $n = 2$ они изображены на рис. 7.3.1). Пусть $x = 0, x_1 x_2 \dots$ — двоичное представление точки $x \in [0, 1]$. Тогда $2x = 0, x_1 x_2 x_3 \dots = 0, x_2 x_3 \dots \pmod{1}$. Таким образом,

$$E_2(x) = 0, x_2 x_3 \dots \pmod{1}. \quad (7.3.1)$$

Это первый и самый легкий пример *кодирования*, которое мы вскоре изучим более детально.

7.3.2. Следствия кодирования. Мы кратко упомянем несколько новых фактов относительно линейных растягивающих отображений, которые лучше всего проверить с использованием кодирования.

1. *Доказательство транзитивности с использованием кодирования.* Для начала получим с помощью кодирования другое доказательство топологической транзитивности. Для этого явно опишем двоичное представление числа, орбита которого при действии итераций отображения E_2 плотна. Рассмотрим целое число k , $0 \leq k \leq 2^n - 1$. Пусть $k_0 \dots k_{n-1}$ — двоичное представление числа k , возможно с несколькими нулями в начале.

Заметим, что $x \in \Delta_n^k$ тогда и только тогда, когда $x_i = k_i$ для $i = 0, \dots, n-1$. Поэтому начиная с этого момента мы будем писать $\Delta_{k_0 \dots k_{n-1}} := \Delta_n^k$. Теперь поместим двоичные представления всех чисел от 0 до $2^n - 1$ (с нулями в начале, если это необходимо) одно за другим и образуем конечную последовательность, которую мы обозначим через ω_n . Другими словами, последовательность ω_n получена объединением всех 2^n двоичных последовательностей длины n . Сделав это для каждого $n \in \mathbb{N}$, поместим последовательности ω_n , $n = 1, 2, \dots$, одну за другой, обозначим получившуюся бесконечную последовательность ω и рассмотрим число x с двоичным представлением $0, \omega$. Поскольку по построению сдвиг последовательности ω влево и удаление первых разрядов дает в различные моменты двоичные представления любого числа, содержащего n цифр, это означает, что орбита точки x при действии итераций отображения E_2 пересекает каждый отрезок $\Delta_{k_0 \dots k_{n-1}}$, а следовательно, плотна.

Эта конструкция обобщается для любого $m \geq 2$. Чтобы построить плотную орбиту для E_m , $m \leq -2$, мы заметим, что $E_m^2 = E_{m^2}$. Очевидно, что орбита любой точки при действии итераций квадрата отображения является подмножеством орбиты при действии итераций самого отображения. Таким образом, если первая орбита плотна, то плотна и вторая. Так что мы применяем нашу конструкцию к отображению E_{m^2} и получаем точку с плотной орбитой при действии E_m .

2. *Экзотическая асимптотика.* Теперь мы используем этот подход, чтобы доказать, что кроме периодических и плотных орбит имеются другие типы асимп-

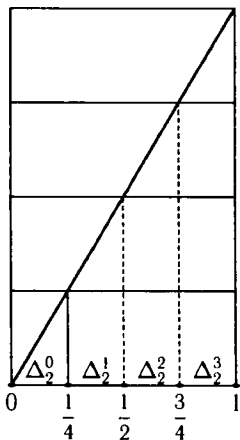


Рис. 7.3.1. Линейное кодирование

тотического поведения орбит растягивающих отображений. Такие орбиты можно построить и для E_2 , но самый простой и наиболее изящный пример получается для отображения E_3 .

Предложение 7.3.1. *Существует такая точка $x \in S^1$, что замыкание орбиты при действии отображения E_3 в аддитивной записи совпадает со стандартным канторовым множеством C . В частности, множество C является E_3 -инвариантным и содержит плотную орбиту.*

Доказательство. Канторово множество C может быть описано как множество всех точек на единичном отрезке, которые имеют представление с основанием 3 только с нулями и двойками в разрядах (см. п. 2.7.1). Подобно отображению, заданному соотношением (7.3.1), отображение E_3 действует как сдвиг разрядов влево в представлении с основанием 3. Из этого следует, что множество C является E_3 -инвариантным. Остается показать, что E_3 имеет плотную орбиту в C .

Каждая точка в C имеет единственное представление с основанием 3 без единиц. Пусть $x \in C$ и

$$0, x_1 x_2 x_3 \dots \quad (7.3.2)$$

— такое представление. Определим $h(x)$ как число, представление которого с основанием два есть $0, \frac{x_1}{2} \frac{x_2}{2} \frac{x_3}{2} \dots$, т.е. $h(x)$ получается из (7.3.2) заменой двоек на единицы. Таким образом, мы построили отображение $h: C \rightarrow [0, 1]$, которое является непрерывным, неубывающим (т.е. если $x > y$, то $h(x) \geq h(y)$) и взаимно однозначным, за исключением того, что двоичные рациональные числа имеют по два прообраза (см. п. 2.7.1 и 4.4.1). Кроме того, $h \circ E_3 = E_2 \circ h$. Предположим, что $D \subset [0, 1]$ — плотное множество точек, которое не содержит двоичных рациональных чисел. Тогда множество $h^{-1}(D)$ плотно в C , потому что если Δ — такой интервал, что $\Delta \cap C \neq \emptyset$, то $h(\Delta)$ является непустым интервалом, полуинтервалом или отрезком, а следовательно, содержит некоторые точки из D . Теперь возьмем любую точку $x \in [0, 1]$, E_2 -орбита которой плотна; тогда E_3 -орбита элемента $h^{-1}(x) \in C$ плотна в C . $\square$

3. Нерекуррентные точки. Другой интересный пример — конструкция нерекуррентной точки, т.е. такой точки x , что для некоторой ее окрестности U все итерации точки x не пересекаются с U (см. определение 6.1.8). Для отображения E_2 имеется плотное множество нерекуррентных точек.

Выберем произвольную фиксированную последовательность $(\omega_0, \dots, \omega_{n-1})$ из нулей и единиц и добавим в конце бесконечную последовательность, состоящую из нулей, если $\omega_{n-1} = 1$, или единиц, если $\omega_{n-1} = 0$. Обозначим получающуюся бесконечную последовательность через ω . Как и прежде, пусть x — число с двоичным представлением $0, \omega$. Таким образом, как мы и хотели показать, $x \in \Delta_{\omega_0 \dots \omega_{n-1}}$, и по построению $x \neq 0$. С другой стороны, $E_2^n x = 0$, а следовательно, $E_2^m x = 0$ для всех $m \geq n$, так что x — нерекуррентная точка.

Таким образом, мы установили, что отображение E_m хаотическое и топологически перемешивающее, что периодические и нерекуррентные орбиты плотны и что отображение E_3 имеет орбиты, замыкание которых есть канторово множество.

7.3.3. Двумерное канторово множество. Теперь мы опишем отображение в плоскости, естественным образом порождающее двумерное канторово множество (мы встречались с ним в упражнении 2.7.5), на котором троичное разложение координат позволяет получить всю информацию относительно динамики. Это *отображение подковы* играет центральную роль в некоторых из наших дальнейших построений.

Рассмотрим отображение, определенное на единичном квадрате $[0, 1] \times [0, 1]$, заданное следующим построением: сначала применим линейное преобразование $(x, y) \mapsto (3x, y/3)$ и получим горизонтальную полосу, левая и правая трети которой будут оставаться неизменными при следующем преобразовании. Удерживая левую треть неподвижно, изогнем и вытянем среднюю треть так, что правая треть попадает, не изменяясь, на верхнюю треть первоначального единичного квадрата. Это приводит к конфигурации, напоминающей латинскую букву «G». Для точек, которые находятся в квадрате и возвращаются в него, это отображение аналитически задается как

$$(x, y) \mapsto \begin{cases} (3x, y/3), & \text{если } x \leq 1/3, \\ (3x - 2, (y + 2)/3), & \text{если } x \geq 2/3. \end{cases}$$

Обратное отображение описывается соотношением

$$(x, y) \mapsto \begin{cases} (x/3, 3y), & \text{если } y \leq 1/3, \\ ((x + 2)/3, 3y - 2), & \text{если } y \geq 2/3. \end{cases}$$

Геометрически конфигурация обратного отображения по форме напоминает букву «e», повернутую против часовой стрелки на 90° .

Итерации этого отображения строятся следующим образом. Мы несколько раз умножаем координату x на 3, при этом все время полагая, что получающееся значение либо не больше $1/3$, либо не меньше $2/3$, иными словами, мы предполагаем, что первый троичный разряд равен 0 или 2, но не 1. (Если разложение не единственно, мы требуем, чтобы такой выбор был возможен.) Сравнивая эту конструкцию с построением троичного канторова множества (см. п. 2.7.1), мы видим, что координата x принадлежит троичному канторову множеству C . Рассматривая обратное отображение, мы аналогичным образом получаем, что для всех определяемых прообразов координаты y также лежат в канторовом множестве. Поэтому это отображение определено для всех положительных и отрицательных итераций на двумерном канторовом множестве $C \times C$. Мы можем непосредственно использовать троичное разложение для кодирования динамики. Для точки (x, y) при отображении троичное разложение координаты x сдвигается на один шаг влево, так что первый разряд опускается, а троичное разложение координаты y сдвигается на один разряд вправо. Теперь естественно избавиться от неопределенности относительно первого разряда смещенного разложения координаты y : мы просто пишем в соответствующей позиции разряд координаты x , который был только что опущен. Эта процедура сохраняет всю информацию, и ее лучше всего представить визуально, если записывать разложение координаты y в обратном порядке перед разложением координаты x . При этом мы получаем

бесконечную в обе стороны строку из нулей и двоек (необходимо помнить, что цифра 1 в разложении не допустима), и эта строка просто перемещается при действии отображения. Конечно, необходимо убедиться, что обратное отображение действует сдвигом в противоположном направлении.

7.3.4. Пространства последовательностей. Теперь мы можем перейти к рассмотрению идеи кодирования. Вообще говоря, под кодированием мы понимаем представление (не обязательно однозначное) точек в фазовом пространстве динамической системы с дискретным временем или в его инвариантном подмножестве с помощью последовательностей символов из некоторого «алфавита»; здесь мы будем использовать символы $0, \dots, N-1$. Так что мы должны познакомиться с этими пространствами.

Обозначим через Ω_N^R пространство последовательностей $\omega = (\omega_i)_{i=0}^\infty$, элементы которых — целые числа между 0 и $N-1$. Определим метрику

$$d_\lambda(\omega, \omega') := \sum_{i=0}^{\infty} \frac{\delta(\omega_i, \omega'_i)}{\lambda^i}, \quad (7.3.3)$$

где $\delta(k, l) = 1$, если $k \neq l$, $\delta(k, k) = 0$ и $\lambda > 2$. То же самое определение годится и для двусторонних последовательностей, если брать суммирование по $i \in \mathbb{Z}$:

$$d_\lambda(\omega, \omega') := \sum_{i \in \mathbb{Z}} \frac{\delta(\omega_i, \omega'_i)}{\lambda^{|i|}} \quad (7.3.4)$$

для некоторого $\lambda > 3$.

Это означает, что две последовательности близки, если они совпадают на длинном отрезке элементов вокруг начала координат.

Рассмотрим симметричный цилиндр, определенный равенством

$$C_{\alpha_1 \dots \alpha_{n-1}} := \{\omega \in \Omega_N : \omega_i = \alpha_i \text{ для } |i| < n\}.$$

Зафиксируем последовательность $\alpha \in C_{\alpha_1 \dots \alpha_{n-1}}$. Если $\omega \in C_{\alpha_1 \dots \alpha_{n-1}}$, то

$$d_\lambda(\alpha, \omega) = \sum_{i \in \mathbb{Z}} \frac{\delta(\alpha_i, \omega_i)}{\lambda^{|i|}} = \sum_{|i| \geq n} \frac{\delta(\alpha_i, \omega_i)}{\lambda^{|i|}} \leq \sum_{|i| \geq n} \frac{1}{\lambda^{|i|}} = \frac{1}{\lambda^{n-1}} \frac{2}{\lambda - 1} < \frac{1}{\lambda^{n-1}}.$$

Таким образом, $C_{\alpha_1 \dots \alpha_{n-1}} \subset B_{d_\lambda}(\alpha, \lambda^{1-n})$, где $B_{d_\lambda}(\alpha, \lambda^{1-n})$ — шар радиуса λ^{1-n} с центром α . Если $\omega \notin C_{\alpha_1 \dots \alpha_{n-1}}$, то

$$d_\lambda(\alpha, \omega) = \sum_{i \in \mathbb{Z}} \frac{\delta(\alpha_i, \omega_i)}{\lambda^{|i|}} \geq \lambda^{1-n},$$

потому что $\omega_i \neq \alpha_i$ для некоторого $|i| < n$. Таким образом, $\omega \notin B_{d_\lambda}(\alpha, \lambda^{1-n})$, и симметричный цилиндр — это шар радиуса λ^{1-n} с центром в любой из точек:

$$C_{\alpha_1 \dots \alpha_{n-1}} = B_{d_\lambda}(\alpha, \lambda^{1-n}). \quad (7.3.5)$$

Поэтому шары в Ω_N описываются указанием симметричной относительно нуля строки чисел.

Такие же рассуждения можно провести и для односторонних последовательностей (необходимо лишь потребовать, чтобы в (7.3.4) выполнялось условие $\lambda > 2$), и λ^{1-n} -шары описываются указанием строки первых n элементов.

При изучении наших примеров (см. (7.3.1)) мы видим, что можно представлять точки в фазовом пространстве последовательностями таким способом, чтобы последовательности, представляющие образ точки, получались из последовательностей, представляющих саму точку, сдвигом символов. Таким образом, данное преобразование соответствует *преобразованию сдвига*

$$\begin{aligned}\sigma: \Omega_N &\rightarrow \Omega_N, & (\sigma\omega)_i &= \omega_{i+1}, \\ \sigma^R: \Omega_N^R &\rightarrow \Omega_N^R, & (\sigma^R\omega)_i &= \omega_{i+1}.\end{aligned}\tag{7.3.6}$$

Сдвиг σ на Ω_N мы обычно будем обозначать σ_N , а сдвиг σ^R на Ω_N^R — σ_N^R . Для обратимых систем с дискретным временем любое кодирование включает последовательности символов, бесконечные в обе стороны, в то время как для необратимых систем используются лишь односторонние последовательности. В п. 7.3.7 эти сдвиги изучаются как динамические системы.

Среди преобразований сдвига, возникающих при кодировании, имеется также новый вид комбинаторной модели для динамической системы, которая описывается возможностью или, наоборот, невозможностью определенных последовательностей событий.

Определение 7.3.2. Пусть $A = (a_{ij})_{i,j=0}^{N-1}$ — матрица $N \times N$, элементы которой a_{ij} — нули или единицы. (Мы называем такую матрицу 0-1 матрицей.) Положим

$$\Omega_A := \{\omega \in \Omega_N : a_{\omega_n \omega_{n+1}} = 1 \text{ для } n \in \mathbb{Z}\}.\tag{7.3.7}$$

Пространство Ω_A замкнуто и инвариантно относительно сдвигов, и ограничение

$$\sigma_N|_{\Omega_A} =: \sigma_A$$

называется *топологической цепью Маркова*, определенной матрицей A .

Это частный случай *сдвига конечного типа*.

7.3.5. Кодирование. Последовательности, представляющие данную точку фазового пространства, называются *кодами* этой точки. У нас уже есть отдельные примеры кодирования: для отображения E_m на всей окружности мы можем осуществлять кодирование последовательностями над *алфавитом* $\{0, \dots, |m| - 1\}$, для ограничения отображения E_3 на троичное канторово множество C — односторонними последовательностями из нулей и единиц, для троичной подковы из п. 7.3.3 — последовательностями из нулей и двоек, бесконечными в двух направлениях. В тех двух случаях, когда мы используем односторонние последовательности, все получающиеся последовательности являются кодами некоторых точек и каждый код представляет только одну точку. Имеется, однако, важное отличие: в первом случае, когда для некоторого положительного числа m используется разложение с основанием m , точка может иметь один или два кода; в последнем же примере получается только один код.

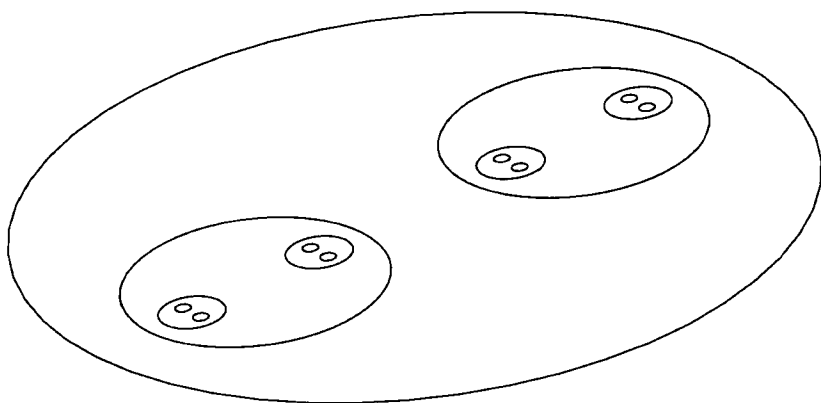


Рис. 7.3.2. Построение канторова множества

Это показывает, что пространство двоичных последовательностей есть канторово множество (см. определение 2.7.4). На самом деле это верно и для других пространств последовательностей.

7.3.6. Сопряжение и факторы. Грубо говоря, в рассмотренном выше случае сдвиг (Ω_2^R, σ^R) «содержит» отображение f с точностью до непрерывной замены координаты. (Мы уже сталкивались с подобной ситуацией в теореме 4.3.20.)

Определение 7.3.3. Предположим, что $g: X \rightarrow X$ и $f: Y \rightarrow Y$ — отображения метрических пространств X и Y и имеется такое непрерывное сюръективное отображение $h: X \rightarrow Y$, что $h \circ g = f \circ h$. Тогда f называется *фактором* отображения g посредством *полусопряжения* или *факторизующего* отображения h . Если h — гомеоморфизм, тогда отображения f и g называются *сопряженными*, а h — *сопряжением*.

Эти понятия кратко упоминались в п. 4.3.5, когда речь шла о моделировании произвольного гомеоморфизма окружности с помощью поворота. Понятие сопряжения является естественным и в то же время центральным: два сопряженных отображения могут быть получены друг из друга непрерывной заменой координат. Следовательно, все свойства и характеристики, которые не зависят от таких замен координат, например количество периодических орбит для каждого периода, неустойчивость по начальным условиям (см. упражнение 7.2.5), топологическая транзитивность, топологическое перемешивание, а следовательно, и свойство хаотичности, остаются неизменными. Такие свойства и характеристики называются топологическими инвариантами. Далее в этой книге мы еще столкнемся с другими важными топологическими инвариантами, например с топологической энтропией (см. определение 8.2.1).

7.3.7. Динамика сдвигов и топологических цепей Маркова. Изучим теперь более подробно свойства сдвигов и топологических цепей Маркова (см. соотношения (7.3.6) и определение 7.3.2). Эти понятия важны, поскольку много интересных динамических систем кодируются сдвигами или топологическими цепями Маркова. Результаты, полученные в этом пункте, могут быть немедленно применены к подобным динамическим системам.

Предложение 7.3.4. *Периодические точки сдвигов σ_N и σ_N^R образуют плотные множества в пространствах Ω_N и Ω_N^R соответственно; $P_n(\sigma_N) = P_n(\sigma_N^R) = N^n$, и оба сдвига σ_N и σ_N^R являются топологически перемешивающими.*

Доказательство. Периодические орбиты для сдвига — это периодические последовательности, т.е. $(\sigma_N)^m \omega = \omega$ тогда и только тогда, когда $\omega_{n+m} = \omega_n$ для всех $n \in \mathbb{Z}$. Для доказательства плотности множеств периодических точек достаточно найти периодическую точку в каждом шаре (симметрическом цилиндре), потому что каждое открытое множество содержит шар. Чтобы найти периодическую точку в $C_{\alpha_{-m}, \dots, \alpha_m}$, возьмем последовательность ω , определенную соотношением $\omega_n = \alpha_{n'}$ для $|n'| \leq m$, $n' \equiv n \pmod{2m+1}$. Она лежит в этом цилиндре и имеет период $2m+1$.

Каждая периодическая последовательность ω с периодом l однозначно определена своими координатами $\omega_0, \dots, \omega_{l-1}$. Имеется N^l различных конечных последовательностей вида $(\omega_0, \dots, \omega_{l-1})$.

Чтобы проверить свойство топологического перемешивания, мы покажем, что $\sigma_N^n(C_{\alpha_{-m}, \dots, \alpha_m}) \cap C_{\beta_{-m}, \dots, \beta_m} \neq \emptyset$ для $n > 2m+1$, скажем, $n = 2m+k+1$, где $k > 0$. Рассмотрим такую последовательность ω , что

$$\omega_i = \alpha_i \text{ для } |i| \leq m, \quad \omega_i = \beta_{i-n} \text{ для } i = m+k+1, \dots, 3m+k+1.$$

Тогда $\omega \in C_{\alpha_{-m}, \dots, \alpha_m}$ и $\sigma_N^n(\omega) \in C_{\beta_{-m}, \dots, \beta_m}$.

Рассуждения для одностороннего сдвига аналогичны. □

Для топологических цепей Маркова имеется полезное геометрическое представление. Соединив i с j стрелкой, если $a_{ij} = 1$, мы получим *марковский граф* G_A с N вершинами и несколькими ориентированными ребрами. Мы скажем, что конечная или бесконечная последовательность вершин графа G_A является *допустимым путем* или *допустимой последовательностью*, если любые две последовательные вершины из этой последовательности соединены ориентированной соответствующим образом стрелкой. Точка из Ω_A соответствует бесконечному (в обе стороны) пути в G_A с отмеченным началом; топологическая цепь Маркова σ_A соответствует перемещению начала в следующую вершину. Следующая простая комбинаторная лемма является ключевой для изучения топологических цепей Маркова.

Лемма 7.3.5. *Для каждой пары $i, j \in \{0, 1, \dots, N-1\}$ число N_{ij}^m допустимых путей длины $m+1$, которые начинаются в точке x_i и кончатся в точке x_j , равно элементу a_{ij}^m матрицы A^m .*

Доказательство. Используем индукцию на m . Из определения графа G_A следует, что $N_{ij}^1 = a_{ij}$. Покажем, что

$$N_{ij}^{m+1} = \sum_{k=0}^{N-1} N_{ik}^m a_{kj}. \quad (7.3.8)$$

Возьмем произвольную точку $k \in \{0, \dots, N-1\}$ и допустимый путь длины $m+1$, соединяющий i и k . Он может быть расширен до допустимого пути длины $m+2$, соединяющего i и j (добавлением точки j) тогда и только тогда, когда $a_{kj} = 1$. Это доказывает соотношение (7.3.8). Теперь, если по индукции предположить, что $N_{ij}^m = a_{ij}^m$ для всех ij , мы получаем $N_{ij}^{m+1} = a_{ij}^{m+1}$ в соответствии с (7.3.8). □

Следствие 7.3.6. *Справедливо соотношение $P_A(\sigma_A) = \text{tr } A^n$.*

Доказательство. Каждому допустимому замкнутому пути длины $m+1$ с отмеченным началом, т.е. пути, который начинается и заканчивается в одной и той же вершине графа G_A , соответствует ровно одна периодическая точка цепи σ_A периода m . □

Поскольку собственное значение с максимальным модулем асимптотически определяет значение следа матрицы, оно определяет и скорость экспоненциального роста.

Предложение 7.3.7. *Справедливо соотношение $p(\sigma_A) = r(A)$, где $r(A)$ — спектральный радиус.*

Доказательство. Очевидно, что $p(\sigma_A) \leq r(A)$. Чтобы доказать обратное неравенство, нам необходимо избежать сокращений: если $\lambda_j = re^{2\pi i \phi_j}$ ($1 \leq j \leq k$) — максимальные по модулю собственные значения, то существует такая последовательность $m_n \rightarrow \infty$, что $m_n \phi_j \rightarrow 0 \pmod{1}$ для всех j (о рекуррентности для сдвигов тора см. § 5.1), следовательно, $\sum \lambda_i^{m_n} \sim r^{m_n}$. □

Пример 7.3.8. Марковский граф, изображенный на рис. 7.3.3, имеет три неподвижные точки $\bar{0}$, $\bar{1}$ и $\bar{4}$. Путям $\bar{01}$ и $\bar{23}$ соответствуют четыре периодические точки с периодом 2. Орбиты периода 3 порождены путями $\bar{011}$, $\bar{001}$, $\bar{234}$.

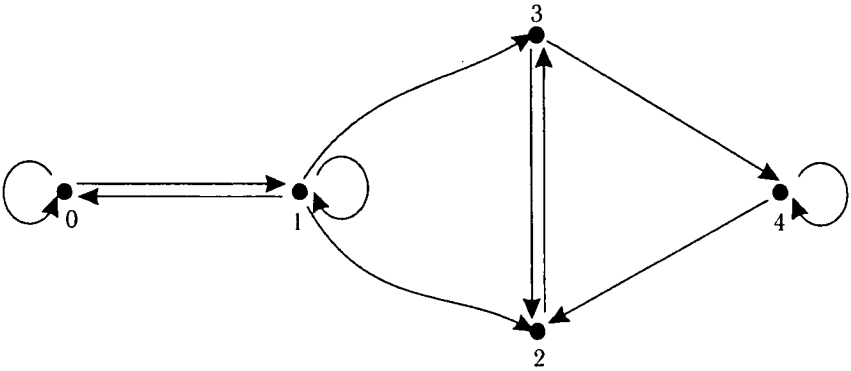


Рис. 7.3.3. Марковский граф

Топологические цепи Маркова можно классифицировать согласно свойствам рекуррентности различных орбит, которые они содержат. Мы будем преимущественно рассматривать те топологические цепи Маркова, которые обладают самыми сильными рекуррентными свойствами.

Определение 7.3.9. Матрица A называется *положительной*, если все ее элементы положительны. Матрица A из нулей и единиц называется *транзитивной*, если матрица A^m положительна для некоторого $m \in \mathbb{N}$. Топологическая цепь Маркова σ_A называется *транзитивной*, если A — транзитивная матрица.

Лемма 7.3.10. Если матрица A^m положительна, тогда матрица A^n также положительна для любого $n \geq m$.

Доказательство. Если $a_{ij}^n > 0$ для всех i, j , тогда для каждого j найдется такое k , что $a_{kj} = 1$. В противном случае $a_{ij}^n = 0$ для любых значений n и i . Далее применим индукцию. Если $a_{ij}^n > 0$ для всех i, j , тогда $a_{ij}^{n+1} = \sum_{k=0}^{N-1} a_{ik}^n a_{kj} > 0$, так как $a_{kj} = 1$ по крайней мере для одного k . $\square$

Лемма 7.3.11. Если матрица A транзитивна и путь $\alpha_{-k}, \dots, \alpha_k$ допустим, т.е. $a_{\alpha_i \alpha_{i+1}} = 1$ для $i = -k, \dots, k-1$, то пересечение $\Omega_A \cap C_{\alpha_{-k}, \dots, \alpha_k, A} =: C_{\alpha_{-k}, \dots, \alpha_k, A}$ непусто и, более того, содержит периодическую точку.

Доказательство. Возьмем такое m , что $a_{\alpha_k \alpha_{-k}}^m > 0$. Тогда можно продолжить последовательность α до допустимой последовательности длины $2k + m + 1$, которая начинается и заканчивается в точке α_{-k} . Повторяя эту последовательность периодически, мы получаем периодическую точку в $C_{\alpha_{-k}, \dots, \alpha_k, A}$. $\square$

Предложение 7.3.12. Если A — транзитивная матрица, тогда топологическая цепь Маркова σ_A является топологическим перемешиванием и ее периодические орбиты плотны в Ω_A ; в частности, преобразование σ_A хаотично, а следовательно, неустойчиво по начальным условиям.

Доказательство. Плотность множества периодических орбит следует из леммы 7.3.11. Для доказательства топологического перемешивания выберем открытые множества $U, V \subset \Omega_A$ и непустые симметричные цилиндры $C_{\alpha_{-k}, \dots, \alpha_k, A} \subset U$ и $C_{\beta_{-k}, \dots, \beta_k, A} \subset V$. Тогда достаточно показать, что $\sigma_A^n(C_{\alpha_{-k}, \dots, \alpha_k, A}) \cap C_{\beta_{-k}, \dots, \beta_k, A} \neq \emptyset$ для любого достаточно большого n . Возьмем $n = 2k + 1 + m + 1$,

где $l \geq 0$, а число m выбрано в соответствии с определением 7.3.9. Тогда $a_{\alpha_k \beta_{-k}}^{m+l} > 0$ по лемме 7.3.10, так что существует допустимая последовательность длины $4k + 2 + m + l$, первые $2k + 1$ символов в которой идентичны символам $\alpha_{-k}, \dots, \alpha_k$, а последние $2k + 1$ символов идентичны символам $\beta_{-k}, \dots, \beta_k$. По лемме 7.3.11 эта последовательность может быть продолжена до периодического элемента в Ω_A , который принадлежит множеству $\sigma_A^n(C_{\alpha_{-k}, \dots, \alpha_k, A}) \cap C_{\beta_{-k}, \dots, \beta_k, A}$. $\square$

Пример 7.3.13. Матрица $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ не транзитивна, потому что все степени этой матрицы остаются верхнетреугольными, а следовательно, не существует пути от 0 до 1. Пространство Ω_A на самом деле счетно и состоит из двух неподвижных точек $(\dots, 0, \dots, 0, \dots)$ и $(\dots, 1, \dots, 1, \dots)$ и единственной гетероклинической орбиты, соединяющей их (эта орбита состоит из последовательностей, у которых все элементы до некоторого места единицы, а после него нули).

Пример 7.3.14. Для матрицы

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

каждая орбита попеременно попадает то в группу элементов, принимающих значения из множества $\{0, 1\}$, то в группу элементов, принимающих значения из множества $\{2, 3\}$, т. е. четность и нечетность чередуются. Поэтому никакая степень этой матрицы не может иметь только положительные элементы.

Упражнения

7.3.1. Докажите, что отображение E_2 имеет непериодическую орбиту, все четные итерации которой лежат в левой половине единичного отрезка.

7.3.2. Докажите, что отображение E_2 имеет несчетное число орбит, для которых ни один отрезок длины 10 не имеет больше одной точки в левой половине единичного отрезка.

7.3.3. Докажите, что линейные отображения, которые сопряжены в смысле линейной алгебры, являются топологически сопряженными в смысле определения 7.3.3.

7.3.4. Выпишите матрицу Маркова для рис. 7.3.3 и проверьте справедливость следствия 7.3.6 вплоть до периода 3.

7.3.5. Рассмотрим метрику

$$d'_\lambda(\alpha, \omega) := \sum_{i \in \mathbb{Z}} \frac{|\alpha_i - \omega_i|}{\lambda^{|i|}} \quad (7.3.9)$$

на Ω_N . Покажите, что если $\lambda > 2N - 1$, то цилиндр $C_{\alpha_1 \dots \alpha_{n-1}}$ является λ^{1-n} -ша-ром для d'_λ .

7.3.6. Прodelайте предыдущее упражнение для односторонних сдвигов (при $\lambda > N$).

7.3.7. Рассмотрим метрику

$$d''_\lambda(\alpha, \omega) := \lambda^{-\max\{n \in \mathbb{N} : \alpha_i = \omega_i, \text{ для } |i| \leq n\}} \quad (7.3.10)$$

($d''_\lambda(\alpha, \alpha) = 0$) на Ω_N . Докажите, что цилиндр $C_{\alpha_1 \dots \alpha_{n-1}}$ является шаром для d''_λ .

7.3.8. Найдите точную верхнюю грань постоянных неустойчивости для транзитивной топологической цепи Маркова с метрикой d''_λ .

7.3.9. Найдите точную верхнюю грань постоянных неустойчивости для транзитивной топологической цепи Маркова с метрикой d'_λ .

7.3.10. Покажите, что для $m < n$ сдвиг на Ω_m является фактором на Ω_n .

7.3.11. Докажите, что квадратичное отображение f_4 на $[0, 1]$ не сопряжено ни с одним из отображений f_λ для $\lambda \in [0, 4)$.

7.3.12. Покажите, что топологические цепи Маркова, определенные матрицами

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix},$$

сопряжены.

7.3.13. Найдите наименьшее из положительных значений $\rho(\sigma_A)$ для транзитивной топологической цепи Маркова с двумя состояниями (т. е. с (2×2) -матрицей A).

Задачи для углубленного изучения

7.3.14. Найдите все факторы иррационального поворота окружности.

7.3.15. Найдите наименьшее значение $\rho(\sigma_A)$ для транзитивной топологической цепи Маркова с тремя состояниями (т. е. с (3×3) -матрицей A).

§ 7.4. Дальнейшие примеры кодирования

Теперь мы построим примеры кодирования для некоторых уже известных динамических систем.

7.4.1. Нелинейные растягивающие отображения. Имеется соответствие между общими (не обязательно линейными) растягивающими отображениями окружности (см. п. 7.1.3) и сдвигами на пространстве последовательностей. В этом можно убедиться, проделав почти такое же построение, как в п. 7.3.1. Для этого потребуются некоторые усилия, но они будут вознаграждены, так как мы получим полную классификацию широкого класса отображений с помощью простого инварианта.

Для простоты обозначений мы рассмотрим растягивающее отображение $f: S^1 \rightarrow S^1$ степени 2. Согласно предложению 7.1.9 отображение f имеет только одну неподвижную точку p . (Для отображений с более высокой степенью мы бы взяли любую из неподвижных точек.) Поскольку $\deg(f) = 2$, имеется в точности одна такая точка $q \neq p$, что $f(q) = p$. Точки p и q делят окружность на две дуги. Начиная с p и следуя в положительном направлении, обозначим первую дугу Δ_0 , а вторую — Δ_1 . Определим кодирование для $x \in S^1$ следующим образом: точка x представляется последовательностью $\omega \in \Omega_2^R$, для которой

$$f^n(x) \in \Delta_{\omega_n}. \quad (7.4.1)$$

Это представление единственно, если только не окажется, что при некотором n выполняется условие $f^n(x) \in \{p, q\} = \Delta_0 \cap \Delta_1$. Отсутствие единственности вполне подобно случаю двоичных рациональных дробей для отображения E_2 . Предположим, что точка x имеет итерацию из множества $\{p, q\}$. Тогда либо $x = p$

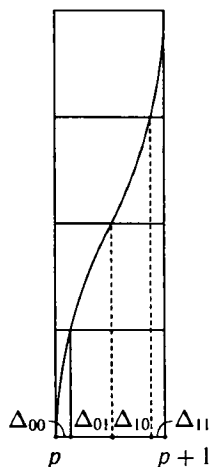


Рис. 7.4.1.
Нелинейное
кодирование

и $f^n(x) = p$ для всех $n \in \mathbb{N}$, либо точка q должна появиться в последовательности итераций раньше, чем p , т. е. $f^n(x) \notin \{p, q\}$ для всех n , меньших некоторого значения k , и тогда $f^k(x) = q$ и $f^{k+1}(x) = p$. Будем считать, что точка p имеет два кода: все нули и все единицы, точка q имеет два кода: $0111111 \dots$ и $1000000 \dots$, и любая такая точка x , что $F^k(x) = q$, имеет два кода, в которых первые $k - 1$ разрядов однозначно определяются из условия (7.4.1), а далее следует любой из кодов точки q .

На самом деле удобнее действовать в обратном направлении.

Предложение 7.4.1. Если $f: S^1 \rightarrow S^1$ — растягивающее отображение степени 2, тогда f является фактором отображения σ^R на Ω_2^R (см. определение 7.3.3), т. е. имеется такое сюръективное непрерывное отображение $h: \Omega_2^R \rightarrow S^1$, что $f^n(h(\omega)) \in \Delta_{\omega_n}$ для всех $n \in \mathbb{N}_0$, т. е. $h \circ \sigma^R = f \circ h$.

Доказательство. То обстоятельство, что область определения отображения h есть Ω_2^R , в частности, означает, что всякая последовательность из нулей и единиц является кодом некоторой точки. Во-первых, f отображает каждый из двух отрезков Δ_0 и Δ_1 на S^1 почти инъективно, отождествление происходит только на концах. Пусть

Γ_{00} — это ядро множества $\Delta_0 \cap f^{-1}(\Delta_0)$,

Γ_{01} — это ядро множества $\Delta_0 \cap f^{-1}(\Delta_1)$,

Γ_{10} — это ядро множества $\Delta_1 \cap f^{-1}(\Delta_0)$,

Γ_{11} — это ядро множества $\Delta_1 \cap f^{-1}(\Delta_1)$.

Ядро определяется следующим образом. Каждое указанное пересечение состоит из отрезка и изолированной точки (p или q), и мы отбрасываем эту постороннюю точку. Каждый из этих четырех отрезков при действии f^2 отображается на S^1 , причем отождествление вновь происходит только на концах. По определению код любой точки из Δ_{ij} начинается с ij . Продолжая рассуждения по индукции, для любой конечной последовательности $\omega_0, \dots, \omega_{n-1}$ мы строим отрезок

$$\Delta_{\omega_0, \dots, \omega_{n-1}} := \text{ядро множества } \Delta_{\omega_0} \cap f^{-1}(\Delta_{\omega_1}) \dots \cap f^{1-n}(\Delta_{\omega_{n-1}}), \quad (7.4.2)$$

который при действии f^n отображается на S^1 с отождествлением концевых точек. Возьмем теперь любую бесконечную последовательность $\omega = \omega_1, \dots \in \Omega_2^R$. Пе-

пересечение $\bigcap_{n=1}^{\infty} \Delta_{\omega_0, \dots, \omega_{n-1}}$ вложенных отрезков $\Delta_{\omega_0, \dots, \omega_{n-1}}$ непусто, и любая точка в этом пересечении кодируется последовательностью ω .

Пока мы использовали только предположение, что f является монотонным отображением степени 2. Чтобы показать, что отображение h определено корректно, мы используем свойство растяжения и с его помощью проверим, что множество $\bigcap_{n=1}^{\infty} \Delta_{\omega_0, \dots, \omega_{n-1}}$ состоит из единственной точки, а следовательно, точка с данным кодом единственна.

Если $g: I \rightarrow S^1$ — инъективное отображение интервала I с неотрицательной производной, тогда по теореме о конечных приращениях А.2.3 мы получаем, что $l(g(I)) = \int_I g'(x) dx = g'(\xi)l(I)$ для некоторого $\xi \in I$. Таким образом, в нашем случае имеется такое ξ_n , что

$$1 = l(S^1) = \int_{\Delta_{\omega_0, \dots, \omega_{n-1}}} (f^n)'(x) dx = (f^n)'(\xi_n) \cdot l(\Delta_{\omega_0, \dots, \omega_{n-1}}).$$

Поскольку отображение f растягивающее, $|(f^n)'| > \lambda^n$ для некоторого $\lambda > 1$, следовательно, $l(\Delta_{\omega_0, \dots, \omega_{n-1}}) < \lambda^{-n} \rightarrow 0$ при $n \rightarrow \infty$ и множество $\bigcap_{n=1}^{\infty} \Delta_{\omega_0, \dots, \omega_{n-1}}$ состоит из единственной точки x_ω . Итак, мы получаем корректно определенное сюръективное отображение $h: \Omega_2^R \rightarrow S^1$, $\omega \mapsto x_\omega$.

Наделим Ω_2^R метрикой d_4 , определенной соотношением (7.3.3). Как было показано в п. 7.3.4, если $\varepsilon = \lambda^{-n}$ и $\delta = 4^{-n}$, то из неравенства $d(\omega, \omega') < \delta$ следует, что $\omega_i = \omega'_i$ для $i < n$, а значит, $|x_\omega - x_{\omega'}| \leq l(\Gamma_{\omega_0, \dots, \omega_{n-1}}) < \lambda^{-n} = \varepsilon$. Таким образом, отображение h непрерывно.

Равенство $h(\sigma^R(\omega)) = f(h(\omega))$ очевидно из построения. $\square$

7.4.2. Классификация посредством кодирования. Из предложения 7.4.1 и предшествующих ему рассуждений сразу следует полусопряженность между односторонним 2-сдвигом и растягивающим отображением f на S^1 .

Предложение 7.4.2. Пусть $f: S^1 \rightarrow S^1$ — растягивающее отображение степени 2. Тогда f является фактором одностороннего 2-сдвига (Ω_2^R, σ_R) посредством полусопряжения $h: \Omega_2^R \rightarrow S^1$. Если $h(\omega) = h(\omega') =: x$, тогда существует такое $n \in \mathbb{N}_0$, что $f^n(x) \in \{p, q\}$, где $p = f(p) = f(q)$, $q \neq p$.

Вторая часть этого утверждения говорит, что отображение h «очень близко» к сопряжению: имеется только счетное число точек, в которых нарушается инъективность.

Важная особенность этого кодирования состоит в том, что оно получено универсальным способом для всех растягивающих отображений и что нарушение инъективности происходит в точках, определенных их динамикой, а именно в неподвижной точке и ее прообразах. Это приводит к той самой классификации, которая была обещана в начале этого параграфа.

Теорема 7.4.3. Если $f, g: S^1 \rightarrow S^1$ — растягивающие отображения степени 2, то f и g топологически сопряжены; в частности, каждое растягивающее отображение окружности S^1 степени 2 сопряжено с E_2 .

Доказательство. Для отображений f и g имеются отображения полусопряжения $h_f, h_g: \Omega_2^R \rightarrow S^1$. Для $x \in S^1$ рассмотрим множество $H_x := h_g(h_f^{-1}(\{x\}))$. Если x — точка инъективности отображения h_f , т. е. множество $h_f^{-1}(\{x\})$ состоит из одной точки, то таким же свойством обладает и H_x . В противном случае x является прообразом неподвижной точки некоторой итерации отображения f и множество $h_f^{-1}(\{x\})$ состоит из набора последовательностей, которые при действии h_g отображаются в одну точку. Поэтому множество H_x всегда состоит из одной точки $h(x)$. Взаимно однозначное отображение $h: S^1 \rightarrow S^1$, определенное таким образом, является сопряжением: $h \circ f = g \circ h$. Оно непрерывно, потому что h_f отображает открытые множества в открытые, т. е. образ последовательности и всех достаточно близких к ней последовательностей содержит малый интервал. Поменяв местами отображения f и g , мы получим, что отображение h^{-1} также непрерывно. $\square$

Это утверждение на самом деле справедливо для отображений любой степени, по модулю большей 1, в чем можно убедиться с помощью подходящего кодирования. Это первый значительный результат относительно сопряжения, который устанавливает определенный вид сопряженности для всех отображений из некоторого класса. Теорема Пуанкаре о классификации 4.3.20 во многом сходна с этим результатом, но все же требует дополнительных предположений (таких, как существование второй производной; см. п. 4.3.3), чтобы сделать вывод о сопряженности с поворотом.

7.4.3. Квадратичные отображения. Для $\lambda > 4$ рассмотрим квадратичное семейство

$$f: \mathbb{R} \rightarrow \mathbb{R}, \quad x \rightarrow \lambda x(1 - x).$$

Если $x < 0$, тогда $f(x) < x$ и $f'(x) > \lambda > 4$, поэтому $f^n(x) \rightarrow -\infty$. Если $x > 1$, то $f(x) < 0$ и, следовательно, $f^n(x) \rightarrow -\infty$. Таким образом, множество точек с ограниченными орбитами — это $\bigcap_{n \in \mathbb{N}_0} f^{-n}([0, 1])$.

Предложение 7.4.4. Если $\lambda > 2 + \sqrt{5}$ и $f: \mathbb{R} \rightarrow \mathbb{R}, x \rightarrow \lambda x(1 - x)$, то существует такой гомеоморфизм $h: \Omega_2^R \rightarrow \Lambda := \bigcap_{n \in \mathbb{N}_0} f^{-n}([0, 1])$, что $h \circ \sigma^R = f \circ h$, т. е. отображение $f|_\Lambda$ сопряжено с 2-сдвигом.

Доказательство. Пусть

$$\Delta_0 = \left[0, \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{1}{\lambda}}\right] \quad \text{и} \quad \Delta_1 = \left[\frac{1}{2} + \sqrt{\frac{1}{4} - \frac{1}{\lambda}}, 1\right].$$

Тогда $f^{-1}([0, 1]) = \Delta_0 \cup \Delta_1$ (в чем можно убедиться, решая квадратное уравнение $f(x) = 1$). Аналогично множество $f^{-2}([0, 1]) = \Delta_{00} \cup \Delta_{01} \cup \Delta_{11} \cup \Delta_{10}$ состоит из четырех отрезков, и т. д. Рассмотрим разбиение множества Λ на Δ_0 и Δ_1 . Эти

части не пересекаются и

$$\begin{aligned} |f'(x)| &= |\lambda(1-2x)| = 2\lambda \left| x - \frac{1}{2} \right| \geq \\ &\geq 2\lambda \sqrt{\frac{1}{4} - \frac{1}{\lambda}} = \sqrt{\lambda^2 - 4\lambda} > \sqrt{(2 + \sqrt{5})^2 - 4(2 + \sqrt{5})} = 1 \end{aligned}$$

на $\Delta_0 \cup \Delta_1$. Таким образом, для любой последовательности $\omega = (\omega_0, \omega_1, \dots)$ диаметр пересечений

$$\bigcap_{n=0}^N f^{-n}(\Delta_{\omega_n})$$

убывает (экспоненциально) при $N \rightarrow \infty$. Это показывает, что для последовательности $\omega = (\omega_0, \omega_1, \dots)$ пересечение

$$h(\{\omega\}) = \bigcap_{n \in \mathbb{N}_0} f^{-n}(\Delta_{\omega_n}) \quad (7.4.3)$$

состоит из одной точки и отображение $h: \Omega_2^R \rightarrow \Lambda$ является гомеоморфизмом. $\square$

Замечание 7.4.5. Оказывается, предложение 7.4.4 справедливо для любого $\lambda > 4$ (предложение 11.4.1), но доказательство в этом случае гораздо более сложно, чем приведенное выше. В случае, когда при отображении отрезок «складывается», мы говорим, что возникает одномерная *подкова*, по аналогии с геометрической ситуацией, рассмотренной в следующем пункте.

7.4.4. Линейная подкова. Опишем теперь знаменитую «подкову» Смейла, которая служит одним из лучших примеров точного кодирования. Частный случай уже был рассмотрен в п. 7.3.3, где кодирование строилось с помощью тричного разложения.

Пусть Δ — прямоугольник в $\mathbb{R}^2$ и $f: \Delta \rightarrow \mathbb{R}^2$ — такой гладкий гомеоморфизм этого прямоугольника на его образ, что пересечение $\Delta \cap f(\Delta)$ состоит из двух «горизонтальных» прямоугольников Δ_0 и Δ_1 и ограничение гомеоморфизма f на компоненты $\Delta^i := f^{-1}(\Delta_i)$, $i = 0, 1$, множества $f^{-1}(\Delta)$ является гиперболическим линейным отображением, сжимающим в вертикальном направлении и растягивающим в горизонтальном. Из этого следует, что множества Δ^0 и Δ^1 являются «вертикальными» прямоугольниками. Этого легче всего добиться, если согнуть прямоугольник Δ в «подкову», или, скорее, в форму постоянного магнита (рис. 7.4.2), хотя при этом возникают некоторые неудобства, связанные с ориентацией. Другой способ, который позволяет избежать этих трудностей, состоит в том, чтобы согнуть прямоугольник Δ так, чтобы получилось что-то похожее на скрепку для бумаги (рис. 7.4.3). Это «утолщенный» вариант тричной подковы из п. 7.3.3. Если горизонтальные и вертикальные прямоугольники лежат строго внутри Δ , тогда максимальное инвариантное подмножество $\Lambda = \bigcap_{n=-\infty}^{\infty} f^{-n}(\Delta)$ множества Δ содержится во внутренней части множества Δ .

Предложение 7.4.6. Отображение $f|_{\Lambda}$ топологически сопряжено с σ_2 .

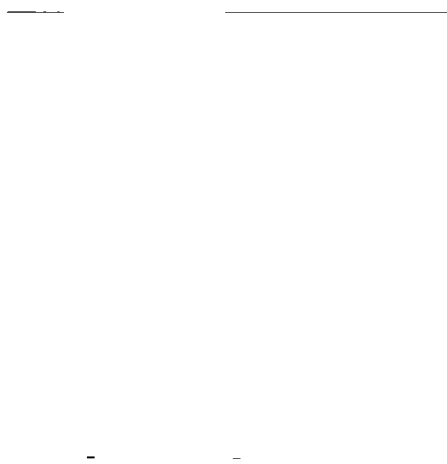


Рис. 7.4.2. Подкова

Рис. 7.4.3. Скрепка для бумаги

Доказательство. При построении кодирования мы будем использовать множества Δ^0 и Δ^1 как элементы разбиения и начнем с положительных итераций. Пересечение $\Delta \cap f(\Delta) \cap f^2(\Delta)$ состоит из четырех тонких горизонтальных прямоугольников: $\Delta_{ij} = \Delta_i \cap f(\Delta_j) = f(\Delta^i) \cap f^2(\Delta^j)$, $i, j \in \{0, 1\}$ (см. рис. 7.4.2).

Продолжая рассуждения по индукции, мы видим, что множество $\bigcap_{i=0}^n f^i(\Delta)$ состоит из 2^n тонких непересекающихся горизонтальных прямоугольников, высоты которых убывают экспоненциально с ростом n . Каждый такой прямоугольник имеет вид $\Delta_{\omega_1, \dots, \omega_n} = \bigcap_{i=1}^n f^i(\Delta^{\omega_i})$, где $\omega_i \in \{0, 1\}$ для $i = 1, \dots, n$. Каждое бесконечное пересечение $\bigcap_{n=1}^{\infty} f^n(\Delta^{\omega_n})$, $\omega_n \in \{0, 1\}$, является горизонтальным отрезком, а пересечение $\bigcap_{n=1}^{\infty} f^n(\Delta)$ есть произведение горизонтального отрезка и канторова множества в вертикальном направлении. Точно так же определяются и изучаются вертикальные прямоугольники $\Delta^{\omega_0, \dots, \omega_{-n}} = \bigcap_{i=0}^n f^{-i}(\Delta^{\omega_{-i}})$, вертикальные отрезки $\bigcap_{n=0}^{\infty} f^{-n}(\Delta^{\omega_{-n}})$ и множество $\bigcap_{n=0}^{\infty} f^{-n}(\Delta)$, которое является произведением отрезка в вертикальном направлении и канторова множества в горизонтальном.

Искомое инвариантное множество $\Lambda = \bigcap_{n=-\infty}^{\infty} f^{-n}(\Delta)$ есть произведение двух канторовых множеств, а следовательно, оно само является канторовым множеством (см. упражнение 2.7.5), и отображение

$$h: \Omega_2 \rightarrow \Lambda, \quad h(\{\omega\}) = \bigcap_{n=-\infty}^{\infty} f^{-n}(\Delta^{\omega_n}),$$

является гомеоморфизмом, который сопрягает сдвиг σ_2 и ограничение диффеоморфизма f на множество Λ . $\square$

Так как периодические точки и топологическое перемешивание являются инвариантами топологической сопряженности, из предложений 7.4.6 и 7.3.4 мы немедленно получим существенную информацию относительно поведения диффеоморфизма f на Λ .

Следствие 7.4.7. *Периодические точки диффеоморфизма f плотны в Λ , $P_n(f|_{\Lambda}) = 2^n$, и ограничение диффеоморфизма f на Λ является топологически перемешивающим.*

Замечание 7.4.8. Любое отображение, для которого имеется точное кодирование, определено на канторовом множестве, потому что точное кодирование устанавливает гомеоморфизм между фазовым пространством и пространством последовательностей, которое является канторовым множеством.

7.4.5. Кодирование автоморфизма тора. Идея кодирования может быть применена к гиперболическому автоморфизму тора. Чтобы упростить обозначения и сделать построения более наглядными, мы рассмотрим стандартный пример. Это первый из наших примеров, в которых кодирование нетривиально, хотя его несложно представить себе геометрически. В § 10.3 описывается конструкция, имеющая сходный динамический смысл, но ее геометрическая интерпретация гораздо более сложна и почти всегда включает фракталы.

Теорема 7.4.9. *Для отображения*

$$F(x, y) = (2x + y, x + y) \pmod{1}$$

двумерного тора из п. 7.1.4 существует такое полусопряжение $h: \Omega_A \rightarrow \mathbb{T}^2$, что

$$F \circ h = h \circ \sigma_5|_{\Omega_A},$$

где

$$A = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix} \quad (7.4.4)$$

Доказательство. Изобразим отрезки двух прямых, порожденных собственными векторами, выходящими из начала координат, и продолжим их до тех пор, пока они не пересекут тор достаточное количество раз и не разделят его на непересекающиеся прямоугольники. Точнее говоря, мы будем продолжать отрезок сжимающейся прямой в четвертый квадрант, пока он не пересечет отрезок растягивающейся прямой дважды в первом квадранте и один раз в третьем (см. рис. 7.4.4). В результате мы получим разбиение тора на два прямоугольника $R^{(1)}$ и $R^{(2)}$. Три пары из семи вершин плоской конфигурации отождествляются, так что имеется только четыре различные точки на торе, которые служат вершинами прямоугольников, — начало координат и три точки пересечения. Хотя прямоугольники $R^{(1)}$ и $R^{(2)}$ и имеют общие точки, можно применить метод, используемый для подковы, считая $R^{(1)}$

Рис. 7.4.4. Разбиение тора

Рис. 7.4.5. Образ разбиения

и $R^{(2)}$ основными прямоугольниками. Растягивающие и сжимающие направления собственных векторов играют роль «горизонтального» и «вертикального» направлений соответственно. Как видно из рис. 7.4.5, образ $F(R^{(i)})$ ($i = 1, 2$) состоит из нескольких «горизонтальных» прямоугольников полной длины. Объединение границ $\partial R^{(1)} \cup \partial R^{(2)}$ состоит из отрезков двух собственных прямых, проходящих через начало координат, которые были только что описаны. Образ сжимающегося отрезка является его частью. Итак, образы прямоугольников $R^{(1)}$ и $R^{(2)}$ должны быть «закреплены» на участках их «вертикальных» сторон, т.е. если один из образов «входит» в какой-либо из прямоугольников $R^{(1)}$ или $R^{(2)}$, он должен полностью пройти через него. Проводя соответствующие построения по сжимающемуся направлению, мы видим, что множество $F(R^{(1)})$ состоит из трех компонент, две из которых лежат в $R^{(1)}$ и одна — в $R^{(2)}$. Образ прямоугольника $R^{(2)}$ имеет две компоненты, по одной в каждом прямоугольнике (см. рис. 7.4.5). Тот факт, что множество $F(R^{(1)})$ имеет две компоненты в $R^{(1)}$, привел бы к трудностям, если бы мы использовали прямоугольники $R^{(1)}$ и $R^{(2)}$ для кодирования (для некоторых последовательностей получалось бы больше одной точки с заданным кодом), но мы используем в качестве составляющих при нашем кодировании пять компонент $\Delta_0, \Delta_1, \Delta_2, \Delta_3, \Delta_4$ (или их прообразы). В этом случае имеется в точности один прямоугольник $\Delta_{\omega_{-1} \dots \omega_0 \omega_1 \dots \omega_k}$, определенный как пересечение $\bigcap_{n=-l}^k F^{-n}(\Delta_{\omega_n})$. (Как и в случае растягивающих отображений в п. 7.3.1, мы

должны удалить «лишние» части, в данном случае это будут отрезки прямой.) Благодаря свойству сжатия отображения F в «вертикальном» направлении «высота» прямоугольника $\Delta_{\omega_{-1} \dots \omega_0 \omega_1 \dots \omega_k}$ меньше чем $\left(\frac{3 - \sqrt{5}}{2}\right)^l$, а из-за сжатия отображения F^{-1} в «горизонтальном» направлении «ширина» прямоугольника $\Delta_{\omega_{-1} \dots \omega_0 \omega_1 \dots \omega_k}$ меньше чем $\left(\frac{3 - \sqrt{5}}{2}\right)^k$. Оба выражения стремятся к нулю при $l \rightarrow \infty$ и $k \rightarrow \infty$, так что пересечение $\bigcap_{n \in \mathbb{Z}} F^{-n}(\Delta_{\omega_n})$ содержит не более одной точки $h(\omega)$. С другой стороны, из-за «марковского» свойства, описанного ранее, т.е. в силу того, что образы проходят через прямоугольники полностью, справедливо следующее утверждение: если $\omega \in \Omega_5$ и $F^{-1}(\Delta_{\omega_n})$ пересекается с $\Delta_{\omega_{n+1}}$ для всех $n \in \mathbb{Z}$, то имеется точка $h(\omega) \in \bigcap_{n \in \mathbb{Z}} F^{-n}(\Delta_{\omega_n})$. Таким образом, мы имеем кодирование, которое, однако, определено не для всех последовательностей из Ω_5 .

Вместо этого мы должны ограничиться подпространством Ω_4 пространства Ω_5 , содержащим только те последовательности, в которых любые два последовательных элемента составляют «допустимый переход», т.е. за элементами 0, 1, 2 может следовать 0, 1 или 3, а за 3 и 4 может следовать 2 или 4. Это в точности топологическая цепь Маркова (см. определение 7.3.2), соответствующая матрице (7.4.4). $\square$

Теорема 7.4.10. *Полусопряжение между отображениями σ_A и F взаимно однозначно на всех периодических точках, кроме неподвижных точек. Число прообразов любой точки, не отрицательно асимптотической к неподвижной точке, ограничено.*

Доказательство. Опишем более детально отождествления, возникающие при полусопряжении, т.е. определим, какие точки на торе имеют больше чем один прообраз. Во-первых, очевидно, что топологическая цепь Маркова σ_A имеет три неподвижные точки, а именно постоянные последовательности, состоящие из нулей, единиц или четверок, в то время как автоморфизм тора F имеет только одну неподвижную точку — начало координат. Легко видеть, что все три неподвижные точки отображаются в начало координат. Как мы уже знаем из предложения 7.1.10, $P_n(F) = \lambda_1^n + \lambda_1^{-n} - 2$, и соответственно $P_n(\sigma_A) = \text{tr } A^n = \lambda_1^n + \lambda_1^{-n} = P_n(F) + 2$ (см. следствие 7.3.6), где $\lambda_1 = (3 + \sqrt{5})/2$ — максимальное собственное значение для (2×2) -матрицы $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ и для (5×5) -матрицы (7.4.4). Чтобы убедиться, что собственные значения совпадают, рассмотрим матрицу $A - \lambda_1 \text{Id}$, вычтем четвертый столбец из первых двух, а пятый — из третьего, затем прибавим первую и вторую строки к четвертой, а третью к пятой:

$$\begin{pmatrix} 1-\lambda & 1 & 0 & 1 & 0 \\ 1 & 1-\lambda & 0 & 1 & 0 \\ 1 & 1 & -\lambda & 0 & 0 \\ 0 & 0 & 1 & -\lambda & 1 \\ 0 & 0 & 1 & 0 & 1-\lambda \end{pmatrix} \rightarrow \begin{pmatrix} -\lambda & 0 & 0 & 1 & 0 \\ 0 & -\lambda & 0 & 1 & 0 \\ 0 & 0 & -\lambda & 1 & 0 \\ \lambda & \lambda & 0 & -\lambda & 1 \\ 0 & 0 & \lambda & 0 & 1-\lambda \end{pmatrix} \rightarrow \begin{pmatrix} -\lambda & 0 & 0 & 1 & 0 \\ 0 & -\lambda & 0 & 1 & 0 \\ 0 & 0 & -\lambda & 1 & 0 \\ 0 & 0 & 0 & 2-\lambda & 1 \\ 0 & 0 & 0 & 1 & 1-\lambda \end{pmatrix}.$$

Кроме того, можно видеть, что каждая точка $q \in \mathbb{T}^2$, положительные и отрицательные итерации которой не попадают на границы $\partial R^{(1)}$ и $\partial R^{(2)}$, имеет единственный прообраз и наоборот. В частности, периодические точки, отличные от начала координат (которые имеют рациональные координаты), попадают в этот класс. Точки пространства Ω_A , образы которых лежат на границах или их итерациях при отображении F , распадаются на три категории, соответствующие трем отрезкам устойчивых и неустойчивых многообразий, проходящих через 0, которые определяют части границы. Таким образом, последовательности отождествляются в следующих случаях: если они имеют постоянный бесконечный правый (будущий) «хвост», состоящий из нулей или четверок, и совпадают на оставшихся местах — это соответствует устойчивой граничной части — или же имеют бесконечный левый (прошлый) «хвост», состоящий из нулей и единиц или четверок, и совпадают на оставшемся участке — это соответствует неустойчивой граничной части. $\square$

Упражнения

7.4.1. Докажите, что для $\lambda \geq 1$ каждая ограниченная орбита квадратичного отображения f_λ лежит в $[0, 1]$.

7.4.2. Приведите детальное доказательство того, что соотношение (7.4.3) определяет гомеоморфизм.

7.4.3. Постройте марковское разбиение для матрицы $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, которое состоит из двух квадратов.

7.4.4. Постройте марковское разбиение и опишите соответствующую топологическую цепь Маркова для автоморфизма F_L , где $L = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}$.

7.4.5. Для данной 0-1 ($n \times n$)-матрицы A опишите систему из n прямоугольников $\Delta_1, \dots, \Delta_n$ в $\mathbb{R}^2$ и такое отображение $f: \Delta := \bigcup_{i=1}^n \Delta_i \rightarrow \mathbb{R}^2$, что его ограничение на множество точек, которые остаются внутри Δ при всех его итерациях, топологически эквивалентно топологической цепи Маркова σ_A .

7.4.6. Убедитесь, что, отбрасывая посторонние точки при кодировании в соответствии с равенством (7.4.2), мы получим $\Delta_{\omega_0, \dots, \omega_{n-1}} = \bigcap_{i=0}^{n-1} \text{Int}(f^{-i}(\Delta_{\omega_i}))$ и $\{h(\omega)\} := \bigcap_{n \in \mathbb{N}} \Delta_{\omega_0, \dots, \omega_{n-1}}$.

Задачи для углубленного изучения

7.4.7. Покажите, что утверждение теоремы 7.4.3 остается справедливым для любого такого отображения f степени 2, что $f' \geq 1$ и $f' = 1$ лишь в конечном числе точек.

7.4.8. Докажите утверждение теоремы 7.4.9 для некоторой 0-1-матрицы A и для любого автоморфизма

$$F_L: \mathbb{T}^2 \rightarrow \mathbb{T}^2, \quad x \mapsto Lx \pmod{1},$$

где L — целочисленная (2×2) -матрица с определителем, равным $+1$ или -1 , и действительными собственными значениями, отличными от ± 1 .

§ 7.5. Равномерное распределение

Теперь мы выясним, имеет ли понятие равномерного распределения орбит, введенное в предыдущих главах для поворотов окружности и параллельных переносов на торе, какой-либо смысл для примеров, изучаемых в настоящей главе, т. е., скажем, для линейных или нелинейных растягивающих отображений окружности, сдвигов в пространстве последовательностей и автоморфизмов тора.

7.5.1. Отсутствие строгой эргодичности. Рассмотрим линейные растягивающие отображения более подробно. Можно определить асимптотические частоты попаданий в отрезок $\Delta \subset S^1$, как это сделано в (4.1.2), а также усреднение Биркгофа функций, как это сделано в (4.1.5). Как мы знаем из предложения 4.1.7 и теоремы 4.1.15, для поворотов эти частоты сходятся равномерно. Это свойство называется строгой эргодичностью (см. определение 4.1.18).

Однако в примерах, изучаемых в настоящей главе, соответствующие выражения не сходятся равномерно к постоянным, т. е. линейные растягивающие отображения не являются строго эргодическими. Действительно, рассмотрим непрерывную функцию, которая обращается в нуль только в нуле. Тогда усреднение Биркгофа неподвижной точки равно 0, в то время как все другие периодические точки имеют ненулевые средние. Более того, имеются орбиты, для которых предел средних частот попаданий в отрезок не существует, как это можно усмотреть из конструкции кодирования. Для отображения E_2 средняя частота попаданий в отрезок $[0, 1/2]$ точки $x \in S^1$ с единственным двоичным представлением равна доле нулей среди первых n разрядов. Пусть x — точка, двоичное представление которой состоит из чередующихся блоков нулей и единиц, причем длина n -го блока равна произведению числа n на общую длину всех предыдущих блоков. Доля нулей в последовательности, полученной после того, как мы допишем n -й блок нулей (который является $(2n - 1)$ -м блоком), больше чем $1 - 1/(2n - 1)$, однако в последовательности, полученной после дописывания n -го блока единиц ($2n$ -й блок), эта доля меньше чем $1/(2n)$. Таким образом, предельные точки для средних частот покрывают весь отрезок $[0, 1]$.

Поэтому мы должны исследовать другой вид сходимости.

7.5.2. Сходимость в среднем. Орбиты-контрпримеры, приведенные выше, имеют довольно специальный вид, так что есть надежда, что «большая часть» орбит все же равномерно распределена или что во всяком случае «в среднем» сходимость наблюдается. Это действительно так в довольно большой общности. Мы покажем это в явном виде для отображения E_2 .

Предложение 7.5.1. Если $\varphi(x) := \chi_{[0, 1/2]} := \begin{cases} 1 & \text{при } x \leq 1/2, \\ 0 & \text{при } x > 1/2, \end{cases}$ — характеристическая функция отрезка $[0, 1/2]$, и

$$\mathcal{B}_n(\varphi)(x) := \frac{1}{n} \sum_{k=0}^{n-1} \varphi(E_2^k(x))$$

— усреднение Биркгофа, то $\int_{S^1} \left| \mathcal{B}_n(\varphi)(x) - \int_{S^1} \varphi(t) dt \right| dx \xrightarrow{n \rightarrow \infty} 0$ экспоненциально быстро.

Замечание 7.5.2. Можно доказать аналогичные результаты для любого двоичного отрезка, и, переходя к линейной комбинации и аппроксимации, мы получим равномерное распределение.

Доказательство. В терминах двоичного представления $x = 0, x_1 x_2 \dots$ мы имеем

$$n\mathcal{B}_n(\varphi)(x) = F_{[0,1/2]}(x, n) := \text{card}\{k: 0 \leq k \leq n-1 \text{ и } E_2^k x \in [0, 1/2]\} = \sum_{k=0}^{n-1} (1 - x_k).$$

Таким образом, среднее Биркгофа постоянно на каждом двоичном отрезке

$$\Delta_n^i = \left[\frac{i}{2^n}, \frac{i+1}{2^n} \right], \quad i = 0, 1, \dots, 2^n - 1.$$

Как уже отмечено прежде, это среднее не может быть равномерно близко к какой-либо постоянной; на самом деле оно может принимать все значения вида i/n , $i = 0, 1, \dots, n$. Однако для большинства отрезков средняя частота попаданий близка к $\ell([0, 1/2]) = 1/2$ (см. рис. 7.5.1). Чтобы показать это, мы используем

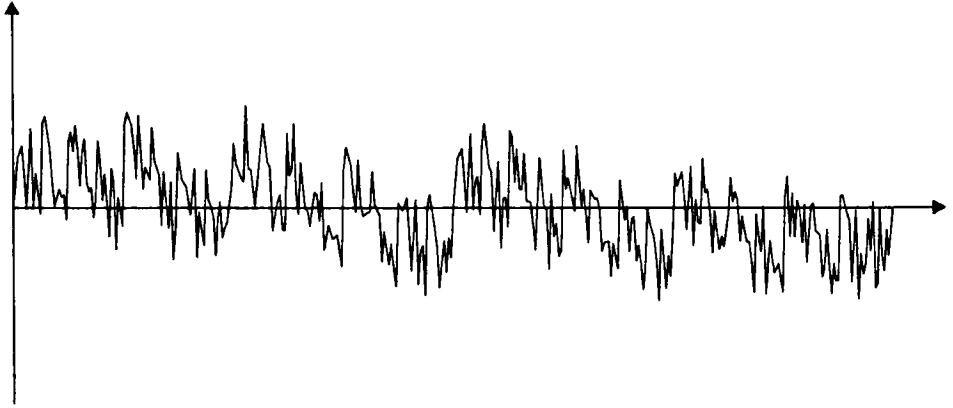


Рис. 7.5.1. Распределение после 10 итераций

то обстоятельство, что число последовательностей нулей и единиц длины n , которые имеют в точности k нулей (это соответствует средней частоте k/n), равно биномиальному коэффициенту $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, и соответственно доля таких последовательностей равна $\binom{n}{k} 2^{-n}$. Таким образом, если $\epsilon > 0$, то

$$\text{card}\{i: |\mathcal{B}_n(\varphi)(x) - 1/2| < \epsilon \text{ для } x \in \Delta_n^i\} = \sum_{k=\lfloor (1/2-\epsilon)n \rfloor+1}^{\lfloor (1/2+\epsilon)n \rfloor} \binom{n}{k}, \quad (7.5.1)$$

где $\lfloor \cdot \rfloor$ обозначает целую часть. Чтобы указать нижнюю оценку, мы оценим сумму оставшихся биномиальных коэффициентов сверху. Поскольку $\binom{n}{k} = \binom{n}{n-k}$,

эта сумма равна

$$2 \sum_{k=0}^{\lfloor (1/2-\epsilon)n \rfloor} \binom{n}{k}.$$

Биномиальные коэффициенты растут с ростом k при $k < n/2$, потому что отношение последовательных коэффициентов равно $(k+1)/(n-k)$. Таким образом, самый большой член в сумме — последний. Поскольку число членов в сумме не превышает $n/2$, мы находим, что

$$2 \sum_{k=0}^{\lfloor (1/2-\epsilon)n \rfloor} \binom{n}{k} \leq n \binom{n}{\lfloor (1/2-\epsilon)n \rfloor} = n \binom{n}{\lfloor \alpha n \rfloor},$$

где $\alpha = (1/2) - \epsilon$. Разделив на 2^n , чтобы найти долю «плохих» последовательностей, мы видим, что необходима верхняя оценка величины $n \binom{n}{\lfloor \alpha n \rfloor} 2^{-n}$.

Мы используем классическую формулу Стирлинга $n! \asymp \sqrt{2\pi n} n^n e^{-n}$, где $f(n) \asymp g(n)$ означает $\lim_{n \rightarrow \infty} f(n)/g(n) = 1$. Обозначая $l := \lfloor \alpha n \rfloor$, мы получаем

$$\binom{n}{\lfloor \alpha n \rfloor} = \frac{n!}{l!(n-l)!} \asymp \frac{n^n e^{-n} \sqrt{2\pi n}}{l^l (n-l)^{n-l} e^{-l} e^{-(n-l)} \sqrt{2\pi l} \sqrt{2\pi(n-l)}},$$

и, следовательно,

$$n \binom{n}{\lfloor \alpha n \rfloor} 2^{-n} \asymp n e^{n \ln n - l \ln l - (n-l) \ln(n-l) - n \ln 2} \sqrt{\frac{n}{2\pi l(n-l)}}. \quad (7.5.2)$$

Чтобы получить верхнюю оценку для величины

$$\begin{aligned} n \ln n - l \ln l - (n-l) \ln(n-l) - n \ln 2 &= \\ &= (n-l) [\ln n - \ln(n-l) - \ln 2] + l [\ln n - \ln l - \ln 2] = \\ &= (n-l) \ln \frac{n}{2(n-l)} + l \ln \frac{n}{2l} = \\ &= (n-l) \ln \left(1 + \frac{2l-n}{2(n-l)} \right) + l \ln \left(1 + \frac{n-2l}{2l} \right), \end{aligned}$$

мы воспользуемся тем, что $\ln(1+x) \leq x$ (потому что логарифм — выпуклая вниз функция), а на самом деле справедливо даже неравенство $\ln(1+x) \leq x - \delta$ при $|x| > \epsilon$. Поскольку $2l = 2\lfloor \alpha n \rfloor = 2\left\lfloor \frac{n}{2} - n\epsilon \right\rfloor \asymp (1-2n)\epsilon$ для больших n , мы получаем $\frac{2l-n}{2(n-l)} \asymp 2\epsilon$ и $\frac{n-2l}{2l} \asymp -2\epsilon$, так что

$$\begin{aligned} (n-l) \ln \left(1 + \frac{2l-n}{2(n-l)} \right) + l \ln \left(1 + \frac{n-2l}{2l} \right) &\leq \\ &\leq (n-l) \left(\frac{2l-n}{2(n-l)} - \delta \right) + l \left(\frac{n-2l}{2l} - \delta \right) = -n\delta, \end{aligned}$$

где $\delta = \delta(\epsilon) > 0$. Взяв δ достаточно малым, чтобы погрешность в соотношении (7.5.2) была мала, мы получим

$$2 \sum_{k=0}^{\lfloor (1/2-\epsilon)n \rfloor} \binom{n}{k} 2^{-n} \leq ne^{-n\delta} \sqrt{\frac{n}{2\pi \lfloor \alpha n \rfloor (n - \lfloor \alpha n \rfloor)}} =: \Delta(n, \epsilon), \quad (7.5.3)$$

а это выражение экспоненциально стремится к 0 при $n \rightarrow \infty$.

Возвращаясь к равенству (7.5.1), мы видим, что для любого фиксированного $\epsilon > 0$ и для любого достаточно большого n доля двоичных отрезков, для которых среднее число нулей отличается от $1/2$ меньше чем на ϵ , ограничена снизу величиной $1 - \Delta(n, \epsilon)$ и, следовательно, экспоненциально близка к 1. Так как все отрезки имеют одну и ту же длину, функция $B_n(\varphi)$ близка к $1/2$ по всей области определения, за исключением объединения отрезков, сумма длин которых экспоненциально стремится к нулю при $n \rightarrow \infty$. Поскольку функция ограничена (сверху и снизу), из этого следует, что среднее отклонение от $1/2 = \ell([0, 1/2]) = \int_{S^1} \varphi(t) dt$ экспоненциально мало. $\square$

Ключевой момент доказательства состоит в сведении оценки полной длины «плохого» множества (для которого средние Биркгофа отличны от средних остального пространства) к комбинаторному вычислению числа отрезков. Мы можем это сделать, потому что все двоичные отрезки данного ранга имеют одну и ту же длину. Из этого, в свою очередь, следует (отметим, что обратное утверждение неверно), что отображение E_2 сохраняет длину в смысле полных образов (см. п. 7.1.1).

7.5.3. Поточечная сходимость средних. Имеется существенная разница между равномерным распределением в среднем (усреднения Биркгофа для больших n близки к среднему по пространству вне множества малой длины) и первоначальным понятием равномерного распределения, которое рассматривалось в гл. 4, где мы вычисляли усреднения Биркгофа отдельных точек. Естественно предположить, что равномерное распределение в среднем влечет сходимость усреднений Биркгофа для «большинства» точек. Задача состоит в том, чтобы придать точный смысл слову «большинство». Даже в самом простом случае, который мы только что рассмотрели, и множество A , состоящее из точек, для которых усреднения Биркгофа сходятся к среднему по пространству, и его дополнение плотны в пространстве, так что характеристическая функция множества A не интегрируема по Риману и A не имеет длины. Однако имеется адекватное понятие малости, которое пригодно в нашем случае.

Определение 7.5.3. Говорят, что множество в $\mathbb{R}$ имеет меру нуль, если для любого $\epsilon > 0$ оно может быть покрыто набором интервалов (не обязательно конечным или непересекающимся), сумма длин которых не превосходит ϵ . Про свойство точек, которое справедливо для всех точек, кроме точек из множества меры нуль, говорят, что оно выполнено *почти всюду* или для *почти каждой* точки.

Любое подмножество множества меры нуль само является множеством меры нуль. Очевидные примеры множеств меры нуль — конечные множества. Даже

счетное множество всегда является множеством меры нуль, поскольку оно (по определению счетности) может рассматриваться как последовательность $(x_n)_{n \in \mathbb{N}}$, покрываемая интервалами $(x_n - \varepsilon 2^{-n-1}, x_n + \varepsilon 2^{-n-1})$, длины которых $\varepsilon 2^{-n}$ в сумме составляют ε . Это означает, что, например, $\mathbb{Q}$ — множество меры нуль, так же как и множество периодических точек отображения E_2 . Счетное объединение $\bigcup_{n \in \mathbb{N}} N_n$ множеств меры нуль имеет меру нуль, так как мы можем покрыть множества N_n интервалами, общая длина которых не больше $\varepsilon 2^{-n}$. Тройное канторово множество имеет меру нуль, потому что оно покрывается объединением S_n из 2^n отрезков длины 3^{-n} каждый (см. п. 2.7.1) и, следовательно, 2^n интервалами несколько большей длины для любого $n \in \mathbb{N}$.

Лемма 7.5.4. *Интервал, полуинтервал или отрезок, отличный от одной точки, не является множеством меры нуль.*

Доказательство. Всякий интервал или полуинтервал содержит отрезок, отличный от точки, так что достаточно доказать лемму для отрезков. Рассмотрим покрытие отрезка $[a, b]$ интервалами. Имеется конечное подпокрытие (в силу компактности, см. определение А.1.18). Рассмотрим все концевые точки интервалов из этого покрытия, которые лежат в (a, b) , и упорядочим их: $x_0 := a < x_1 < x_2 < \dots < x_k < b =: x_{k+1}$. Отрезок $[a, b]$ распадается в объединение k полуинтервалов $I_1 = [a, x_1]$, $I_2 = [x_1, x_2]$, $\dots$, $I_k = [x_{k-1}, x_k]$ и отрезка $I_{k+1} = [x_k, b]$. Каждый полуинтервал или отрезок I_j покрыт m_j раз интервалами из нашего конечного покрытия, и, следовательно, сумма длин этих интервалов не меньше чем $m_j(x_{j+1} - x_j) > x_{j+1} - x_j$. Поэтому сумма длин всех интервалов из этого конечного подпокрытия, а следовательно, и всех интервалов из первоначального покрытия ограничена снизу величиной $(b - x_k) + (x_k - x_{k-1}) + \dots + (x_1 - a) = b - a$. $\square$

Следствие 7.5.5. *Дополнение к множеству меры нуль плотно.*

Доказательство. В противном случае множество меры нуль содержало бы отрезок, отличный от точки. $\square$

Теорема 7.5.6. *В обозначениях предложения 7.5.1 выполняется условие $\mathcal{B}_n(\varphi)(x) \xrightarrow{n \rightarrow \infty} \int_{S^1} \varphi(t) dt$ почти всюду.*

Замечание 7.5.7. Сравнивая этот результат с предложением 7.5.1, следует отметить, что здесь отсутствует экспоненциальная сходимость, но зато мы получаем контроль над почти каждой точкой.

Доказательство. Множество точек x , в которых имеется сходимости, равно

$$\left\{x: \mathcal{B}_n(\varphi)(x) \rightarrow \frac{1}{2}\right\} = \bigcap_{m=1}^{\infty} \bigcup_{N \in \mathbb{N}} \bigcap_{n \geq N} \left\{x: \left|\mathcal{B}_n(\varphi)(x) - \frac{1}{2}\right| < \frac{1}{m}\right\},$$

так что мы должны показать, что («плохое») множество

$$B := \left\{x: \mathcal{B}_n(\varphi)(x) \not\rightarrow \frac{1}{2}\right\} = \bigcup_{m=1}^{\infty} \bigcap_{N \in \mathbb{N}} \bigcup_{n \geq N} \left\{x: \left|\mathcal{B}_n(\varphi)(x) - \frac{1}{2}\right| \geq \frac{1}{m}\right\}$$

является множеством меры нуль. Для этого заметим, что множество

$$\{x: |B_n(\varphi)(x) - 1/2| \geq 1/m\}$$

может быть покрыто набором (двоичных) интервалов, сумма длин которых не превосходит $\Delta(n, 1/m)$, где величина Δ определена соотношением (7.5.3) и экспоненциально мала по n для любого заданного m . Поэтому множество

$$\bigcup_{n \geq N} \{x: |B_n(\varphi)(x) - 1/2| \geq 1/m\}$$

может быть покрыто интервалами, сумма длин которых не превосходит $\bar{\Delta}(N, 1/m) := \sum_{n \geq N} \Delta(n, 1/m)$. Этот ряд сходится (геометрически, см. (7.5.3)), так что $\bar{\Delta}(N, 1/m) \xrightarrow{N \rightarrow \infty} 0$. Найдется такое число N_0 , что $\bar{\Delta}(N_0, 1/m) < \epsilon 2^{-m}$, так что множество $\bigcap_{N \in \mathbb{N}} \bigcup_{n \geq N} \{x: |B_n(\varphi)(x) - 1/2| \geq 1/m\}$ может быть покрыто объединением интервалов, сумма длин которых не превосходит $\epsilon 2^{-m}$. Объединяя эти покрытия, мы получим покрытие множества B интервалами, общая длина которых не превосходит ϵ . $\square$

7.5.4. Закон больших чисел. Два вида равномерного распределения, а именно $\int_{S^1} |B_n(\varphi)(x) - \int_{S^1} \varphi(t) dt| dx \rightarrow 0$ (предложение 7.5.1) и $B_n(\varphi)(x) \xrightarrow{n \rightarrow \infty} \int_{S^1} \varphi(t) dt$ почти всюду (теорема 7.5.6), называются соответственно слабым и сильным законами больших чисел, потому что они описывают тот факт, что любое начальное распределение вероятности (представленное непрерывной функцией φ) стремится к равномерному распределению после повторных применений отображения, т. е. большое число итераций стремится в конечном счете сделать любое распределение равномерным.

Оба эти понятия равномерного распределения слабее, чем строго равномерное распределение, которое наблюдалось для иррациональных поворотов в теореме 4.1.15 и, как мы показали, было связано с понятием строгой эргодичности (см. определение 4.1.18). Подобным образом слабый и сильный законы больших чисел связаны с понятием *эргодичности*, которое мы здесь не определяем, потому что это требует знакомства с теорией меры. Следствием этой теории (а именно эргодической теоремы Биркгофа) является то, что, несмотря на кажущееся различие, оба введенных понятия равномерного распределения эквивалентны.

7.5.5. Распределение периодических точек. Другое интересное заключение относительно средних для некоторых орбит можно сделать, анализируя доказательство теоремы 7.5.6.

Теорема 7.5.8. Для любого $\epsilon > 0$ величина

$$\frac{\text{card}\{p: E_2^n(p) = p, |\lim_{m \rightarrow \infty} B_m(\varphi)(p) - 1/2| \geq \epsilon\}}{2^n - 1}$$

сходится к 0 экспоненциально.

Доказательство. Имеется $2^n - 1$ периодических точек с периодом n для преобразования E_2 , по одной в каждом двоичном отрезке Δ_n^i , $i = 0, \dots, 2^n - 1$, где отрезки Δ_n^0 и $\Delta_n^{2^n-1}$ имеют общий конец 0. Обозначим n -периодическую точку из отрезка Δ_n^i через p_n^i . Поскольку функция $\mathcal{B}_n$ постоянна на Δ_n^i , значения $\mathcal{B}_n(p_n^i)$ для всех этих $2^n - 1$ периодических точек, кроме экспоненциально малой их доли, находятся на расстоянии не более чем ϵ от $1/2$. Но для любой n -периодической точки p выполняется равенство $\lim_{m \rightarrow \infty} \mathcal{B}_m(p) = \mathcal{B}_n(p)$, так как средние берутся по периодической последовательности. $\square$

Этот результат не является следствием теоремы 7.5.6, потому что множество периодических точек имеет меру нуль. Скорее, это довольно естественное следствие того факта, что множество периодических точек «равномерно распределено».

Упражнения

7.5.1. Докажите, что множество точек, для которых при отображении E_2 имеются асимптотические частоты, есть счетное объединение нигде не плотных множеств.

7.5.2. Докажите утверждение, аналогичное предложению 7.5.1, для отображения E_3 .

7.5.3. Докажите утверждение, аналогичное предложению 7.5.1, для отображения E_m .

7.5.4. Докажите утверждение, аналогичное теореме 7.5.6, для отображения E_3 .

7.5.5. Докажите утверждение, аналогичное теореме 7.5.6, для отображения E_m .

7.5.6. Докажите утверждение, аналогичное теореме 7.5.8, для отображения E_m .

§ 7.6. Независимость, энтропия, перемешивание

Исследование свойств распределения орбит представляет собой вероятностный подход к динамике. Так, рассмотренные примеры, для которых имеется сходимости к средним, но равномерная сходимости полностью отсутствует, наводят на мысль, что даже детерминированные динамические системы обладают свойствами, которые отражают явления случайного характера. Рассмотрим более подробно динамические системы, в которых такие свойства играют ключевую роль.

7.6.1. Модель подбрасывания монеты. Рассмотрим конструкцию кодирования для растягивающих отображений, введенную в п. 7.3.1, с несколько иной точки зрения. Будем считать, что двоичное представление числа из отрезка $[0, 1]$ является записью результатов бесконечной последовательности подбрасываний монеты: мы пишем нуль, если монета выпадает гербом вверх, и один — если решкой. Если монета *идеальная*, т.е. орел и решка выпадают равновероятно, и наши испытания *независимы*, то вероятность любой последовательности нулей и единиц в первых n испытаниях равна 2^{-n} . Мы называем появление любой

фиксированной последовательности орлов и решек в n испытаниях *элементарным событием*. Вероятность любого исхода после n испытаний (не обязательно первых или даже последовательных) равна полной длине системы двоичных отрезков, которые соответствуют всем элементарным событиям, совместимым с данным исходом. Таким образом, вычисления длин из § 7.5 имеют вероятностную интерпретацию. Например, чтобы найти вероятность выпадения k решек в n испытаниях, надо взять число последовательностей нулей и единиц длины n с k единицами, которое равно $\binom{n}{k}$, и разделить на 2^n . Подобным образом, из равенства (7.5.1) мы находим число испытаний, для которых отклонение среднего числа выпадений решки от $1/2$ не больше чем ϵ . Мы даже можем придать новую интерпретацию вычислениям усреднений Биркгофа характеристической функции φ для отображения E_2 , приведенным в § 7.5, и получим оценки вероятностей различных событий в модели с подбрасыванием монеты: вероятность того, что среднее число выпадений решки отличается от $1/2$ больше чем на произвольное заданное число, сходится к нулю экспоненциально, если число испытаний стремится к бесконечности.

В п. 7.3.5 объясняется, как кодирование отображения приводит к сдвигу на пространстве последовательностей. Кодирование для отображения E_2 с использованием разбиения на левую и правую половины единичного отрезка соответствует в точности модели с подбрасыванием монеты; оно также проводится на пространстве последовательностей из нулей и единиц.

7.6.2. Схемы Бернулли. Более общая схема случайных испытаний с несколькими возможными исходами (не обязательно равновероятными или независимыми) также может описываться в терминах сдвига на пространстве символов. Вероятности элементарных событий могут быть не равны. Элементы этого пространства соответствуют возможным исходам бесконечных последовательностей испытаний, и преобразование сдвига соответствует одному шагу вперед. Самый простой класс таких схем — *стационарная схема Бернулли*. Зафиксируем вероятности $p_0, \dots, p_{N-1}$ символов $0, \dots, N-1$, встречающихся в испытаниях, и предположим, что последовательные испытания взаимно независимы, т. е. вероятность каждого исхода любого испытания не зависит от результатов предыдущего. Это означает, что вероятность любой конечной последовательности исходов равна произведению вероятностей исходов из этой последовательности.

Рассмотрим случай $N = 2$, где $p := p_0 \neq p_1 = 1 - p$, т. е. подбрасывание деформированной монеты, и проведем вычисления подобно тому, как это было сделано в § 7.5. Вероятность элементарного события, при котором появляется k нулей и $n - k$ единиц, равна $\binom{n}{k} p^k (1 - p)^{n-k}$. Таким образом, если φ — характеристическая функция множества последовательностей, первый элемент которых равен 0, тогда среднее по пространству функции φ равно p и для большинства конечных последовательностей доля вхождений нуля должна быть приблизительно равна p . В этом можно убедиться, если проделать такие же выкладки, как в § 7.5, применительно к нашей ситуации.

Предложение 7.6.1. Для стационарной схемы Бернулли с весами p и $1 - p$

$$P\{i: |\mathcal{B}_n(\varphi)(x) - p| < \varepsilon \text{ для } x \in \Delta_n^i\} \rightarrow 1$$

экспоненциально при $n \rightarrow \infty$.

Доказательство. Соотношение (7.5.1) заменяется на равенство

$$P\{i: |\mathcal{B}_n(\varphi)(x) - p| < \varepsilon \text{ для } x \in \Delta_n^i\} = \sum_{k=\lfloor (p-\varepsilon)n \rfloor + 1}^{\lfloor (p+\varepsilon)n \rfloor} \binom{n}{k} p^k (1-p)^{n-k}. \quad (7.6.1)$$

Чтобы получить оценку снизу, оценим сверху сумму оставшихся биномиальных коэффициентов. Одна часть этой суммы имеет вид

$$\sum_{k=0}^{\lfloor (p-\varepsilon)n \rfloor} \binom{n}{k} p^k (1-p)^{n-k},$$

а оценку для соответствующей суммы, начинающейся с номера $\lfloor (p+\varepsilon)n \rfloor + 1$, предлагается получить читателю. Для достаточно большого n последний член является самым большим, потому что следующий член получается из предыдущего умножением на $\frac{n-k}{k} \frac{p}{1-p}$, а это выражение больше 1 при $k < pn$. Как и в § 7.5, это приводит к оценке сверху

$$\begin{aligned} \sum_{k=0}^{\lfloor (p-\varepsilon)n \rfloor} \binom{n}{k} p^k (1-p)^{n-k} &\leq (l+1) \binom{n}{l} p^l (1-p)^{n-l} \asymp \\ &\asymp (l+1) e^{n \ln n - l \ln l - (n-l) \ln(n-l)} \sqrt{\frac{n}{2\pi l(n-l)}} p^l (1-p)^{n-l} = \\ &= e^{n \ln n - l \ln l - (n-l) \ln(n-l) + l \ln p + (n-l) \ln(1-p)} (l+1) \sqrt{\frac{n}{2\pi l(n-l)}}. \end{aligned}$$

Показатель степени

$$\begin{aligned} (n-l)[\ln n - \ln(n-l) + \ln(1-p)] + l[\ln n - \ln l + \ln p] &= \\ = (n-l) \ln \frac{n(1-p)}{n-l} + l \ln \frac{np}{l} &= \\ = (n-l) \ln \left(1 + \frac{n(1-p) - n + l}{n-l}\right) + l \ln \left(1 + \frac{np-l}{l}\right) \end{aligned}$$

можно оценить с использованием свойства выпуклости логарифма: для $l = \lfloor (p-\varepsilon)n \rfloor$ и $\varepsilon > 0$ имеется такое $\delta > 0$, что

$$\begin{aligned} (n-l) \ln \left(1 + \frac{n(1-p) - n + l}{n-l}\right) + l \ln \left(1 + \frac{np-l}{l}\right) &\leq \\ \leq (n-l) \left(\frac{n(1-p) - n + l}{n-l} - \delta\right) + l \left(\frac{np-l}{l} - \delta\right) &= -n\delta. \end{aligned}$$

Таким образом, показатель степени отрицательный, и верхняя граница стремится к нулю экспоненциально при $n \rightarrow \infty$.

Проводя аналогичную оценку оставшейся части суммы, мы получаем, что правая часть равенства (7.6.1) стремится к 1 экспоненциально, как и требуется. $\square$

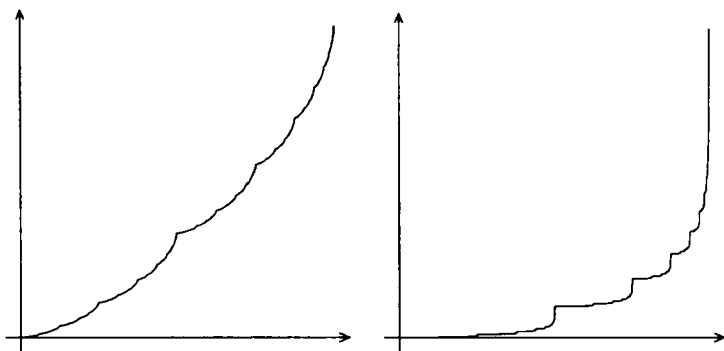
7.6.3. Энтропия. Элементарные события, которые составляют большую часть возможных исходов, т. е. события, среднее число появлений которых близко к среднему по пространству, не являются наиболее вероятными сами по себе, но они имеют определенный экспоненциальный размер. Для любого такого события S логарифм вероятности этого события, разделенный на n , приблизительно равен $p \ln p + (1 - p) \ln(1 - p)$. Эта величина называется *энтропией* распределения вероятности $(p, 1 - p)$ и тесно связана со степенью неопределенности, порожденной случайностью схемы. Вышеупомянутая экспоненциальная асимптотика для вероятности типичных исходов — наиболее элементарный случай знаменитой теоремы Шеннона—Макмиллана, которая является одним из краеугольных камней теории информации³.

7.6.4. Меры Бернулли на отрезке. Возвращаясь к растягивающему отображению E_2 , мы можем интерпретировать вероятности элементарных событий как аналог длин соответствующих двоичных отрезков. Чтобы отличать эти вероятности от длин, мы будем называть вероятность *мерой Бернулли* соответствующего отрезка. Например, мера отрезка $[0, 1/4]$ равна $m_p([0, 1/4]) = p^2$, и $m_p([1/4, 1/2]) = m_p([1/2, 3/4]) = p(1 - p)$, $m_p([3/4, 1]) = (1 - p)^2$. В качестве меры объединения двоичных отрезков естественно взять сумму мер слагаемых. Нетрудно проверить, что любой способ приближения отрезка, интервала или полуинтервала объединениями двоичных отрезков изнутри или извне приводит к одному и тому же пределу соответствующих мер, так что мы получаем меру $m_p(I)$ любого отрезка, интервала или полуинтервала I и, следовательно, их конечных объединений. Как и при определении обычного интеграла, мы можем использовать эту меру для определения интеграла любой непрерывной функции (на самом деле и некоторых других функций, например функций с конечным числом разрывов). Как и прежде, *множество меры ноль* — это такое множество, что для заданного $\epsilon > 0$ имеется счетное покрытие интервалами, сумма мер Бернулли которых меньше ϵ .

Так же как и в определении 4.2.2, удобно представить эту меру ее *функцией распределения* $\Psi_p(x) = m_p([0, x])$. Очевидно, что $\Psi_p(0) = 0$, $\Psi_p(1) = 1$ и функция Ψ_p неубывающая. На самом деле она возрастает, так как каждый двоичный отрезок имеет положительную меру и между любыми двумя числами найдется двоичный отрезок. Аналогично эта функция непрерывна для $p \notin \{0, 1\}$, потому что меры коротких двоичных отрезков малы. Ясно, что $\Psi_{1/2}(x) = x$, но, оказывается, для $p \neq 1/2$ функция Ψ_p не дифференцируема во многих точках. Однако она имеет интересное свойство самоподобия: $\Psi_p(x/2) = p\Psi_p(x)$. На рис. 7.6.1 изображены графики функций $\Psi_{1/3}$ и $\Psi_{1/10}$.

7.6.5. Перемешивание. Точно так же, как топологическое перемешивание является более сильным свойством, чем топологическая транзитивность, имеется более сильное свойство, соответствующее равномерному распределению (или

³См., например, теорему 2.3 в книге: K. Petersen. Ergodic theory. Cambridge: Cambridge University Press, 1983. (Cambridge Studies in Advanced Mathematics, v. 2).

Рис. 7.6.1. Функции $\Psi_{1/3}$ и $\Psi_{1/10}$

закону больших чисел), которое называется *перемешиванием*. Мы рассмотрим это понятие в конкретной ситуации кусочно непрерывных отображений окружности.

Для этого начнем с функции распределения: предположим, что $\Psi: [0, 1] \rightarrow [0, 1]$ — такая непрерывная неубывающая функция, что $\Psi(0) = 0$, и определим меру отрезка $[a, b]$ соотношением $m([a, b]) := \Psi(b) - \Psi(a)$. Это определение продолжается на конечные объединения отрезков по аддитивности. В частности, если мы представим окружность как отрезок $[0, 1]$ с отождествленными концами, тогда мера дуги, содержащей 0, определяется как сумма мер двух частей с обеих сторон от 0. Меру открытой или полуоткрытой дуги будем считать равной мере ее замыкания.

Рассмотрим кусочно непрерывное кусочно монотонное отображение окружности $f: S^1 \rightarrow S^1$ (или отрезка). Это понятие было введено в определении 4.2.2, где мы отметили, что отображение $m_f(I) := m(f^{-1}(I))$ корректно определено. Если $m_f = m$, тогда мы говорим, что мера m *инварианта* относительно f . Это не первый случай сохранения меры; в § 6.1 мы отметили, что ньютоновы системы сохраняют объем (см. следствие 6.2.3), который является естественной мерой в многомерном случае.

Определение 7.6.2. Рассмотрим кусочно непрерывное кусочно монотонное отображение $f: S^1 \rightarrow S^1$ и предположим, что мера m инвариантна относительно f . Тогда отображение f называется *перемешивающим* (относительно m), если $m(\Delta_1 \cap f^{-n}(\Delta_2)) \xrightarrow{n \rightarrow \infty} m(\Delta_1) \cdot m(\Delta_2)$ для любых двух дуг Δ_1, Δ_2 .

Поскольку $\Delta_1 \cap f^{-n}(\Delta_2)$ есть конечное объединение отрезков, мера определена корректно.

Предложение 7.6.3. Если f — перемешивающее отображение относительно меры m , функция распределения которой возрастает, то оно является топологически перемешивающим.

Доказательство. Из предположения относительно меры m следует, что всякий отрезок имеет положительную меру. Если U и V — открытые множества,

тогда найдутся отрезки $\Delta_1 \subset V$, $\Delta_2 \subset U$. Поскольку отображение f перемешивающее относительно m , найдется такое $N \in \mathbb{N}$, что $m(\Delta_1 \cap f^{-n}(\Delta_2)) > 0$ и, следовательно, $\Delta_1 \cap f^{-n}(\Delta_2) \neq \emptyset$ для $n \geq N$. Применяя отображение f^n , мы получаем, что $\Delta_1 \cap f^n(\Delta_2) \neq \emptyset$ для всех $n \geq N$, следовательно, $V \cap f^n(U) \neq \emptyset$ для $n \geq N$. $\square$

Предложение 7.6.4. Мера Бернулли m_p на S^1 из п. 7.6.4 инвариантна относительно E_2 , и отображение E_2 является перемешивающим относительно m_p .

Доказательство. Для доказательства инвариантности заметим, что любая дуга может быть сколь угодно близко приближена объединением непересекающихся двоичных отрезков. Поэтому достаточно проверить, что всякий двоичный отрезок имеет ту же меру, что и его прообраз. Двоичный отрезок Δ определен конечной строкой $0, x_1 x_2 \dots x_n$ двоичных разрядов. Обозначим $\mu(0) = p$ и $\mu(1) = 1 - p$, тогда $m(\Delta) = \prod_{i=1}^n \mu(x_i)$. Прообраз состоит из двух двоичных отрезков с двоичными строками $0, 0x_1 x_2 \dots x_n$ и $0, 1x_1 x_2 \dots x_n$, сумма мер которых равна

$$pm(\Delta) + (1 - p)m(\Delta) = m(\Delta).$$

Свойство перемешивания достаточно также проверить только для двоичных отрезков. Предположим, что Δ_1 и Δ_2 — двоичные отрезки, заданные строками $0, \alpha_1 \dots \alpha_m$ и $0, \omega_1 \dots \omega_r$ соответственно. Тогда $E_2^{-n}(\Delta_2)$ — непересекающееся объединение 2^n двоичных отрезков, определенных строками $0, x_1 \dots x_n \omega_1 \dots \omega_r$ со всеми возможными комбинациями $x_1 \dots x_n$. Если $n > m$, тогда множество $\Delta_1 \cap E_2^{-n}(\Delta_2)$ представляет собой непересекающееся объединение двоичных отрезков со всеми возможными строками $0, \alpha_1 \dots \alpha_m x_{m+1} \dots x_n \omega_1 \dots \omega_r$. Его мера равна

$$\sum_{x_1 \dots x_n} \prod_{i=1}^m \mu(\alpha_i) \prod_{j=1}^{n-m} \mu(x_{m+j}) \prod_{k=1}^r \mu(\omega_k) = m(\Delta_1) m(\Delta_2) \sum_{x_1 \dots x_n} \prod_{j=1}^{n-m} \mu(x_{m+j}) = m(\Delta_1) m(\Delta_2),$$

поскольку последняя сумма равна сумме мер всех двоичных отрезков для строк длины $n - m$, а следовательно, мере окружности, т. е. 1. $\square$

Две другие известные хаотические системы, которые мы рассмотрели, также перемешивающие; это двусторонний сдвиг и гиперболический линейный автоморфизм тора. Конечно, это утверждение требует определений перемешивания и меры в этом контексте.

Предложение 7.6.5. Двусторонний сдвиг с мерой Бернулли, определенной так же, как в п. 7.6.4, является перемешивающим, где свойство перемешивания понимается так же, как и в определении 7.6.2, только дуги заменяются на цилиндры.

При доказательстве используется такое же рассуждение, как и в случае отображения E_2 , и оно предлагается в качестве упражнения (упражнение 7.6.1).

Предложение 7.6.6. *Гиперболический автоморфизм F двумерного тора T^2 , индуцированный линейным отображением L с матрицей $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, является перемешивающим относительно площади (меры Лебега), где перемешивание понимается так же, как и в определении 7.6.2, только дуги заменяются на параллелограммы.*

Доказательство. Поскольку отображение F обратимо, мы можем заменить условие $m(\Delta_1 \cap F^{-n}(\Delta_2)) \xrightarrow{n \rightarrow \infty} m(\Delta_1) \cdot m(\Delta_2)$ условием $m(B \cap F^n(A)) \xrightarrow{n \rightarrow \infty} m(A) \cdot m(B)$. Для удобства мы будем использовать в качестве «испытательных множеств» частные виды параллелограммов вместо дуг, применявшихся в случае отображения E_2 . Вместо дуги A мы используем параллелограмм A , стороны которого параллельны собственным направлениям; обозначим длину сторон вдоль растягивающей собственной прямой через a_1 , а длину других сторон через a_2 . Вместо дуги B мы используем параллелограмм B с двумя вертикальными сторонами длины b_2 и двумя сторонами длины b_1 , параллельными собственному направлению для собственного значения λ , большего чем единица. Обозначим косинус угла с горизонтальным направлением через c . Рассмотрим теперь параллелограмм $F^n(A)$ для некоторого большого n . Его нижняя сторона длины a_1 отображается в отрезок длины $\lambda^n a_1$, который пересекает вертикальную сторону параллелограмма B приблизительно в $cb_2 \lambda^n a_1$ точках, потому что точки пересечения — это образы под действием отображения последования для линейного потока вдоль направления собственного вектора, и это отображение последования — иррациональный поворот, который имеет свойство равномерного распределения. Чтобы определить меру пересечения $F^n(A) \cap B$, заметим, что оно состоит (кроме, быть может, двух полос вблизи ребер параллелограмма B и двух частей, не проходящих полностью через B) из такого числа полос ширины $\lambda^{-n} a_2$ и длины b_1 , что их общая площадь составляет $(c \lambda^n a_1 b_2)(\lambda^{-n} a_2) b_1 = (a_1 a_2)(c b_1 b_2)$, т. е. в точности равна $m(A)m(B)$ (см. рис. 7.6.2). $\square$

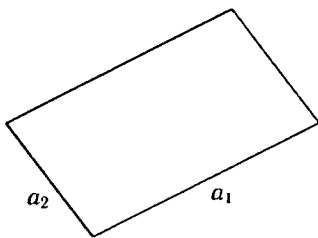


Рис. 7.6.2. Перемешивание для автоморфизма тора

Это рассуждение применимо для любого линейного автоморфизма тора.

Предложение 7.6.7. *Любой гиперболический линейный автоморфизм тора является перемешивающим относительно площади как меры.*

Упражнения

7.6.1. Покажите, что сдвиг σ_2 является перемешивающим относительно любой меры Бернулли.

7.6.2. Оцените вторую часть суммы в доказательстве предложения 7.6.1.

7.6.3. Покажите, что если $0 < p < q < 1$, то имеется такое множество A , что A является множеством меры нуль для m_p , а его дополнение — множеством меры нуль для m_q .

7.6.4. Докажите утверждение, аналогичное теореме 7.5.6, для меры Бернулли m_p .

ЭНТРОПИЯ И ХАОС

В этой главе мы изучаем два взаимосвязанных понятия, которые являются важными параметрами хаотических динамических систем. Первое — это фрактальная размерность множества. Если допустить, что размерность пространства может не являться целым числом, то топологическое понятие размерности можно обобщить, скажем, на канторовы множества. Хотя все канторовы множества гомеоморфны, они могут выглядеть более «толстыми» или более «тонкими» в зависимости от параметров, участвующих в их построении. Фрактальная размерность является критерием «толщины» этих множеств. Если данное канторово множество возникает как инвариантное множество гиперболической динамической системы, его размерность тесно связана с другими динамически важными величинами, особенно с коэффициентами сжатия и растяжения в системе. Эта связь является предметом активных исследований, и мы проиллюстрируем ее на примере подковы Смейла.

Другое понятие — энтропия. Она служит для измерения глобальной сложности орбит в экспоненциальном масштабе и тесно связана с коэффициентом экспоненциального роста периодических точек и коэффициентами растяжения и сжатия. Как инвариант топологического сопряжения, она также позволяет различать не сопряженные динамические системы.

Значения размерности и энтропии инвариантного множества динамической системы связаны, а также связаны конструкции, используемые для их определения. В основе каждого из этих определений лежит понятие *емкости* множества, с которого мы начинаем данную главу.

§ 8.1. Размерность компактного пространства

8.1.1. Емкость. Для компактного метрического пространства имеется понятие «размера» или «емкости», которое обобщает понятие объема. Предположим, что X — компактное пространство с метрикой d . Тогда множество $E \subset X$ называется r -плотным, если $X \subset \bigcup_{x \in E} B_d(x, r)$, где $B_d(x, r)$ — шар радиуса r относительно метрики d с центром в точке x (см. п. 2.6.1). Определим r -емкость пространства (X, d) как минимальное число элементов $S_d(r)$ в его r -плотном множестве.

Например, если X — это отрезок $[0, 1]$ с обычной метрикой, тогда значение $S_d(r)$ приблизительно равно $1/(2r)$, потому что необходимо $1/(2r)$ шаров (т.е. интервалов), чтобы покрыть единичный отрезок, и $\lfloor 2 + 1/(2r) \rfloor$ шаров с центрами в точках $ir(2 - r)$, $0 \leq i \leq \lfloor 1 + 1/(2r) \rfloor$, вполне годятся для покрытия. В качестве другого примера возьмем единичный квадрат $X = [0, 1]^2$. Тогда значение $S_d(r)$ имеет порядок r^{-2} , потому что требуется по крайней мере $1/(\pi r^2)$ шаров радиуса r , чтобы покрыть единичный квадрат, а с другой стороны, $(1 + 1/r)^2$ шаров с центрами в точках (ir, jr) служат таким покрытием. Аналогично $(1 + 1/r)^3$ шаров радиуса r покрывают единичный куб.

В случае трюичного канторова множества с обычной метрикой мы можем немного упростить себе задачу и использовать замкнутые шары. Тогда мы получим, что $S_d(3^{-i}) = 2^i$. Если же использовать обыкновенные открытые шары, то $S_d((3 - 1/i)^{-i}) = 2^i$.

8.1.2. Энтропийная размерность. Один интересный аспект емкости — соотношение между ее зависимостью от r (т. е. скорость степенного роста $S_d(r)$ как функции от r) и размерностью.

Если $X = [0, 1]$, то

$$\lim_{r \rightarrow 0} \left(-\frac{\ln S_d(r)}{\ln r} \right) \geq \lim_{r \rightarrow 0} \left(-\frac{\ln(1/2r)}{\ln r} \right) = \lim_{r \rightarrow 0} \frac{\ln 2 + \ln r}{\ln r} = 1$$

и

$$\lim_{r \rightarrow 0} \left(-\frac{\ln S_d(r)}{\ln r} \right) \leq \lim_{r \rightarrow 0} \left(-\frac{\ln[2 + 1/2r]}{\ln r} \right) \leq \lim_{r \rightarrow 0} \left(-\frac{\ln(1/r)}{\ln r} \right) = 1,$$

так что $\lim_{r \rightarrow 0} (-\ln S_d(r)/\ln r) = 1 = \dim X$. Если $X = [0, 1]^2$, то $\lim_{r \rightarrow 0} (-\ln S_d(r)/\ln r) = 2 = \dim X$, и если $X = [0, 1]^3$, то $\lim_{r \rightarrow 0} (-\ln S_d(r)/\ln r) = 3 = \dim X$. Из этих примеров видно, что мы можем трактовать число $\lim_{r \rightarrow 0} (-\ln S_d(r)/\ln r)$ как размерность.

Определение 8.1.1. Если X — вполне ограниченное метрическое пространство (см. определение А.1.20), тогда число $\text{bdim}(X) := \lim_{r \rightarrow 0} \left(-\frac{\ln S_d(r)}{\ln r} \right)$ называется *энтропийной размерностью* пространства X ¹.

8.1.3. Примеры. Попробуем найти энтропийную размерность для менее тривиальных пространств.

1. *Троичное канторово множество.* Если C — троичное канторово множество, то

$$\text{bdim}(C) = \lim_{r \rightarrow 0} \left(-\frac{\ln S_d(r)}{\ln r} \right) = \lim_{n \rightarrow \infty} -\frac{\ln 2^n}{\ln 3^{-n}} = \frac{\ln 2}{\ln 3}.$$

Если множество C_α строится удалением среднего интервала относительного размера $1 - (2/\alpha)$ на каждом шаге, тогда $\text{bdim}(C_\alpha) = \ln 2 / \ln \alpha$. Эта величина возрастает и стремится к 1 при $\alpha \rightarrow 2$ (так как мы удаляем все меньшие интервалы) и убывает к 0 при $\alpha \rightarrow \infty$ (так как мы удаляем все большие интервалы). Таким образом, если при построении канторова множества размер остающихся отрезков быстро уменьшается с каждой итерацией, то энтропийная размерность мала.

Отсюда, кстати, следует, что энтропийная размерность множества может изменяться при гомеоморфизме, потому что рассмотренные канторы множества попарно гомеоморфны.

2. *Ковер Серпинского.* Несложно рассмотреть и другие множества, подобные канторову, например ковер Серпинского S из п. 2.7.2. Для квадратного ковра Серпинского мы можем, как и при вычислении емкости троичного канторова множества, упростить себе задачу и использовать для покрытия замкнутые шары (имеющие общий центр с одним из остающихся малых кубов на некотором шаге). Тогда $S_d(3^{-i}/\sqrt{2}) = 8^i$ и

$$\text{bdim}(S) = \lim_{n \rightarrow \infty} \left(-\frac{\ln 8^i}{\ln 3^{-i}/\sqrt{2}} \right) = \frac{\ln 8}{\ln 3} = \frac{3 \ln 2}{\ln 3},$$

а это число равно утроенной размерности троичного канторова множества (но все еще меньше 2). Для треугольного ковра Серпинского мы подобным образом получаем, что его энтропийная размерность равна $\ln 3 / \ln 2$.

3. *Снежинка Коха.* Для снежинки Коха K из п. 2.7.2 мы имеем $S_d(3^{-i}) = 4^i$, так как ее можно покрыть (замкнутыми) шарами с центрами на ребрах i -го многоугольника. Таким образом,

$$\text{bdim}(K) = \lim_{n \rightarrow \infty} \left(-\frac{\ln 4^i}{\ln 3^{-i}} \right) = \frac{\ln 4}{\ln 3} = \frac{2 \ln 2}{\ln 3},$$

причем это число меньше размерности ковра Серпинского. Это соответствует тому, что итерации для снежинки Коха выглядят более «тонкими». Заметим, однако, что эта размерность больше 1, т. е. больше размерности кривой. Во всех этих примерах размерность — дробное число. Такие множества называются *фракталами*.

¹В англоязычной литературе для энтропийной размерности используют термин «box dimension» или «box-counting dimension».

4. *Подкова Смейла.* Предположим (без ограничения общности), что в построении подковы Смейла (см. п. 7.4.4) коэффициент растяжения на линейных частях равен $\lambda > 2$, а коэффициент сжатия равен $\mu < 1/\lambda$. Для данного числа $n \in \mathbb{N}$ инвариантное множество $\Lambda = \bigcap_{n=-\infty}^{\infty} f^{-n}(\Delta)$ содержится в множестве $\Lambda = \bigcap_{i=-n}^n f^{-i}(\Delta)$, которое состоит из 4^n прямоугольников со сторонами λ^{-n} и μ^n и поэтому может быть покрыто примерно $4^n/(\lambda^n \mu^n)$ квадратами со сторонами μ^n . Таким образом, $S_d(\mu^{-n}) \asymp 4^n/(\lambda^n \mu^n)$, и

$$\text{bdim}(\Lambda) = \lim_{n \rightarrow \infty} \left(-\frac{\ln S_d(\mu^{-n})}{\ln \mu^{-n}} \right) = \lim_{n \rightarrow \infty} \left(-\frac{n(\ln 4 - \ln \lambda - \ln \mu)}{n \ln \mu} \right) = 1 + \frac{\ln 4 - \ln \lambda}{-\ln \mu}.$$

5. *Пространства последовательностей.* Рассмотрим пространство двусторонних последовательностей Ω_N с метрикой d_λ , определенной равенством (7.3.4). Согласно соотношению (7.3.5) имеется покрытие N^{2n-1} непересекающимися шарами радиуса λ^{1-n} , а именно цилиндрами $C_{\alpha_1 \dots \alpha_{n-1}} = \{\omega \in \Omega_N : \omega_i = \alpha_i \text{ для } |i| < n\}$. Поэтому $S_{d_\lambda}(\lambda^{1-n}) = N^{2n-1}$ и, следовательно, энтропийная размерность равна

$$\text{bdim}(\Omega_N, d_\lambda) = \lim_{r \rightarrow 0} \left(-\frac{\ln S_d(r)}{\ln r} \right) = \lim_{n \rightarrow \infty} \left(-\frac{\ln N^{2n-1}}{\ln \lambda^{1-n}} \right) = \lim_{n \rightarrow \infty} \frac{2n-1}{n-1} \frac{\ln N}{\ln \lambda} = 2 \frac{\ln N}{\ln \lambda}.$$

Как и в случае канторова множества, энтропийная размерность уменьшается при увеличении λ , что соответствует быстрому уменьшению радиуса цилиндров (как функции длины строки) для больших λ .

8.1.4. **Зависимость от метрики.** Еще один вопрос, связанный с понятием емкости $S_d(r)$, — это ее зависимость от метрики при заданном r . Если заменить метрику большей (так сказать, увеличить разрешающую способность), тогда шары станут меньше и, следовательно, $S_d(r)$ увеличится. Скорость, с которой это происходит, — новая мера разрешающей способности метрики. Простой пример — умножение метрики на положительное число a . Ясно, что $S_{ad}(ar) = S_d(r)$ и

$$\lim_{r \rightarrow 0} \left(-\frac{\ln S_{ad}(r)}{\ln(ar)} \right) = \lim_{r \rightarrow 0} \left(-\frac{\ln S_{ad}(ar)}{\ln(ar)} \right) = \lim_{r \rightarrow 0} \left(-\frac{\ln S_d(r)}{\ln(ar)} \right) = \lim_{r \rightarrow 0} \left(-\frac{\ln S_d(r)}{\ln a + \ln r} \right) = \lim_{r \rightarrow 0} \left(-\frac{\ln S_d(r)}{\ln r} \right).$$

Таким образом, умножение метрики на постоянное число не меняет энтропийной размерности. Однако можно исследовать асимптотическое поведение емкости $S_{d_i}(r)$ для последовательности d_i метрик при $i \rightarrow \infty$ для фиксированного r . Мы сделаем это при изучении энтропии.

Упражнения

8.1.1. Докажите, что число элементов в минимальном покрытии не всегда совпадает с минимальным числом элементов в покрытии.

8.1.2. Вычислите энтропийную размерность множества $\mathbb{Q} \cap [0, 1]$.

8.1.3. Для подковы Смейла покажите, что $0 < \text{bdim}(\Lambda) < 2$.

8.1.4. Для подковы в форме буквы S с тремя пересечениями вычислите энтропийную размерность инвариантного множества и докажите, что она заключена в границах между 0 и 2.

8.1.5. Найдите размерность метрики d_λ'' , введенной равенством (7.3.9), на Ω_N и Ω_N^R .

8.1.6. Покажите, что размерность $\text{bdim}(\Omega_N^R, d_\lambda)$ пространства односторонних последовательностей Ω_N^R с метрикой d_λ равна $\ln N / \ln \lambda$.

8.1.7. Покажите, что энтропийная размерность треугольного ковра Серпинского равна $\ln 3 / \ln 2$.

8.1.8. Постройте на отрезке канторовы множества, имеющие энтропийные размерности 0 и 1.

8.1.9. Определите энтропийную размерность множества точек из отрезка $[0, 1]$, которые имеют двоичное разложение без последовательных нулей.

§ 8.2. Топологическая энтропия

8.2.1. Критерии сложности и инварианты. Мы рассмотрели несколько характеристик сложности динамической системы, включая топологическую транзитивность, минимальность, плотность множества периодических точек, хаос и топологическое перемешивание. В частности, топологическое перемешивание служит признаком того, что имеется переплетение и разбегание различных орбит. Это все качественные («да-нет») показатели сложности. Единственной количественной мерой сложности до сих пор являлась экспоненциальная скорость роста периодических орбит. Хотя простые рациональные повороты имеют бесконечно много периодических точек, для хаотических примеров характерен экспоненциальный рост конечного количества периодических точек.

1. *Энтропия.* Вслед за ростом периодических орбит можно в некотором смысле измерить и рост всех орбит. Это делается с помощью наиболее важного численного инварианта, связанного с ростом орбит, — топологической энтропии. Она представляет собой скорость экспоненциального роста числа отрезков орбит, различимых с произвольно хорошей, но конечной точностью. Можно сказать, что топологическая энтропия грубым, но убедительным образом описывает полную экспоненциальную сложность орбитальной структуры единственным числом. Действительно, мы увидим, что для изученных нами хаотических систем характерно наличие положительной энтропии и топологическая энтропия не меньше чем скорость роста периодических орбит. Поэтому естественно рассматривать энтропию как количественный критерий хаоса в динамической системе.

2. *Инварианты.* Пришло время объяснить, для чего необходимо изучение инвариантов динамических систем. Инварианты — величины, связанные с динамической системой, которые совпадают для двух динамических систем, эквивалентных в смысле сопряжения (см. определение 7.3.3). При исследовании новой динамической системы естественно задаться вопросом, является ли она эквивалентной какой-либо ранее изученной системе, что сэкономило бы много сил; или же можно попытаться установить, являются ли некоторые семейства динамических систем попарно эквивалентными или можно ли их разбить на классы эквивалентности (относительно топологического сопряжения). В любом случае необходимо знать, имеется ли сопряжение между двумя данными системами. Если после многих попыток найти сопряжение все же не удастся, очевидно, что нужно найти способ показать, что никакого сопряжения нет. Это можно сделать с помощью инвариантов: если одна система транзитивна, а другая — нет, тогда они не могут быть сопряжены. Если один гомеоморфизм окружности имеет число вращения α , а другой имеет число вращения $\beta \neq \pm\alpha$, тогда эти два гомеоморфизма

топологически не сопряжены. Энтропию также удобно использовать в качестве инварианта (см. следствие 8.2.3) по той причине, что она имеет действительные значения (в противоположность «да-нет» критериям), а следовательно, позволяет более точно различать динамические системы, чем транзитивность, перемешивание и т. д. С другой стороны, она определена для весьма широкого класса динамических систем, в отличие, например, от числа вращения, определенного только для отображений окружности.

8.2.2. Первое определение энтропии. Чтобы определить энтропию, мы измеряем скорость роста емкости $S_d(r)$ для фиксированного r по мере динамически значимого увеличения метрики. Этим энтропия отличается от энтропийной размерности, при определении которой мы исследовали изменение емкости как функции от r для фиксированной метрики. Предположим, что $f: X \rightarrow X$ — непрерывное отображение компактного метрического пространства X с метрикой d , и определим следующую возрастающую последовательность метрик d_n^f , $n = 1, 2, \dots$, начинающуюся с $d_1^f = d$:

$$d_n^f(x, y) = \max_{0 \leq i \leq n-1} d(f^i(x), f^i(y)). \quad (8.2.1)$$

Другими словами, d_n^f — расстояние между отрезками орбит $\mathcal{O}_n(x) = \{x, \dots, f^{n-1}x\}$ и $\mathcal{O}_n(y)$. Обозначим открытый шар $\{y \in X: d_n^f(x, y) < r\}$ через $B_f(x, r, n)$.

Определение 8.2.1. Пусть $S_d(f, r, n)$ обозначает r -емкость метрики d_n^f . Точнее говоря, множество $E \subset X$ называется r -плотным относительно d_n^f , или (n, r) -плотным, если $X \subset \bigcup_{x \in E} B_f(x, r, n)$. В этом случае $S_d(f, r, n)$ — минимальное число элементов (n, r) -плотного множества. Это минимальное число начальных условий, поведение которых вплоть до момента времени n аппроксимирует поведение *любого* начального состояния до момента r . Рассмотрим скорость экспоненциального роста

$$h_d(f, r) := \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \ln S_d(f, r, n) \quad (8.2.2)$$

последовательности $S_d(f, r, n)$. Очевидно, что $h_d(f, r)$ не уменьшается с уменьшением r , так что мы можем определить

$$h_d(f) := \lim_{r \rightarrow 0} h_d(f, r). \quad (8.2.3)$$

Назовем $h(f) := h_{\text{top}}(f) := h_d(f)$ *топологической энтропией* отображения f .

Заметим, что мы берем двойной предел: сначала по n и затем по r . Изучение предела по n наиболее важно, поскольку при переходе к этому пределу мы встречаемся с динамикой. Во многих интересных случаях переход к пределу по r на самом деле тривиален, потому что $h_d(f, r)$ не зависит от r (для малых r).

Значение $h_d(f)$ а priori может зависеть от метрики d . На самом деле это не так, если метрика меняется на гомеоморфную (см. определение A.1.17). Поэтому мы можем опустить зависимость от метрики в соотношении (8.2.3).

Предложение 8.2.2. Если d' — метрика на X , эквивалентная метрике d , то $h_{d'}(f) = h_d(f)$.

Доказательство. Тожественное отображение $\text{Id}: (X, d) \rightarrow (X, d')$ является гомеоморфизмом по предположению и равномерно непрерывно в обоих направлениях в силу компактности пространства X . Таким образом, для $r > 0$ существует такое $\delta(r) > 0$, что если $d'(x_1, x_2) < \delta$, то $d(x_1, x_2) < r$, т. е. любой δ -шар в метрике d' содержится в r -шаре в метрике d . Согласно (8.2.1) это также справедливо для d_n^I и d_n^I . Таким образом, $S_{d'}(f, \delta, n) \geq S_d(f, r, n)$ для каждого n , так что $h_{d'}(f, \delta) \geq h_d(f, r)$ и $h_{d'}(f) \geq \lim_{\delta \rightarrow 0} h_{d'}(f, \delta) \geq \lim_{r \rightarrow 0} h_d(f, r) = h_d(f)$. При замене d на d' мы получаем $h_d(f) \geq h_{d'}(f)$, а следовательно, справедливо равенство. $\square$

Следствие 8.2.3. Топологическая энтропия инвариантна относительно топологического сопряжения.

Доказательство. Предположим, что отображения $f: X \rightarrow X$ и $g: Y \rightarrow Y$ топологически сопряжены посредством гомеоморфизма $h: X \rightarrow Y$ (см. определение 7.3.3). Зафиксируем метрику d на X и определим d' на Y как образ метрики d , т. е. $d'(y_1, y_2) = d(h^{-1}(y_1), h^{-1}(y_2))$ (см. п. 2.6.1). Тогда гомеоморфизм h становится изометрией, так что $h_d(f) = h_{d'}(g)$. $\square$

8.2.3. Субэкспоненциальный рост. В качестве первого примера применения этого понятия рассмотрим ситуации с относительно простой динамикой.

Предложение 8.2.4. Топологическая энтропия сжатий и изометрий равна нулю. В частности, любой сдвиг T_γ тора имеет нулевую энтропию.

Доказательство. Если X — компактное метрическое пространство и отображение $f: X \rightarrow X$ удовлетворяет условию Липшица с константой 1, тогда $d_n^I = d$ для всех n и, следовательно, $S_n(f, r, n)$ не зависит от n , так что $h(f) = 0$. $\square$

Это отсутствие какого-либо роста противоположно случаю положительной топологической энтропии. Между этими двумя экстремальными случаями имеется множество ситуаций «медленного», т. е. субэкспоненциального роста данных величин. Примером является линейное закручивание $T: S^1 \times [0, 1] \rightarrow S^1 \times [0, 1]$, $T(x, y) = (x + y, y)$, введенное в п. 6.1.1. В этом случае мы можем указать d_n^I - r -плотное множество nr^2 шаров с центрами, расположенными равномерно на расстоянии r друг от друга по горизонтали и равномерно на расстоянии nr друг от друга по вертикали. Тогда центры также $r/2$ -отделены друг от друга в метрике d_n^I .

8.2.4. Энтропия через покрытия. Топологическая энтропия не всегда легко вычисляется, и полезно иметь альтернативные определения, чтобы в зависимости от ситуации выбрать наиболее удобное (мы скоро воспользуемся этим методом в предложении 8.2.9).

Имеются несколько величин, подобных величине $S_d(f, r, n)$, которые можно использовать для определения топологической энтропии. Допустим, что

$D_d(f, r, n)$ — это минимальное число множеств, диаметр которых в метрике d_n^f меньше чем r и объединение которых покрывает X .

Лемма 8.2.5. *Предел $\tilde{h}_d(f, r) := \lim_{n \rightarrow \infty} (1/n) \ln D_d(f, r, n)$ существует для любого $r > 0$.*

Доказательство. Если A — множество, d_n^f -диаметр которого меньше чем r , и B — множество, d_m^f -диаметр которого меньше чем r , тогда множество $A \cap f^{-n}(B)$ имеет d_{m+n}^f -диаметр меньше чем r . Таким образом, если $\mathfrak{A}$ — покрытие пространства X , состоящее из $D_d(f, r, n)$ множеств d_n^f -диаметра меньше чем r , и $\mathfrak{B}$ — покрытие пространства X , состоящее из $D_d(f, r, m)$ множеств d_m^f -диаметра меньше чем r , тогда покрытие всеми множествами $A \cap f^{-n}(B)$, где $A \in \mathfrak{A}$, $B \in \mathfrak{B}$, содержит не больше $D_d(f, r, n) \cdot D_d(f, r, m)$ множеств и является покрытием множествами d_{m+n}^f -диаметра меньше r . Таким образом,

$$D_d(f, r, m+n) \leq D_d(f, r, n) \cdot D_d(f, r, m)$$

для всех m, n . Для $a_n = \ln D_d(f, r, n)$ это означает, что $a_{m+n} \leq a_n + a_m$, и, следовательно, $\lim_{n \rightarrow \infty} a_n/n$ существует согласно лемме 4.3.7. $\square$

Предложение 8.2.6. *Если $\underline{h}_d(f, r) := \lim_{n \rightarrow \infty} (1/n) \ln S_d(f, r, n)$, то*

$$\lim_{r \rightarrow 0} \tilde{h}_d(f, r) = \lim_{r \rightarrow 0} \underline{h}_d(f, r) = \lim_{r \rightarrow 0} h_d(f, r) = h(f). \quad (8.2.4)$$

Доказательство. Диаметр r -шара не превосходит $2r$, так что каждое покрытие r -шарами является покрытием множествами, диаметр которых не превосходит $2r$, т. е.

$$D_d(f, 2r, n) \leq S_d(f, r, n). \quad (8.2.5)$$

С другой стороны, любое множество, диаметр которого не превосходит r , содержится в r -шаре с центром в каждой из своих точек, поэтому

$$S_d(f, r, n) \leq D_d(f, r, n). \quad (8.2.6)$$

Таким образом,

$$\tilde{h}_d(f, 2r) \leq \underline{h}_d(f, r) \leq h_d(f, r) \leq \tilde{h}_d(f, r). \quad \square$$

8.2.5. Топологическая энтропия через отделенные множества. Топологическую энтропию можно также определить через максимальное число $N_d(f, r, n)$ точек пространства X , находящихся на d_n^f -расстоянии не меньше r друг от друга. Мы называем такое множество точек (n, r) -отделенным (см. рис. 8.2.1). Такие точки порождают максимальное число отрезков орбит длины n , различных с точностью r .

Предложение 8.2.7. *Справедливы равенства*

$$h_{\text{top}}(f) = \lim_{r \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \ln N_d(f, r, n) = \lim_{r \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \ln N_d(f, r, n). \quad (8.2.7)$$

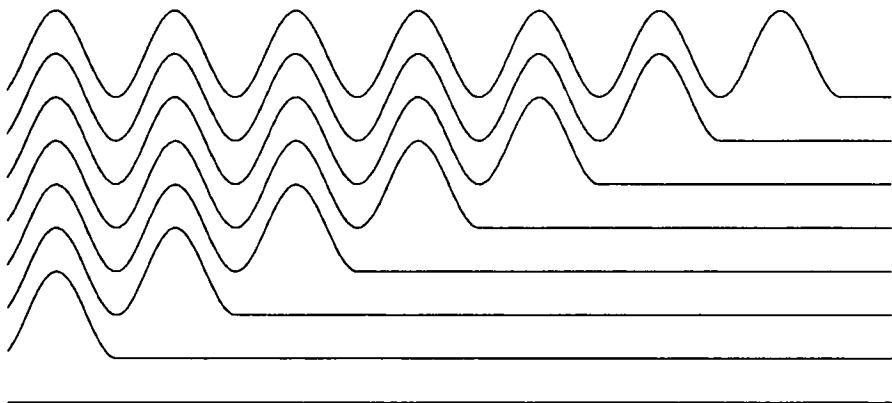


Рис. 8.2.1. Отделенное множество

Замечание 8.2.8. Это объясняет, почему в начале параграфа мы дали описание энтропии как скорости экспоненциального роста числа отрезков орбит, различных с произвольной конечной точностью.

Доказательство. Максимальное (n, r) -отделенное множество является (n, r) -плотным, т.е. для любого такого множества точек r -шары с центрами в этих точках покрывают X , потому что иначе можно было бы увеличить наше множество, прибавив любую не покрытую точку. Таким образом,

$$S_d(f, r, n) \leq N_d(f, r, n). \quad (8.2.8)$$

С другой стороны, ни один $r/2$ -шар не может содержать две точки, находящиеся на расстоянии r друг от друга. Таким образом,

$$N_d(f, r, n) \leq S_d\left(f, \frac{r}{2}, n\right). \quad (8.2.9)$$

Используя неравенства (8.2.8) и (8.2.9), мы получаем

$$h_d(f, r) \leq \varliminf_{n \rightarrow \infty} \frac{1}{n} \ln N_d(f, r, n) \leq \varlimsup_{n \rightarrow \infty} \frac{1}{n} \ln N_d(f, r, n) \leq h_d\left(f, \frac{r}{2}\right). \quad (8.2.10)$$

Далее результат следует из предложения 8.2.6. □

8.2.6. Некоторые свойства энтропии. Следующее предложение описывает некоторые стандартные элементарные свойства топологической энтропии. Доказательства показывают, насколько полезно переходить от одного из трех определений к другому.

Предложение 8.2.9. 1) Если Λ — замкнутое f -инвариантное множество, то $h_{\text{top}}(f|_{\Lambda}) \leq h_{\text{top}}(f)$.

2) Если $X = \bigcup_{i=1}^m \Lambda_i$, где Λ_i ($i = 1, \dots, m$) — замкнутые f -инвариантные множества, тогда $h_{\text{top}}(f) = \max_{1 \leq i \leq m} h_{\text{top}}(f|_{\Lambda_i})$.

3) Справедливо равенство $h_{\text{top}}(f^m) = |m| h_{\text{top}}(f)$.

4) Если g — фактор отображения f , тогда $h_{\text{top}}(g) \leq h_{\text{top}}(f)$.

5) Справедливо равенство $h_{\text{top}}(f \times g) = h_{\text{top}}(f) + h_{\text{top}}(g)$, где $f: X \rightarrow X$, $g: Y \rightarrow Y$ и отображение $f \times g: X \times Y \rightarrow X \times Y$ определено соотношением $(f \times g)(x, y) = (f(x), g(y))$.

Доказательство. Утверждение 1) очевидно, так как каждое покрытие пространства X множествами d_n^f -диаметра меньше чем r является также покрытием множества Λ .

Чтобы доказать утверждение 2), заметим, что $D_d(f, r, n) \leq \sum_{i=1}^m D_d(f|_{\Lambda_i}, r, n)$, поскольку объединение покрытий множеств $\Lambda_1, \dots, \Lambda_m$ множествами диаметра меньше чем r является покрытием пространства X . Таким образом,

$$D_d(f|_{\Lambda_i}, r, n) \geq \frac{1}{m} D_d(f, r, n)$$

по крайней мере для одного i . Поскольку имеется только конечное число значений i , по крайней мере одно i подходит для бесконечного множества номеров n . Для этого $i \in \{1, \dots, m\}$ мы имеем

$$\overline{\lim}_{n \rightarrow \infty} \frac{\ln D_d(f|_{\Lambda_i}, r, n)}{n} \geq \overline{\lim}_{n \rightarrow \infty} \frac{\ln D_d(f, r, n) - \ln m}{n} = \bar{h}_d(f, r).$$

Это доказывает утверждение 2).

Если число m положительно, тогда утверждение 3) следует из двух замечаний. Во-первых,

$$d_n^m(x, y) = \max_{0 \leq i \leq n-1} d(f^{im}(x), f^{im}(y)) \leq \max_{0 \leq i \leq mn-1} d(f^i(x), f^i(y)) = d_{nm}^f(x, y),$$

так что любой r -шар в метрике d_n^m содержит r -шар в метрике d_{mn}^f и

$$S_d(f^m, r, n) \leq S_d(f, r, mn). \quad (8.2.11)$$

Следовательно, $h_{\text{top}}(f^m) \leq m h_{\text{top}}(f)$. С другой стороны, для каждого $r > 0$ имеется такое $\delta(r) > 0$, что $B(x, \delta(r)) \subset B_f(x, r, m)$ для всех $x \in X$. Таким образом,

$$B_{f^m}(x, \delta(r), n) = \bigcap_{i=0}^{n-1} f^{-im} B(f^{im}(x), \delta(r)) \subset \bigcap_{i=0}^{n-1} f^{-im} B_f(f^{im}(x), r, m) = B_f(x, r, mn).$$

Следовательно,

$$S_d(f, r, mn) \leq S_d(f^m, \delta(r), n)$$

и $m h_{\text{top}}(f) \leq h_{\text{top}}(f^m)$. Если отображение f обратимо, тогда $B_f(x, r, n) = B_{f^{-1}}(f^{n-1}(x), r, n)$ и $S_d(f, r, n) = S_d(f^{-1}, r, n)$, так что $h_{\text{top}}(f) = h_{\text{top}}(f^{-1})$.

Если m отрицательно, тогда утверждение 3) следует из соответствующего утверждения для $m > 0$ и $n = -1$.

Докажем утверждение 4). Рассмотрим такие отображения $f: X \rightarrow X$, $g: Y \rightarrow Y$, $h: X \rightarrow Y$, что $h \circ f = g \circ h$ и $h(X) = Y$ (см. определение 7.3.3). Обозначим через d_X , d_Y метрики в пространствах X и Y соответственно.

Отображение h равномерно непрерывно, поэтому для любого $r > 0$ найдется такое $\delta(r) > 0$, что если $d_X(x_1, x_2) < \delta(r)$, то $d_Y(h(x_1), h(x_2)) < r$. Значит, образ любого шара радиуса $\delta(r)$ в метрике $(d_X)_n^f$ лежит внутри шара радиуса r в метрике $(d_Y)_n^g$, т. е.

$$S_{d_X}(f, \delta(r), n) \geq S_{d_Y}(g, r, n).$$

Логарифмируя обе части и переходя к пределу, мы получаем утверждение 4).

Чтобы доказать утверждение 5), мы используем метрику произведения $d((x_1, y_1), (x_2, y_2)) = \max(d_X(x_1, x_2), d_Y(y_1, y_2))$ в $X \times Y$. Шары в метрике произведения — это произведения шаров в X и Y . То же самое справедливо для шаров в метрике $d_n^{f \times g}$. Таким образом, $S_d(f \times g, r, n) \leq S_{d_X}(f, r, n) S_{d_Y}(g, r, n)$ и $h_{\text{top}}(f \times g) \leq h_{\text{top}}(f) + h_{\text{top}}(g)$. С другой стороны, произведение любого (n, r) -отделенного множества в X для f и любого (n, r) -отделенного множества в Y для g является (n, r) -отделенным множеством для $f \times g$. Поэтому

$$N_d(f \times g, r, n) \geq N_{d_X}(f, r, n) \times N_{d_Y}(g, r, n)$$

и, следовательно, $h_{\text{top}}(f \times g) \geq h_{\text{top}}(f) + h_{\text{top}}(g)$. □

Упражнения

8.2.1. Вычислите топологическую энтропию отображения $f(x) = x(1 - x)$ на $[0, 1]$.

8.2.2. Вычислите топологическую энтропию линейной подковы.

8.2.3. Предположим, что $f: S^1 \rightarrow S^1$ — сохраняющий ориентацию C^2 -диффеоморфизм без периодических точек. Найдите $h_{\text{top}}(f)$.

Задачи для углубленного изучения

8.2.4. Рассмотрим отображение $f: \mathbb{T}^3 \rightarrow \mathbb{T}^3$, $f(x, y, z) = (x, x + y, y + z)$. Найдите $h_{\text{top}}(f)$.

8.2.5. Предположим, что множество $X = \bigcup_i X_i$ компактно, каждое из множеств X_i замкнуто и $f: X \rightarrow X$ — такое отображение, что каждое из множеств X_i является f -инвариантным. Покажите, что $h_{\text{top}}(f) = \sup h_{\text{top}}(f|_{X_i})$.

8.2.6. Для данных отображений $f: X \rightarrow X$, $g: Y \rightarrow Y$ предположим, что $h \circ f = g \circ h$, где $h: X \rightarrow Y$ — такое непрерывное сюръективное отображение, что всякий элемент $y \in Y$ имеет конечное число прообразов. Покажите, что $h_{\text{top}}(f) = h_{\text{top}}(g)$.

§ 8.3. Приложения и обобщения

8.3.1. Растягивающие отображения. Первый из рассмотренных нами примеров, для которого характерна сложная структура орбит, — это растягивающие отображения E_m . Поскольку одной из особенностей такой структуры является экспоненциальный рост периодических орбит (см. предложение 7.1.2), естественно ожидать, что полная экспоненциальная сложность структуры орбит, которая и измеряется топологической энтропией, тоже будет положительна.

Предложение 8.3.1. Если $m \in \mathbb{N}$, $|m| \geq 2$, то $h_{\text{top}}(E_m) = \ln |m| = p(E_m)$.

Доказательство. Для отображения E_m , а на самом деле и для любого растягивающего отображения расстояние между итерациями любых двух точек растет, пока не становится больше некоторой константы, зависящей от отображения ($1/(2|m|)$ для отображения E_m). Для простоты обозначений предположим, что $m > 0$. Если $d(x, y) < m^{-n}/2$, тогда $d_n^{E_m}(x, y) = d(E_m^{n-1}(x), E_m^{n-1}(y))$, так что если $d_n^{E_m}(x, y) \geq r$, то $d(x, y) \geq rm^{-n}$. Положив $r = m^{-k}$, мы получаем, что $\{im^{-n-k}: i = 0, \dots, m^{n+k} - 1\}$ есть максимальное множество точек, попарные расстояния между которыми в метрике $d_n^{E_m}$ не меньше чем m^{-k} , т. е.

$$N_d(E_m, m^{-k}, n) = m^{n+k},$$

и, следовательно,

$$h(E_m) = \lim_{k \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \frac{\ln N_d(E_m, m^{-k}, n)}{n} = \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{n+k}{n} \ln m = \ln m.$$

Случай $m < 0$ совершенно аналогичен. □

Поскольку топологическая энтропия инвариантна относительно топологического сопряжения (см. следствие 8.2.3) и каждое растягивающее отображение степени m топологически сопряжено с отображением E_m (см. теорему 7.4.3), из предложения 8.3.1 мы получаем такой результат.

Следствие 8.3.2. Если $f: S^1 \rightarrow S^1$ — растягивающее отображение степени m , то

$$h_{\text{top}}(f) = p(f) = \ln |m|.$$

8.3.2. Сдвиги и топологические цепи Маркова.

Предложение 8.3.3. Для любой топологической цепи Маркова σ_A справедливы равенства $h_{\text{top}}(\sigma_A) = p(\sigma_A) = \ln |\lambda_A^{\max}|$.

Доказательство. Рассмотрим сначала сдвиг σ_N . Так же как и в п. 7.3.4, любой цилиндр

$$C_{\alpha_{-m}, \dots, \alpha_{n+m}}^{-m, \dots, n+m} := \{\omega \in \Omega_N: \omega_i = \alpha_i \text{ для } -m \leq i \leq m+n\} \quad (8.3.1)$$

является в то же время шаром радиуса $r_m = \lambda^{-m}/2$ с центром в каждой из своих точек относительно метрики $d_n^{\sigma_N}$, ассоциированной со сдвигом σ_N (поскольку $\lambda > 3$). Таким образом, любые два шара радиуса r_m в метрике $d_n^{\sigma_N}$ либо совпадают, либо не пересекаются, и имеется в точности N^{n+2m+1} различных шаров вида (8.3.1), так что $S_{d_\lambda}(\sigma_N, r_m, n) = N^{n+2m+1}$ и

$$h(\sigma_N) = \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{n} \ln N^{n+2m+1} = \ln N.$$

Точно так же, если σ_A — топологическая цепь Маркова, тогда $S_d(\sigma_A, r_m, n)$ — это число тех цилиндров вида (8.3.1), которые имеют непустое пересечение с Ω_A . Предположим, что каждая строка матрицы A содержит по крайней мере одну единицу. Поскольку число допустимых путей длины n , которые начинаются в точке i и кончаются в точке j , равно компоненте a_{ij}^n матрицы A^n (см. лемму 7.3.5), число непустых цилиндров ранга $n+1$ в Ω_A равно $\sum_{i,j=0}^{N-1} a_{ij}^n < C \cdot \|A^n\|$

для некоторой константы C . С другой стороны, $\sum_{i,j=0}^{N-1} a_{ij}^n > c \|A^n\|$ для другой константы $c > 0$, поскольку все числа a_{ij}^n неотрицательные, и, следовательно, левая часть — это норма $\sum_{i,j=0}^{N-1} a_{ij}^n$ матрицы A^n , которая эквивалентна обычной норме, потому что все нормы на $\mathbb{R}^{N^2}$ эквивалентны. Таким образом, мы получаем

$$S_{d_\lambda}(\sigma_A, r_m, n) = \sum_{i,j=0}^{N-1} a_{ij}^{n+2m} \quad (8.3.2)$$

и

$$\begin{aligned}
 h(\sigma_A) &= \lim_{m \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \ln S_{d_n}(\sigma_A, r_m, n) = \\
 &= \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{n} \ln \|A^{n+2m}\| = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \|A^n\| = \ln r(A) = \ln |\lambda_A^{\max}|, \quad (8.3.3)
 \end{aligned}$$

где $r(A)$ — спектральный радиус матрицы A (см. определение 3.3.1). Наше утверждение следует из равенства (8.3.3) и предложения 7.3.7. $\square$

8.3.3. Гиперболический автоморфизм тора. При вычислении энтропии автоморфизма тора мы воспользуемся как кодированием, так и известными нам результатами о скорости роста периодических точек.

Предложение 8.3.4. Если отображение $F_L: \mathbb{T}^2 \rightarrow \mathbb{T}^2$ задается соотношением $F_L(x, y) = (2x + y, x + y) \pmod{1}$, то

$$h(F_L) = \rho(F_L) = \frac{3 + \sqrt{5}}{2}.$$

Доказательство. В п. 7.4.5 мы показали, что отображение

$$F_L(x, y) = (2x + y, x + y) \pmod{1}$$

является фактором топологической цепи Маркова σ_A , где

$$A = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix},$$

и что максимальное собственное значение матрицы A равно $\lambda_A^{\max} = \frac{3 + \sqrt{5}}{2}$. Как следует из утверждения 4) предложения 8.2.9,

$$h(F_L) \leq h(\sigma_A) = \ln \frac{3 + \sqrt{5}}{2}. \quad (8.3.4)$$

С другой стороны, покажем теперь, что множество n -периодических точек F_L является $(n, 1/4)$ -отделенным для любого $n \in \mathbb{N}$. В соответствии с предложением 7.1.10 из этого следует, что $N_d(F_L, 1/4, n) \geq P_n(F_L)$ и

$$h(F_L) \geq \rho(F_L) = \ln \frac{3 + \sqrt{5}}{2}.$$

Тогда справедливость нашего результата будет следовать из неравенства (8.3.4).

Если p, q — n -периодические точки и $d(p, q) < 1/4$, то имеется однозначно определенный минимальный прямоугольник R с вершинами p, s, q, t , образованный отрезками растягивающихся и сжимающихся прямых, проходящих через p и q (см. рис. 8.3.1). Под действием отображения F_L стороны ps и qt растягиваются с коэффициентом $\lambda_1 = \frac{3 + \sqrt{5}}{2} > 2$, в то время как другие две стороны сжимаются с коэффициентом λ_1^{-1} .

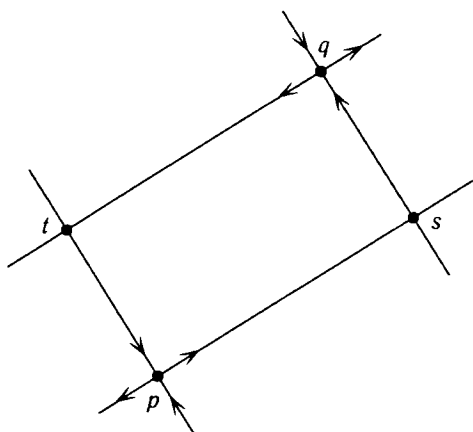


Рис. 8.3.1. Гетероклинические точки

Из этого следует, что $F_L^n(R) \neq R$, потому что отображение F_L^n не может оставлять все четыре стороны инвариантными при одновременном их растяжении и сжатии. Другими словами, $F_L^{-n}(R) \neq R$. Поэтому найдется номер $k \leq n$, для которого $F_L^k(R)$ не может быть минимальным прямоугольником. Тогда для наименьшего такого k мы имеем $d(F_L^k(p), F_L^k(q)) > 1/4$, потому что прямоугольник с диагоналями меньше чем $1/4$ минимален. Таким образом, периодические точки периода n образуют $(n, 1/4)$ -отделенное множество. $\square$

Замечание 8.3.5. В случае растягивающих отображений E_m и для топологических цепей Маркова σ_A можно также показать, что периодические точки образуют (n, r_0) -отделенные множества для некоторого r_0 . Это позволяет нам получить неравенство $h_{\text{top}} \geq \rho$ одним и тем же способом для всех трех случаев.

8.3.4. Периодические точки и энтропия. В наших примерах мы можем наблюдать интересную модель поведения. Для двух гладких примеров со сложной экспоненциально растущей структурой орбит, а именно для растягивающих отображений (см. предложение 8.3.1) и гиперболических автоморфизмов тора (см. предложение 8.3.4), две естественные меры экспоненциального роста орбит — экспоненциальная скорость роста периодических точек ρ и топологическая энтропия h_{top} — совпадают. Это довольно широко распространенное явление, хотя и не универсальное. Оно связано с локальной гиперболической структурой, т. е. с наличием растяжения и сжатия в этих примерах. (Об этом подробно говорится в гл. 10.) Для топологических цепей Маркова скорость роста периодических точек и топологическая энтропия также совпадают (предложение 8.3.3). Гиперболическость и в этом случае является подходящим объяснением, поскольку в силу предложения 7.4.6 топологические цепи Маркова топологически сопряжены с ограничением некоторых гладких систем на специальные инвариантные множества, которые имеют гиперболическое поведение.

8.3.5. Топологическая энтропия для потоков. Определение топологической энтропии $h_{\text{top}}(\Phi)$ для потока $\Phi = (\varphi^t)_{t \in \mathbb{R}}$ полностью аналогично определению

для дискретного случая. Единственное изменение состоит в том, что метрики в соотношении (8.2.1) заменяются неубывающим семейством метрик

$$d_T^\Phi(x, y) = \max_{0 \leq t \leq T} d(\varphi^t(x), \varphi^t(y)).$$

Схожесть определений приводит к особенно полезному следствию, аналогичному утверждению 3) предложения 8.2.9.

Предложение 8.3.6. *Справедливо равенство $h_{\text{top}}(\Phi) = h_{\text{top}}(\varphi^1)$.*

Доказательство. Если $r > 0$, тогда в силу компактности и непрерывности найдется такое $\delta(r) > 0$, что если $d(x, y) \leq \delta(r)$, то $\max_{0 \leq t \leq 1} d(\varphi^t(x), \varphi^t(y)) < r$. Тогда

любой r -шар в метрике d_T^Φ содержит $\delta(r)$ -шар в метрике $d_{[T]}^\Phi$. С другой стороны, $d_n^\Phi \geq d_n^1$. Из этих двух замечаний и следует нужное утверждение. $\square$

Таким образом, топологическая энтропия для потока инвариантна относительно эквивалентности потоков, т.е. совпадает для двух потоков, для которых сдвиги за время t топологически сопряжены, причем сопряжение одно и то же для всех t . Она изменяется при замене времени (см. определение 9.4.12), а следовательно, и при орбитальной эквивалентности (эквивалентность потоков с заменой времени) довольно сложным образом. Можно показать, что для потока без неподвижных точек любая замена времени сохраняет нулевую топологическую энтропию, т.е. при замене времени в потоке с нулевой энтропией мы также получаем поток с нулевой энтропией. Если топологическая энтропия для отображения или потока равна нулю, субэкспоненциальная асимптотика любой из величин, используемых для ее определения, может помочь при изучении сложности структуры орбит.

8.3.6. Локальная энтропия как мера неустойчивости. Как мы заметили во введении к этому параграфу, энтропия может рассматриваться как мера хаоса в системе. Теперь мы покажем в явном виде, как с помощью энтропии получить количественную характеристику неустойчивости по начальным условиям в динамической системе. Для этого мы введем близкое понятие *локальной энтропии* и объясним, как она измеряет неустойчивость с одной стороны и как она связана с топологической энтропией с другой стороны.

Зафиксируем точку x и «микроскопическое» значение ϵ , а также «макроскопическое» r и допустим, что $N_d(f, r, n, x, \epsilon)$ — максимальное число точек в $B_d(x, \epsilon)$, попарные расстояния между которыми в метрике d_n^f не меньше чем r . Если это число велико, то это указывает на весьма неустойчивое по начальным условиям поведение.

Определение 8.3.7. Если

$$h_{d,x,r}(f) := \lim_{\epsilon \rightarrow 0} \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \ln N_d(f, r, n, x, \epsilon),$$

тогда величина

$$h_{d,x}(f) := \lim_{r \rightarrow 0} h_{d,x,r}(f)$$

называется *локальной энтропией* отображения f в точке x .

Замечание 8.3.8. Пределы существуют, потому что величина $N_d(f, r, n, x, \epsilon)$ возрастает с ростом ϵ и убывает с ростом r .

Предложение 8.3.9. *Справедливо неравенство $h_{d,x}(f) \leq h_{\text{top}}(f)$.*

Доказательство. Топологическая энтропия соответствует случаю неизменного ϵ , когда $B(x, \epsilon)$ совпадает со всем пространством. Поэтому любая точка с сильной неустойчивостью в этом смысле обязательно порождает большую топологическую энтропию. $\square$

С другой стороны, имеется связь между локальной энтропией и величиной $h_d(f, r)$ (см. (8.2.2)).

Предложение 8.3.10. Для $r > 0$ существует такая точка x , что

$$h_{d,x,r}(f) \geq h_d(f, r).$$

Доказательство. Если $S_d(f, r, n, x, \epsilon)$ — минимальное число r -шаров в метрике d_n^f , покрывающих $B_d(x, \epsilon)$, тогда имеется такая точка x , что

$$S_d(f, r, n) \leq S_d(\epsilon) S_d(f, r, n, x, \epsilon), \quad (8.3.5)$$

поскольку мы можем взять покрытие пространства шарами в метрике $S_d(\epsilon)$ радиуса ϵ и, обозначая их центры через x_j , получим

$$S_d(f, r, n) \leq \sum_{j=1}^{S_d(\epsilon)} S_d(f, r, n, x_j, \epsilon),$$

следовательно,

$$S_d(f, r, n, x, \epsilon) \geq S_d(f, r, n) / S_d(\epsilon),$$

где x совпадает с одной из точек x_j .

При $n \rightarrow \infty$ мы получаем последовательность таких точек x_n , удовлетворяющих неравенству (8.3.5) для соответствующих значений n . Возьмем предельную точку x этой последовательности и рассмотрим 2ϵ -шар с центром в этой точке. Для достаточно больших n мы имеем $B_d(x_n, \epsilon) \subset B_d(x, 2\epsilon)$, и, следовательно,

$$S_d(f, r, n, x, 2\epsilon) \geq S_d(f, r, n) / S_d(\epsilon)$$

для всех n , а значит,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln S_d(f, r, n, x, 2\epsilon) \geq \lim_{n \rightarrow \infty} \frac{1}{n} \ln (S_d(f, r, n) / S_d(\epsilon)) = h_d(f, r).$$

Используя такие же рассуждения, как и прежде, мы можем заменить $S_d(f, r, n, x, 2\epsilon)$ соответствующим числом r -плотных точек и, переходя к пределу при $\epsilon \rightarrow 0$, получим

$$h_{d,x,r}(f) \geq h_d(f, r)$$

для всех r . $\square$

Замечание 8.3.11. Поскольку $h_d(f, r) \xrightarrow{r \rightarrow 0} h(f)$, это показывает, что имеются точки, для которых значение $h_{d,x,r}(f)$ произвольно близко к топологической энтропии. Таким образом, верхняя грань локальных энтропий по пространству равна топологической энтропии, и топологическая энтропия действительно измеряет неустойчивость по начальным условиям.

Упражнения

8.3.1. Докажите следствие 8.3.2, не используя топологическое сопряжение.

8.3.2. Постройте отображение с положительной топологической энтропией, которое не имеет периодических точек.

8.3.3. Постройте топологически транзитивное отображение компактного метрического пространства, которое имеет бесконечную топологическую энтропию.

8.3.4. Докажите, что локальная энтропия отображения E_m не зависит от точки и равна топологической энтропии.

8.3.5. Докажите, что локальная энтропия сдвига σ_m не зависит от точки и равна топологической энтропии.

8.3.6. Докажите, что локальная энтропия автоморфизма тора, индуцированного матрицей $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, не зависит от точки и равна топологической энтропии.

Задачи для углубленного изучения

8.3.7. Рассмотрим замкнутый единичный диск в $\mathbb{R}^2$ и отображение f_λ на нем, определенное в полярных координатах соотношением $f_\lambda(re^{i\theta}) = \lambda re^{2i\theta}$, где $0 \leq \lambda \leq 1$. Покажите, что $h_{\text{top}}(f_1) \geq \ln 2$ и что $h_{\text{top}}(f_\lambda) = 0$ для $\lambda < 1$.

8.3.8. Докажите, что $h_{\text{top}}(\varphi^t) = |t|h_{\text{top}}(\varphi^1)$ для любого потока φ^t .

8.3.9. Приведите пример топологически транзитивного отображения, для которого локальная энтропия не постоянна.

Часть II

Динамические системы: обзор современных результатов

В этой части книги результаты и темы, рассмотренные в основном курсе, получают дальнейшее развитие. Изложение опирается на предшествующий курс, но в то же время является самостоятельным описанием избранных более современных достижений. Выбор материала обусловлен тематикой основного курса, а также тем, насколько подробно соответствующие результаты освещены в литературе.

Начиная с гл. 10 мы используем гораздо менее строгий и детальный стиль изложения, чем в основном курсе. Многие результаты снабжены по крайней мере наброском доказательства, но наличие слова «доказательство» не означает, что рассуждения столь же полны и детальны, как в первой части книги. Основное внимание уделяется дальнейшему развитию и применению изложенных ранее идей. В тексте приводится много ссылок. Глава 9, однако, за исключением отсутствия упражнений написана на том же уровне строгости, как и основной курс, и при желании может рассматриваться как его часть.

Главы в значительной мере независимы друг от друга, их можно читать избирательно и в любом порядке.

ПРОСТАЯ ДИНАМИКА КАК СРЕДСТВО АНАЛИЗА

§ 9.1. Введение

9.1.1. Приложения принципа сжимающих отображений. Простые динамические системы со сложной структурой орбит, представленные в гл. 7 и повторно рассмотренные в § 8.3, являются примерами *гиперболических* динамических систем. Многие результаты гиперболической динамики получаются (более или менее) непосредственно из принципа сжимающих отображений, который впервые был рассмотрен как пример динамической системы с простой динамикой в гл. 2. Хотя мы уже использовали его в § 2.5 как средство для изучения других динамических систем, в связи с его ведущей ролью в гиперболической динамике следует дать более полное представление о его приложениях. Соответственно главная задача этой главы состоит в том, чтобы предоставить примеры использования принципа сжимающих отображений, т. е. показать, как один важный результат, касающийся определенного класса систем с простой динамикой, при использовании его в некотором вспомогательном пространстве позволяет провести анализ (сложных) динамических систем. Поскольку результаты, которые мы получим, довольно важны, мы поговорим и об их дальнейшем развитии, особенно в связи основами теории дифференциальных уравнений. В этой главе мы сохраним тот же уровень строгости доказательств, что и в основном курсе.

Как и в предшествующих главах, результаты, полученные здесь, имеют интересные приложения, связанные с изучением динамических систем. В случае вопроса о существовании и единственности решений дифференциальных уравнений это очевидно. Но и все другие представленные здесь результаты также являются стандартными средствами в динамике. Эта глава не может охватить все такие приложения, но некоторые из них представлены в других главах. Так, например, в гл. 10 приведена лемма Аносова о замыкании (теорема 10.2.2), которая следует из принципа сжимающих отображений и теоремы о гиперболической неподвижной точке (теорема 9.5.4). В настоящей же главе мы рассмотрим еще один замечательный пример — теорему об устойчивом многообразии (теорема 9.5.2).

9.1.2. Краткий обзор. Мы начнем с вывода двух важных результатов из математического анализа — теорем об обратной и неявной функции. Из теоремы о неявной функции следует и еще один результат относительно самого принципа сжимающих отображений: неподвижная точка сжимающего отображения гладко зависит от отображения. Первое и самое непосредственное приложение этих теорем — приведенный в § 9.3 результат об устойчивости трансверсальных неподвижных точек. Он состоит в том, что простое условие на линейную часть отображения в неподвижной точке может гарантировать, что неподвижная точка

сохраняется при возмущении отображения. Подобная ситуация наблюдалась для сжимающих отображений (см. предложение 2.2.20), и она очень хорошо соответствует подходу, основанному на линеаризации (о котором говорилось в § 2.1, в начале гл. 3 и, например, в п. 6.2.2.7). В этих первых приложениях пространство, на котором определяется сжимающее отображение, совпадает с тем пространством, в котором поставлена задача. Однако в приложениях принципа сжимающих отображений к теореме существования и единственности решений дифференциальных уравнений (см. § 9.4), теореме об устойчивом многообразии (§ 9.5) и другим важным теоремам анализа ситуация сводится к поиску неподвижной точки в некотором пространстве функций, которое имеет бесконечную размерность, а не в евклидовом пространстве.

9.1.3. Сведение задачи к принципу сжимающих отображений. Общая черта всех результатов — применение принципа сжимающих отображений в некотором вспомогательном пространстве, однако в одних примерах это пространство найти несложно, а в других это требует значительных усилий. Наиболее непосредственным и самым ранним из изученных здесь приложений являются приближения Пикара (§ 9.4), хотя для них рассматриваемое пространство не столь просто, как в приведенных ранее примерах. Первый шаг доказательства теоремы об обратной функции (теорема 9.2.2) близок к методу Ньютона, хотя и требует больше изобретательности. В доказательстве устойчивости трансверсальных неподвижных точек (предложение 9.3.1) нет такого же очевидного обоснования для первого шага, но для него характерны некоторые особенности, являющиеся общими для других приложений принципа сжимающих отображений в динамике. В основе доказательства лежит комбинация трансверсальности и близости (малости возмущения). При этом строится обратимое отображение, после чего берется композиция обратного к нему с сильно сжимающим отображением, которое связано с возмущением. (Тонкость состоит в том, что необходимо добиться, чтобы нужный объект являлся неподвижной точкой получающегося сжимающего отображения.)

Все приведенные в этой главе приложения принципа сжимающих отображений, за исключением приближений Пикара, связаны с линеаризацией. Это также является типичным для приложений в теории гладких динамических систем.

§ 9.2. Теоремы о неявной и об обратной функции в евклидовом пространстве

9.2.1. Теорема об обратной функции. Теорема об обратной функции говорит, что если дифференцируемое отображение имеет обратимую производную в некоторой точке, тогда отображение обратимо вблизи этой точки. Этот результат связан с линеаризацией: если мы налагаем некоторое условие качественного характера (условие типа «да — нет») на линейную часть (в данном случае — обратимость), тогда оно остается справедливым и для нелинейного отображения, по крайней мере в некоторой окрестности. Вариант этой теоремы для случая действительной прямой хорошо знаком из курса математического анализа.

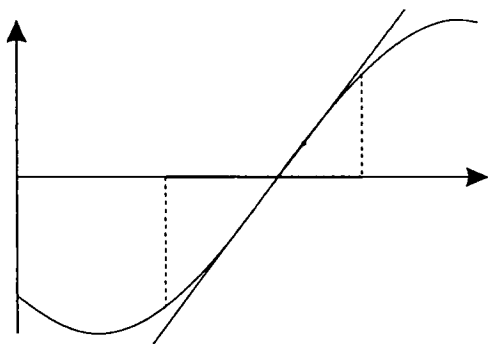


Рис. 9.2.1. Теорема об обратной функции

Теорема 9.2.1. *Предположим, что $I \subset \mathbb{R}$ — интервал и $f: I \rightarrow \mathbb{R}$ дифференцируемая функция. Если $a \in I$ — такое число, что $f'(a) \neq 0$ и производная f' непрерывна в точке a , тогда функция f обратима в окрестности U точки a и $(f^{-1})'(y) = 1/f'(x)$, где $y = f(x)$.*

Обычно утверждение об обратимости считается легкой частью, а формула для производной — трудной, потому что основные примеры обратимых вещественных функций из математического анализа задаются формулами, в которых обратимость довольно очевидна. Однако главная сущность этого результата состоит в том, что мы можем сделать вывод об обратимости на основе знания линейной части отображения только в одной точке, без какой-либо дополнительной информации. После этого формула для производной — легкое следствие. Легко получаются и производные высших порядков. Этот результат довольно прост при обобщении на пространство $\mathbb{R}^n$, однако мы сначала приведем его доказательство для простого случая одной переменной.

Доказательство. Мы хотим для данного y решить уравнение $f(x) = y$ относительно x , или, что то же самое, найти корень функции $F_y(x) := y - f(x)$. Для этого сначала получим подходящее сжимающее отображение.

Пространство. Пространство, на котором действует сжимающее отображение, — действительная прямая.

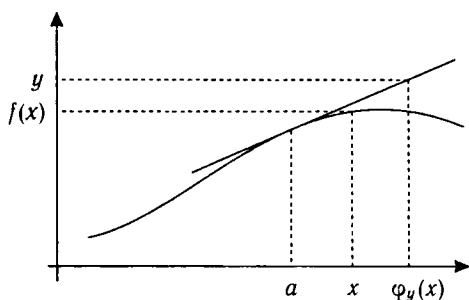
Определение сжимающего отображения. В соответствии с методом Ньютона из п. 2.2.8 мы выберем начальное приближение x (значение y в данный момент фиксировано) и будем улучшать приближения, многократно применяя отображение

$$F_y(x) = x - \frac{F_y(x)}{F_y'(x)} = x + \frac{y - f(x)}{f'(x)}.$$

Чтобы проверить, что это отображение является сжимающим, необходимо оценить вторую производную функции f , о существовании которой мы ничего не предполагаем. Вместо этого удобно рассматривать отображение

$$\varphi_y(x) := x + \frac{y - f(x)}{f'(a)}$$

на I . Его неподвижные точки являются решениями нашей задачи, поскольку $\varphi_y(x) = x$ тогда и только тогда, когда $f(x) = y$.

Рис. 9.2.2. Отображение φ_y

Свойство сжатия. Теперь мы покажем, что φ_y — сжимающее отображение некоторого замкнутого подмножества окрестности точки a . Тогда согласно принципу сжимающих отображений оно имеет единственную неподвижную точку, а следовательно, существует единственное такое значение x , что $f(x) = y$.

Для этого положим $A := f'(a)$ и $\alpha = |A|/2$. Из непрерывности функции f' в точке a следует существование такого $\varepsilon > 0$, что $W := (a - \varepsilon, a + \varepsilon) \subset I$ и $|f'(x) - A| < \alpha$ для $x \in \bar{W}$, где $\bar{W}$ — замыкание множества W .

Чтобы заключить, что отображение φ_y является сжимающим на $\bar{W}$, заметим, что если $x \in \bar{W}$, то

$$|\varphi'_y(x)| = \left| 1 - \frac{f'(x)}{A} \right| = \left| \frac{A - f'(x)}{A} \right| < \frac{\alpha}{|A|} = 1/2.$$

Используя предложение 2.2.3, мы получаем $|\varphi_y(x) - \varphi_y(x')| \leq |x - x'|/2$ для $x, x' \in \bar{W}$.

Мы должны также показать, что $\varphi_y(\bar{W}) \subset \bar{W}$ для значений y , достаточно близких к $b := f(a)$. Допустим, что $\delta = A\varepsilon/2$ и $V = (b - \delta, b + \delta)$. Тогда для $y \in V$ мы имеем

$$|\varphi_y(a) - a| = \left| a + \frac{y - f(a)}{A} - a \right| = \left| \frac{y - b}{A} \right| < \left| \frac{\delta}{A} \right| = \frac{\varepsilon}{2},$$

так что если $x \in \bar{W}$, то

$$|\varphi_y(x) - a| \leq |\varphi_y(x) - \varphi_y(a)| + |\varphi_y(a) - a| < \frac{x - a}{2} + \frac{\varepsilon}{2} \leq \varepsilon$$

и, следовательно, $\varphi_y(x) \in \bar{W}$.

Поэтому, применяя предложение 2.2.21 к отображению $\varphi_y: \bar{W} \rightarrow \bar{W}$ для $y \in V$, мы получаем единственную неподвижную точку $g(y) \in \bar{W}$, которая непрерывно зависит от y .

Теперь докажем, что обратное отображение дифференцируемо. Пусть $y = f(x) \in V$. Мы хотим показать, что производная $g'(y)$ существует и является обратной величиной к $B := f'(g(y))$.

Пусть $U := g(V) = W \cap f^{-1}(V)$ (прообраз относительно f), так что множество U открыто. Возьмем $y + k = f(x + h) \in V$. Тогда

$$\frac{|h|}{2} \geq |\varphi_y(x + h) - \varphi_y(x)| = \left| h + \frac{f(x) - f(x + h)}{A} \right| = \left| h - \frac{k}{A} \right| \geq |h| - \left| \frac{k}{A} \right|,$$

следовательно,

$$\frac{|h|}{2} \leq \left| \frac{k}{A} \right| < \frac{|k|}{\alpha} \quad \text{и} \quad \frac{1}{|k|} < \frac{2}{\alpha|h|}.$$

Поскольку $g(y+k) - g(y) - k/B = h - k/B = -(f(x+h) - f(x) - Bh)/B$, мы получаем

$$\frac{|g(y+k) - g(y) - k/B|}{|k|} < \frac{2}{|B|\alpha} \frac{|f(x+h) - f(x) - Bh|}{|h|} \xrightarrow{|h| \leq 2|k|/\alpha \rightarrow 0} 0,$$

что доказывает равенство $g'(y) = 1/B = 1/f'(g(y))$.

Наконец, предположим, что $f \in C^r$. Покажем по индукции, что $g \in C^r$. Для этого предположим, что $g \in C^k$ для некоторого $k < r$ (базис индукции при $k = 0$). Тогда $f'(g(y)) \in C^k$, и, следовательно, функция g' также принадлежит классу C^k . Таким образом, $g \in C^{k+1}$. $\square$

Теперь обобщим эти рассуждения на произвольное пространство $\mathbb{R}^m$.

Теорема 9.2.2 (об обратной функции). *Предположим, что множество $O \subset \mathbb{R}^m$ открыто, отображение $f: O \rightarrow \mathbb{R}^m$ дифференцируемо и производная Df обратима в точке $a \in O$ и непрерывна в этой точке. Тогда существуют такие окрестности $U \subset O$ точки a и V точки $b := f(a) \in \mathbb{R}^m$, что f является биекцией из U в V (т. е. отображение f взаимно однозначно на U и $f(U) = V$). Обратное отображение $g: V \rightarrow U$ для f дифференцируемо, причем $Dg(y) = (Df(g(y)))^{-1}$. Кроме того, если f — отображение класса C^r (т. е. все частные производные отображения f до порядка r включительно существуют и непрерывны) на U , то таково же и обратное отображение.*

Доказательство. Доказательство практически такое же, как и прежде. Нужно только заменить некоторые числа линейными отображениями и некоторые абсолютные величины нормами.

Пространство. Сжимающее отображение действует в $\mathbb{R}^m$.

Отображение. Для любого заданного $y \in \mathbb{R}^m$ рассмотрим отображение

$$\varphi_y(x) := x + Df(a)^{-1}(y - f(x))$$

на O . Заметим, что $\varphi_y(x) = x$ тогда и только тогда, когда $f(x) = y$, так что мы будем искать единственную неподвижную точку отображения φ_y . Нам необходимо определить множество W , на котором это отображение является сжимающим.

Свойство сжатия. Пусть $A := Df(a)$ и $\alpha < \|A^{-1}\|^{-1}/2$. Используя непрерывность отображения Df в точке a , возьмем такое $\epsilon > 0$, что $\|Df(x) - A\| < \alpha$, если x принадлежит замыканию множества $W := B(a, \epsilon)$. Чтобы заключить, что отображение φ_y является сжимающим, заметим, что

$$\|D\varphi_y(x)\| = \|\text{Id} - A^{-1}Df(x)\| = \|A^{-1}(A - Df(x))\| < \|A^{-1}\|\alpha =: \lambda < 1/2$$

для $x \in W$, и применяя следствие 2.2.15, получим, что $\|\varphi_y(x) - \varphi_y(x')\| \leq \lambda\|x - x'\|$ для $x, x' \in W$. Поэтому в соответствии с предложением 2.2.20 найдется такая

окрестность V точки b , что φ_y является сжатием $\overline{W}$ для всех $y \in V$ и имеет единственную неподвижную точку $g(y) \in W$ (которая непрерывно зависит от y). Множество $U := g(V) = W \cap f^{-1}(V)$ открыто.

Детерминант отображения $Df(x)$ непрерывно зависит от Df и, следовательно, непрерывен в точке a как функция от x . Таким образом, выбирая, если это необходимо, окрестность V (а следовательно, и U) меньшего размера, мы можем считать, что $\det Df \neq 0$ на U и поэтому отображение $Df(x)$ обратимо на U .

Мы хотим показать, что если $y = f(x) \in V$, то отображение $Dg(y)$ существует и является обратным для $B := Df(g(y))$. Возьмем $y + k = f(x + h) \in V$. Тогда

$$\begin{aligned} \frac{\|h\|}{2} &\geq \|\varphi_y(x + h) - \varphi_y(x)\| = \|h + A^{-1}(f(x) - f(x + h))\| = \\ &= \|h - A^{-1}k\| \geq \|h\| - \|A^{-1}\| \|k\|, \end{aligned} \quad (9.2.1)$$

так что

$$\frac{\|k\|}{\alpha} > \|A^{-1}\| \|k\| \geq \frac{\|h\|}{2} \quad \text{и} \quad \frac{1}{\|k\|} < \frac{2}{\alpha \|h\|}.$$

Поскольку $g(y + k) - g(y) - B^{-1}k = h - B^{-1}k = -B^{-1}(f(x + h) - f(x) - Bh)$, мы получаем

$$\frac{\|g(y + k) - g(y) - B^{-1}k\|}{\|k\|} < \frac{\|B^{-1}\|}{\alpha/2} \frac{\|f(x + h) - f(x) - Bh\|}{\|h\|} \xrightarrow{\|h\| \leq 2\|k\|/\alpha \rightarrow 0} 0,$$

откуда следует, что $Dg(y) = B^{-1}$.

Наконец, предположим, что $f \in C^r$ и $g \in C^k$ для некоторых k и r , $k < r$. Тогда $Df(g(y)) \in C^k$, и обратное отображение Dg также принадлежит классу C^k (это можно показать, используя формулу для обратной матрицы — элементы матрицы A^{-1} являются многочленами от элементов матрицы A , деленными на $\det A \neq 0$). Таким образом, $g \in C^{k+1}$. $\square$

9.2.2. Теорема о неявной функции. Эта теорема тесно связана с теоремой об обратной функции. Ее легко вывести из теоремы об обратной функции, а поэтому она является косвенным приложением принципа сжимающих отображений. Кроме того, как мы увидим в следующем пункте, из нее следует важный результат относительно самого принципа сжимающих отображений, касающийся зависимости неподвижной точки сжимающего отображения от самого отображения (см. также рис. 2.2.3).

Подобно теореме об обратной функции, теорема о неявной функции переносит информацию относительно линейной части отображения на само отображение. Чтобы понять, как это происходит, рассмотрим задачу, которую решает теорема о неявной функции в случае линейного отображения. Предположим, что $A: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ — линейное отображение, и запишем его в виде $A = (A_1, A_2)$, где $A_1: \mathbb{R}^n \rightarrow \mathbb{R}^n$ и $A_2: \mathbb{R}^m \rightarrow \mathbb{R}^n$ — линейные отображения. Выберем $k \in \mathbb{R}^m$ и попробуем найти такое $h \in \mathbb{R}^n$, что $A(h, k) = 0$. Чтобы увидеть, когда это можно сделать, перепишем это равенство в виде $A_1 h + A_2 k = 0$ и заключим, что если отображение A_1 обратимо, то

$$A(h, k) = 0 \Leftrightarrow h = -(A_1)^{-1} A_2 k. \quad (9.2.2)$$

Можно интерпретировать этот результат, сказав, что уравнение $A(h, k) = 0$ неявно определяет такое отображение $h = Lk$, что $A(Lk, k) = 0$. Теорема о неявной функции говорит, что если это утверждение справедливо для линейной части отображения, то оно справедливо и для самого отображения: при некоторых условиях, соответствующих обратимости отображения A_1 , уравнение $f(x, y) = 0$ неявно определяет такое отображение $x = g(y)$, что $f(g(y), y) = 0$. Чтобы сформулировать эти условия на отображение $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$, мы запишем $Df = (D_1f, D_2f)$, где $D_1f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ и $D_2f: \mathbb{R}^m \rightarrow \mathbb{R}^n$, аналогично тому, как это было сделано выше.

Теорема о неявной функции гласит, что если мы можем решить данное уравнение для частного значения параметра, то имеется решение и для близлежащих значений параметра.

Теорема 9.2.3 (о неявной функции). Пусть $O \subset \mathbb{R}^n \times \mathbb{R}^m$ — открытое множество и $f: O \rightarrow \mathbb{R}^n$ — отображение класса C^r . Если имеется такая точка $(a, b) \in O$, что $f(a, b) = 0$ и отображение $D_1f(a, b)$ обратимо, тогда найдутся такая открытая окрестность $U \subset O$ точки (a, b) и такая открытая окрестность $V \subset \mathbb{R}^m$ точки b , что для каждого $y \in V$ существует единственная точка $x =: g(y) \in \mathbb{R}^n$, для которой $(x, y) \in U$ и $f(x, y) = 0$. Более того, g является отображением класса C^r и $Dg(b) = -(D_1f(a, b))^{-1} D_2f(a, b)$.

Доказательство. Отображение $F(x, y) := (f(x, y), y): O \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ принадлежит классу C^r , и если $A = Df(a, b)$, то $DF(a, b)(h, k) = (A(h, k), k)$ (по правилу дифференцирования сложной функции). Это выражение равно нулю, только если $k = 0$ и $A(h, k) = 0$, следовательно, $(h, k) = 0$ в соответствии с формулой (9.2.2). Поэтому отображение DF обратимо, и по теореме об обратной функции (теорема 9.2.2) имеются такая открытая окрестность $U \subset O$ точки (a, b) и такая открытая окрестность $W \subset \mathbb{R}^n \times \mathbb{R}^m$ точки $(0, b)$, что отображение $F: U \rightarrow W$ обратимо, причем обратное отображение $G = F^{-1}: W \rightarrow U$ принадлежит классу C^r . Таким образом, для любого $y \in V := \{y \in \mathbb{R}^m: (0, y) \in W\}$ существует такая точка $x =: g(y) \in \mathbb{R}^n$, что $(x, y) \in U$ и $F(x, y) = (0, y)$, т. е. $f(x, y) = 0$.

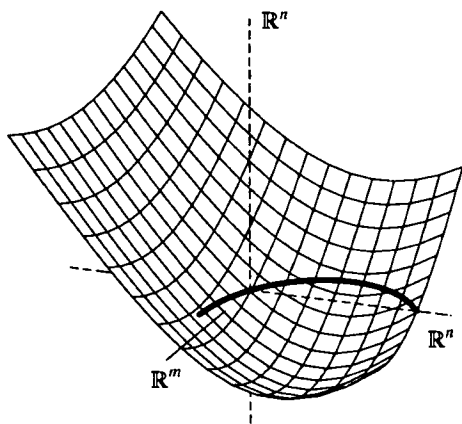


Рис. 9.2.3. Теорема о неявной функции

Далее, $(g(y), y) = (x, y) = G(0, y)$, и, следовательно, отображение g принадлежит классу C^r . Чтобы найти $Dg(b)$, положим $\gamma(y) := (g(y), y)$. Тогда $f(\gamma(y)) \equiv 0$, и, следовательно, $Df(\gamma(y))D\gamma(y) = 0$ по правилу дифференцирования сложной функции. Для $y = b$ из этого следует, что $D_1f(a, b)Dg(b) + D_2f(a, b) = Df(a, b)D\gamma(b) = 0$, и это завершает доказательство. $\square$

9.2.3. Гладкий принцип сжимающих отображений. Возвращаясь к динамике, мы применим гладкость неявной функции g к принципу сжимающих отображений и покажем, что неподвижная точка сжимающего отображения гладко зависит от самого этого отображения (см. рис. 2.2.3). Чтобы выразить эту зависимость, мы будем считать, что наши сжимающие отображения зависят от параметра.

Теорема 9.2.4. *Предположим, что $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ — отображение класса C^r и существует такое $\lambda < 1$, что $d(f(x, y), f(x', y)) \leq \lambda d(x, x')$ для $x, x' \in X$. Тогда для каждого $y \in Y$ имеется единственная неподвижная точка $g(y)$ отображения $x \mapsto f(x, y)$ и отображение g принадлежит классу C^r .*

Доказательство. Существование неподвижной точки $g(y)$ следует из принципа сжимающих отображений. Запишем теперь $F(x, y) := f(x, y) - x$ и заметим, что эта функция из класса C^r удовлетворяет условию теоремы о неявной функции (теорема 9.2.3): она равна нулю в точке $(a, b) = (g(y), y)$ (y — произвольная точка) и $\|D_1Fv\| = \|D_1fv - v\| \geq \|v\| - \|D_1fv\| \geq (1 - \lambda)\|v\| > 0$ для $v \neq 0$, так что отображение D_1F обратимо. Таким образом, $g \in C^r$. $\square$

Замечание 9.2.5. Вместо области $\mathbb{R}^n \times \mathbb{R}^m$ можно также взять произведение $A \times O$, где множество $O \subset \mathbb{R}^m$ открыто и A является, скажем, замыканием открытого множества. (Чтобы применить принцип сжимающих отображений, необходимо иметь замкнутое множество, но еще необходимо иметь достаточно хорошее множество, чтобы можно было продифференцировать отображение r раз.)

Замечание 9.2.6. Предположим, что отображение f_λ гладко зависит от λ и $f := f_0$, как в предложении 2.2.20. Покажите, что имеется гладкое семейство $\lambda \mapsto x_\lambda$, где x_0 определяется в соответствии с предложением 2.2.20 и $f_\lambda(x_\lambda) = x_\lambda$.

§ 9.3. Устойчивость трансверсальных неподвижных точек

Для неподвижной точки сжимающего отображения одновременно характерны два вида устойчивости. Как притягивающая неподвижная точка она асимптотически устойчива. Из предложений 2.2.20 и 2.6.14 (а также теоремы 9.2.4) следует, что она также устойчива при возмущениях самого отображения, т. е. возмущения отображения имеют в близлежащей окрестности единственную неподвижную точку. Это важное свойство устойчивости локальной динамики, и сейчас мы, используя принцип сжимающих отображений, опишем общие условия, при которых справедливо подобное утверждение. Этот результат служит примером непосредственного и простого использования в динамике принципа сжимающих отображений и теоремы о неявной функции, причем принцип сжимающих отобра-

ражений применяется к некоторому вспомогательному отображению в том же самом пространстве.

Напомним, что два C^1 -отображения f и g считаются близкими в смысле метрики в C^1 , если величина $|f - g| + \|Df - Dg\|$ равномерно мала.

Предложение 9.3.1. *Если p — периодическая точка периода t для отображения f класса C^1 и дифференциал Df_p^m не имеет собственного значения 1 (в этом случае p называется трансверсальной периодической точкой), тогда для каждого отображения g , достаточно близкого к f в C^1 , имеется единственная периодическая точка периода t , близкая к точке p .*

Заметим, что в размерности один условие на производную попросту означает, что она не равна единице. Соответственно в примере простейшей бифуркации из п. 2.3.2 (см. рис. 2.3.2) единственная неподвижная точка появляется или исчезает в точности тогда, когда имеется касание с диагональю, т. е. производная отображения равна 1. Это показано на рис. 9.3.1. Ось независимой переменной

Рис. 9.3.1. Устойчивость неподвижной точки

направлена вправо, вертикальная ось — ось значений, а на оси, направленной от нас, отмечаются значения параметра, в соответствии с которым изменяется отображение. Плоскость представляет диагональ для различных параметров, и графики возмущенных отображений объединяются в поверхность, которая пересекает диагонали в совокупности неподвижных точек.

Доказательство.

Пространство. Мы определяем сжимающее отображение в окрестности точки p .

Отображение. Введем локальные координаты вблизи точки p , принимая p за начало координат. В этих координатах Df_0^m становится матрицей. Поскольку 1 не является собственным значением, отображение $F = f^m - \text{Id}$, определенное локально в этих координатах, локально обратимо согласно теореме об обратной функции (теорема 9.2.2). Пусть теперь g — отображение, близкое к f

в C^1 . Вблизи 0 можно записать $g^m = f^m - H$, где отображение H мало вместе с своими первыми производными. Неподвижную точку для g^m можно найти из уравнения $x = g^m(x) = (f^m - H)(x) = (F + \text{Id} - H)(x)$, или $(F - H)(x) = 0$, т. е.

$$x = F^{-1}H(x).$$

Свойство сжатия. Поскольку отображение F^{-1} имеет ограниченные производные и H имеет очень малые первые производные, можно показать, что $F^{-1}H$ — сжимающее отображение. Точнее, пусть $\|\cdot\|_0$ обозначает норму в C^0 , $\|dF^{-1}\|_0 = L$ и

$$\max(\|H\|_0, \|dH\|_0) \leq \varepsilon.$$

Тогда, поскольку $F(0) = 0$, мы получаем $\|F^{-1}H(x) - F^{-1}H(y)\| \leq \varepsilon L\|x - y\|$ для всех x, y , близких к 0, и $\|F^{-1}H(0)\| \leq L\|H(0)\| \leq \varepsilon L$, следовательно, $\|F^{-1}H(x)\| \leq \|F^{-1}H(x) - F^{-1}H(0)\| + \|F^{-1}H(0)\| \leq \varepsilon L\|x\| + \varepsilon L$. Таким образом, если $\varepsilon \leq \frac{R}{L(1+R)}$, диск $X = \{x: \|x\| \leq R\}$ отображается под действием $F^{-1}H$ в себя и отображение $F^{-1}H: X \rightarrow X$ является сжимающим. Из принципа сжимающих отображений следует, что существует единственная неподвижная точка этого отображения в диске X , которая является единственной неподвижной точкой для g^m вблизи 0. $\square$

Замечание 9.3.2. Легко показать, что трансверсальная неподвижная точка изолирована.

Замечание 9.3.3. Если f является отображением класса C^1 с гиперболической неподвижной точкой p , т. е. $Df|_p$ не имеет собственных значений на единичной окружности, и отображение g достаточно C^1 -близко к f , тогда g имеет единичную неподвижную точку вблизи p и эта неподвижная точка является гиперболической неподвижной точкой отображения g .

§ 9.4. Решения дифференциальных уравнений

Дифференциальные уравнения естественным образом связаны с задачами динамики. Использование дифференциальных уравнений в естественных науках основано на том, что они описывают систему детерминированным образом. Это означает, что для любого допустимого начального условия существует решение, которое описывает дальнейшее развитие системы от этого начального состояния. Кроме того, детерминированность означает, что решение должно быть единственным, — если бы это было не так, тогда начальные данные не определяли бы однозначно эволюцию системы и модель нельзя было бы использовать для предсказания поведения системы.

Сейчас мы исследуем основной результат о существовании и единственности решений. Конечно же, в дальнейшем удобно считать существование и единственность решений само собой разумеющимися, но все же здесь будет интересно получить этот результат как еще одно приложение принципа сжимающих отображений. Такое доказательство существования решений имеет то преимущество, что существуют красивые и полезные следствия из гладкой зависимости неподвижной

точки сжимающего отображения от самого отображения, касающиеся поведения решений дифференциального уравнения при изменении начального условия: небольшие изменения в начальном условии лишь слегка изменяют само решение.

9.4.1. Равномерный случай. Использование принципа сжимающих отображений в данной ситуации называется *итерационным процессом Пикара*. Здесь мы впервые используем принцип сжимающих отображений в функциональном пространстве. Идея заключается в том, чтобы записать дифференциальное уравнение с данным начальным условием как интегральное уравнение и затем применить интеграл к непрерывным функциям, которые могут быть решениями. Эта операция окажется сжимающим отображением, а следовательно, наши приближения будут стремиться к решению.

Теорема 9.4.1. *Предположим, что $f: \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ — непрерывная функция, которая удовлетворяет условию Липшица для $y \in \mathbb{R}^n$ с константой M . Для любой пары $(a, b) \in \mathbb{R} \times \mathbb{R}^n$ и $\delta < 1/M$ имеется единственное такое решение $\varphi_{a,b}: (a - \delta, a + \delta) \rightarrow \mathbb{R}^n$ дифференциального уравнения $\dot{y} = f(t, y)$, что $\varphi_{a,b}(a) = b$.*

Доказательство.

Пространство. Ниже мы построим сжимающее отображение, определенное на пространстве непрерывных функций (кандидатов на роль решения). Условие на функцию f означает, что $\|f(t, y') - f(t, y)\| \leq M\|y' - y\|$ для $t \in \mathbb{R}$, $y, y' \in \mathbb{R}^n$. Рассмотрим множество непрерывных функций $\varphi: [a - \delta, a + \delta] \rightarrow \mathbb{R}^n$ и положим $\|\varphi\| := \max_{|t-a| \leq \delta} \|\varphi(t)\|$. Тем самым мы определим полное метрическое пространство в силу теоремы А.1.13.

Отображение. Применим принцип сжимающих отображений к оператору Пикара, который определяется соотношением

$$\mathcal{P}_{a,b}(\varphi)(t) := b + \int_a^t f(x, \varphi(x)) dx.$$

Свойство сжатия. Заметим, что

$$\|\mathcal{P}_{a,b}(\varphi_1) - \mathcal{P}_{a,b}(\varphi_2)\| = \max_{|t-a| \leq \delta} \left\| \int_0^t f(x, \varphi_1(x)) - f(x, \varphi_2(x)) dx \right\| \leq M\delta \|\varphi_1 - \varphi_2\|,$$

т. е. $\mathcal{P}_{a,b}$ — сжимающее отображение, и, следовательно, оно имеет единственную неподвижную точку. Остается показать, что неподвижные точки отображения $\mathcal{P}_{a,b}$ являются решениями уравнения $\dot{y} = f(t, y)$, где $\varphi(a) = b$ (и наоборот). Для этого продифференцируем равенство $\varphi_{a,b}(t) = b + \int_a^t f(x, \varphi_{a,b}(x)) dx$ по t и в соответствии с известной из курса математического анализа теоремой получим $\dot{\varphi}_{a,b}(t) = f(t, \varphi_{a,b}(t))$. Очевидно, что неподвижные точки оператора $\mathcal{P}_{a,b}$ удовлетворяют условию $\varphi_{a,b}(a) = b$. Чтобы убедиться в обратном, т. е. доказать, что решения являются неподвижными точками, подставим решение в условие для неподвижной точки и заметим, что подынтегральное выражение равно $\dot{\varphi}$, откуда мы по той же теореме получаем неподвижную точку. Таким образом,

из существования и единственности неподвижной точки следует существование и единственность решения. $\square$

На самом деле решения в этом случае определены для всех значений переменной t (см. предложение 9.4.7), в чем можно убедиться, объединив друг с другом полученные локальные решения.

Пример 9.4.2. Можно в явном виде выполнить эту итерационную процедуру для дифференциального уравнения $\dot{y} = y$, $y(0) = 1$, взяв в качестве начального приближения $y_0(t) = 1$. Тогда $y_1(t) = 1 + \int_0^t y(x) dx = 1 + \int_0^t dx = 1 + t$ и $y_2(t) = 1 + \int_0^t (1 + x) dx = 1 + t + t^2/2$. По индукции можно заключить, что $y_k(t) = \sum_{n=0}^k t^n/n!$, а следовательно, $y(t) = \sum_{n=0}^{\infty} t^n/n! = e^t$.

Пикар изобрел эту схему задолго до того, как был сформулирован и доказан принцип сжимающих отображений, и использование этого метода последовательного приближения было основано на доказательстве того, что погрешность убывает достаточно быстро.

9.4.2. Неравномерный случай. Может случиться так, что константа Липшица для правой части дифференциального уравнения зависит от t и что правая часть даже определена не для всех значений переменной времени и не на всем пространстве $\mathbb{R}^n$. В этом случае все же имеется результат, подобный теореме 9.4.1, но необходимо наложить условие, что решения не покидают область определения правой части уравнения.

Теорема 9.4.3. *Предположим, что $I \in \mathbb{R}$ — интервал, множество $O \subset \mathbb{R}^n$ открыто и $f: I \times O \rightarrow \mathbb{R}^n$ — непрерывная функция, которая удовлетворяет условию Липшица с константой M по переменной $y \in O$ для любого фиксированного $t \in I$. Тогда для любой пары $(a, b) \in I \times O$ найдется такое $\delta > 0$, что существует единственное решение $\varphi_{a,b}: (a - \delta, a + \delta) \rightarrow \mathbb{R}^n$ дифференциального уравнения $\dot{y} = f(t, y)$, удовлетворяющее условию $\varphi_{a,b}(a) = b$.*

Доказательство.

Пространство. Условие на f означает, что $\|f(t, y') - f(t, y)\| \leq M\|y' - y\|$ для $t \in I$, $y, y' \in O$. Возьмем замкнутое ограниченное подмножество K множества O и отрезок $I' \subset I$, содержащий точку a . Допустим, что $B > \sup_{t \in I', x \in K} \|f(t, x)\|$, и выберем такое $\delta \in (0, 1/M)$, что $[a - \delta, a + \delta] \subset I'$ и шар $B(b, B\delta)$ содержится в K . Теперь рассмотрим множество $\mathcal{C}$ таких непрерывных функций $\varphi: [a - \delta, a + \delta] \rightarrow O$, что $\|\varphi - b\| < B\delta$, где норма снова определяется соотношением $\|\varphi\| := \max_{|t-a| \leq \delta} \|\varphi(t)\|$. Множество $\mathcal{C}$ является замкнутым подмножеством полного метрического пространства всех непрерывных функций на $[a - \delta, a + \delta]$ (с этой нормой), а следовательно, само полно.

Отображение. Оператор Пикара, как и раньше, определяется равенством

$$\mathcal{P}_{a,b}(\varphi)(t) := b + \int_a^t f(x, \varphi(x)) dx.$$

Тогда $\|\mathcal{P}_{a,b}(\varphi) - b\| \leq \max_{|t-a| \leq \delta} \left\| \int_a^t f(x, \varphi(x)) dx \right\| < B\delta$, так что $\mathcal{P}_{a,b}(\varphi) \in \mathcal{C}$ для $\varphi \in \mathcal{C}$, т. е. оператор $\mathcal{P}_{a,b}$ определен корректно.

Свойство сжатия. Поскольку

$$\|\mathcal{P}_{a,b}(\varphi) - \mathcal{P}_{a,b}(\varphi')\| = \max_{|t-a| \leq \delta} \left\| \int_0^t f(x, \varphi(x)) - f(x, \varphi'(x)) dx \right\| \leq M\delta \|\varphi - \varphi'\|,$$

оператор $\mathcal{P}_{a,b}$ является сжимающим на $\mathcal{C}$ и, следовательно, имеет единственную неподвижную точку. Как и прежде, неподвижные точки соответствуют решениям дифференциального уравнения. $\square$

Замечание 9.4.4. Заметим, что здесь мы получаем только локальные решения. Глобальные решения можно получить, объединяя локальные решения; в силу единственности любые два локальных решения должны совпадать на пересечении областей их определения. На самом деле единственное препятствие для продолжения решений состоит в том, что они могут натолкнуться на границу множества O , вне которого наше дифференциальное уравнение не определено. Мы детально рассмотрим эту ситуацию в п. 9.4.7.

9.4.3. Непрерывная зависимость. Поскольку оператор $\mathcal{P}_{a,b}$ непрерывно зависит от a и b и $\mathcal{P}_{a,b'}(\mathcal{C}) \subset \mathcal{C}$ для всех точек b' , достаточно близких к b , решения непрерывно зависят от начального значения b в силу предложения 2.6.14.

Предложение 9.4.5. В условиях теоремы 9.4.3 решения непрерывно зависят от начального значения, т. е. для данного $\varepsilon > 0$ существует такое $\eta > 0$, что если $\|b' - b\| < \eta$, то выполнено условие $\max_{|t-a| \leq \delta} \|\varphi_{a,b'}(t) - \varphi_{a,b}(t)\| < \varepsilon$.

Доказательство. Очевидно, что мы должны выбрать такое η , что $B(b', B\delta) \subset K$ (см. начало предыдущего доказательства) при условии $\|b' - b\| < \eta$, чтобы функция $\varphi_{a,b'}$ была определена для $|t - a| < \delta$. Если это так, то доказательство предложения (возможно, для меньшего η) сводится к переформулировке свойства непрерывной зависимости неподвижной точки сжимающего отображения от параметра, в этом случае относительно нормы $\|\varphi\| := \max_{|t-a| \leq \delta} \|\varphi(t)\|$. $\square$

9.4.4. Гладкая зависимость. Отображение $\mathcal{P}: \mathcal{C} \times \mathbb{R} \times O \rightarrow \mathcal{C}$ принимает значения в векторном пространстве, в котором определена операция дифференцирования (см. определение А.2.1). Это отображение линейно (а значит, гладко) зависит от $b \in O$, а зависимость от $\varphi \in \mathcal{C}$ осуществляется через f и, следовательно, имеет ту же гладкость, что и f . Для того чтобы убедиться в этом, рассмотрим первую производную. Из формулы Лагранжа следует, что

$$\begin{aligned} \mathcal{P}_{a,b}(\varphi)(t) - \mathcal{P}_{a,b}(\psi)(t) &= \int_a^t f(x, \varphi(x)) dx - \int_a^t f(x, \psi(x)) dx = \\ &= \int_a^t f(x, \varphi(x)) - f(x, \psi(x)) dx = \int_a^t (\partial f / \partial y)(x, c_x)(\varphi(x) - \psi(x)) dx \approx \\ &\approx \int_a^t (\partial f / \partial y)(x, \varphi(x))(\varphi(x) - \psi(x)) dx. \end{aligned}$$

Таким образом, первая производная определяется соотношением $D\mathcal{P}_{a,b}(\varphi)(\eta)(t) = \int_a^t (\partial f / \partial y)(x, \varphi(x)) \eta(x) dx$. Из следствия 2.2.15 мы получаем такой результат.

Предложение 9.4.6. *Если в условиях предложения 9.4.5 функция f принадлежит классу C^r , то решения принадлежат классу C^{r+1} и C^r -гладко зависят от начального значения b , т. е. отображение $b \mapsto \varphi_{a,b}(a+t)$ принадлежит классу C^r для всех $t \in (-\delta, \delta)$.*

Тот факт, что сами решения принадлежат классу C^{r+1} , индуктивно следует из дифференциального уравнения, в соответствии с которым функция $\dot{y}$ принадлежит классу C^k , если y и f принадлежат классу C^k .

9.4.5. Отсутствие решений и неединственность. Чтобы увидеть, что все условия действительно необходимы, рассмотрим рис. 9.4.1. На этом рисунке

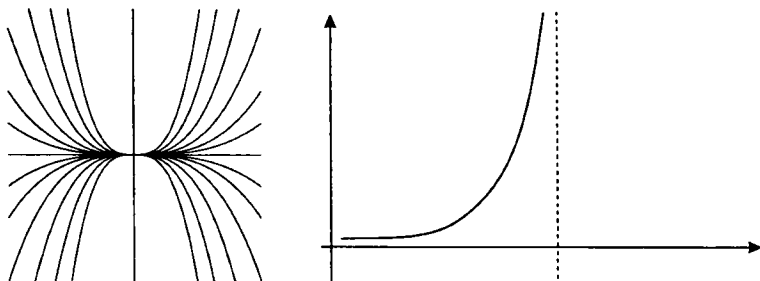


Рис. 9.4.1. Сложности, возникающие в дифференциальных уравнениях

изображены решения $x = ct^2$ уравнения $t\dot{x} = 2x$, для которого единственность нарушается при начальном условии $a = b = 0$, а существование нарушается при любом начальном условии $a = 0, b \neq 0$. Справа изображена ситуация, показывающая, что решения не всегда продолжаются на все значения t , хотя правая часть уравнения $f(t, x)$ определена для всех t : решения $x = -1/(t+c)$ уравнения $\dot{x} = x^2$ имеют особенности при конечных значениях t . Существование решений можно доказать, используя лишь непрерывность правой части дифференциального уравнения. Однако возможное нарушение единственности показывает, что непрерывную зависимость от начального значения нельзя доказать без использования условия Липшица.

9.4.6. Продолжение решений. По причинам, о которых говорилось в п. 3.2.6 и которые получают свое объяснение в предложении 9.4.11, мы ограничимся рассмотрением дифференциальных уравнений вида $\dot{x} = f(x)$, т. е. дифференциальных уравнений, правая часть которых не зависит от времени. (Они называются *автономными* дифференциальными уравнениями, а правая часть тогда называется *векторным полем*, порождающим поток.) Физически это отражает то обстоятельство, что мы считаем законы природы неизменными. Нам бы хотелось не беспокоиться о том, что решения могут быть определены только до некоторого момента. Обычно это действительно так.

Предложение 9.4.7. Если функция f определена на всем пространстве $\mathbb{R}^n$ и липшицева, то решения уравнения $\dot{x} = f(x)$ определены для всех t .

Доказательство. Как следует из теоремы 9.4.1, для любого начального условия $y(0) = b$ имеется решение $\varphi_{0,b}: [-\delta, \delta] \rightarrow \mathbb{R}^n$, удовлетворяющее начальному условию $\varphi_{0,b}(0) = b$. Для начального условия $\varphi_{0,b}(\delta) =: b'$ имеется решение (см. рис. 9.4.2) $\varphi_{\delta, \varphi_b(\delta)}$ на отрезке $[0, 2\delta]$, т. е.

$$\varphi_{\delta, \varphi_b(\delta)}(t) = f(\varphi_{\delta, \varphi_b(\delta)}) \quad \text{и} \quad \varphi_{\delta, \varphi_b(\delta)}(0) = b.$$

В то же самое время $\varphi_{0,b}(t) = f(\varphi_{0,b})$ и $\varphi_{0,b}(0) = b$, так что $\varphi_{\delta, \varphi_b(\delta)}(t) = \varphi_{0,b}(t)$ для $t \in [0, \delta]$ в силу единственности. Поэтому имеется единственное решение на

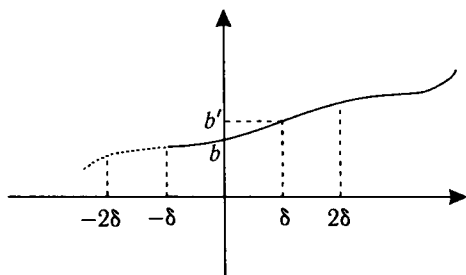


Рис. 9.4.2. Продолжение решений

отрезке $[-\delta, 2\delta]$. Осуществляя подобное продолжение от точки $-\delta$, мы получим решение на отрезке $[-2\delta, 2\delta]$, которое, в свою очередь, может быть продолжено на отрезок $[-3\delta, 3\delta]$, и т. д. Таким образом, решения определены на всей оси времени, независимо от начального условия. $\square$

Применяя предложение 9.4.5 примерно T/δ раз, мы получим такой результат.

Предложение 9.4.8. Решения непрерывно зависят от начального значения для любого конечного промежутка времени, т. е. для заданных $T, \epsilon > 0$ существует такое $\delta > 0$, что если $\|b' - b\| < \delta$, то выполнено неравенство

$$\max_{|t-a| \leq T} \|\varphi_{a,b'}(t) - \varphi_{a,b}(t)\| < \epsilon.$$

9.4.7. Поток. Изучим теперь отображения, порождаемые решениями дифференциальных уравнений. В первой лемме содержится утверждение, что эволюция системы в течение промежутка времени данной длины не зависит от начального момента времени. Затем мы заключаем, что эта эволюция определяет *однопараметрическую группу* некоторых дифференцируемых отображений.

Лемма 9.4.9. Отображение $\varphi_a^t: b \mapsto \varphi_{a,b}(a+t)$ из предложения 9.4.6 определено на всем пространстве $\mathbb{R}^n$ для любого значения t и является отображением класса C^1 , если f принадлежит этому классу. Оно также не зависит от a .

Доказательство. Предложение 9.4.7 показывает, что отображение ϕ_a^t определено на $\mathbb{R}^n$ для любого t , а в соответствии с предложением 9.4.6 оно имеет тот же класс гладкости, что и f .

Для данных $a, a' \in \mathbb{R}$ и $b \in \mathbb{R}^n$ рассмотрим решения $\varphi_{a,b}$ и $\varphi_{a',b}$ дифференциального уравнения, удовлетворяющие условиям $\varphi_{a,b}(a) = b$ и $\varphi_{a',b}(a') = b$. Тогда $\phi_a^t(b) = \varphi_{a,b}(a+t)$ и $\phi_{a'}^t(b) = \varphi_{a',b}(a'+t)$. Мы должны показать, что эти решения совпадают. Определив $\psi_1(t) := \varphi_{a,b}(t+a)$ и $\psi_2(t) := \varphi_{a',b}(t+a')$, мы получим, что

$$\dot{\psi}_1(t) = f(\psi_1(t)), \quad \psi_1(0) = b \quad \text{и} \quad \dot{\psi}_2(t) = f(\psi_2(t)), \quad \psi_2(0) = b.$$

В силу единственности $\phi_{a'}^t(b) = \varphi_{a',b}(t+a') = \psi_2(t) = \psi_1(t) = \varphi_{a,b}(a+t) = \phi_a^t(b)$. $\square$

В дальнейшем мы будем опускать нижний индекс a и с этого момента будем писать $\phi^t(b) = \varphi_{a,b}(a+t)$ (полагая по умолчанию $a=0$).

Определение 9.4.10. Семейство отображений $(\phi^t)_{t \in \mathbb{R}}$, для которого отображение $(t, x) \mapsto \phi^t(x)$ принадлежит классу C^r , называется C^r -*поток*ом, если $\phi^{s+t} = \phi^s \circ \phi^t$ для всех $s, t \in \mathbb{R}$.

Это «групповое свойство» справедливо в нашей ситуации.

Предложение 9.4.11. Дифференциальное уравнение $\dot{x} = f(x)$, правая часть которого $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ — функция класса гладкости C^r , причем норма $\|Df\|$ ограничена, определяет C^r -поток на $\mathbb{R}^n$.

Доказательство. Для заданного $t \in \mathbb{R}$ функции $\psi_1(s) := \varphi_{0,b}(s+t)$ и $\psi_2(s) := \varphi_{0,\varphi_{0,b}(t)}(s)$ являются решениями нашего дифференциального уравнения и $\psi_2(0) = \varphi_{0,\varphi_{0,b}(t)}(0) = \varphi_{0,b}(t) = \psi_1(0)$, так что $\psi_1 = \psi_2$ в силу единственности. Следовательно,

$$\phi^s \circ \phi^t(b) = \phi^s(\varphi_{0,b}(t)) = \varphi_{0,\varphi_{0,b}(t)}(s) = \varphi_{0,b}(s+t) = \phi^{s+t}(b).$$

Положив $s = -t$, мы, в частности, получим, что отображение ϕ^t обратимо, причем обратным для него является ϕ^{-t} . Таким образом, эти отображения ϕ^t являются диффеоморфизмами класса C^r . $\square$

Понятие гладкого потока является центральным в теории динамических систем с непрерывным временем. Оно связывает динамику и дифференциальные уравнения. Это понятие позволяет описать динамику как изучение асимптотического поведения (дискретных или непрерывных) однопараметрических групп преобразований.

В п. 2.4.1 приведено полное описание динамики потока, порожденного дифференциальным уравнением $\dot{x} = f(x)$ на прямой, где f — липшицева функция: имеется замкнутое множество неподвижных точек, а поток монотонен на каждом интервале из дополнения к этому множеству, причем все орбиты стремятся к одной концевой точке, а под действием отрицательного потока — к другой.

При изменении величины правой части без изменения направления орбиты сохраняются, изменяется только скорость вдоль них.

Определение 9.4.12. Потоки, порожденные уравнениями $\dot{x} = f(x)$ и $\dot{x} = a(x)f(x)$ для некоторой непрерывной не обращающейся в нуль скалярной функции a , называются связанными *заменой времени*.

§ 9.5. Гиперболичность

Мы начинаем этот параграф с красивого примера линеаризации, который непосредственно связан с описанием качественных особенностей нелинейного отображения в терминах его линейной части.

9.5.1. Гиперболические неподвижные точки. Напомним наше описание динамики гиперболического линейного отображения в $\mathbb{R}^2$, приведенное в § 3.1. Гиперболическое линейное отображение в $\mathbb{R}^2$ — это линейное отображение с собственными значениями $\lambda \in (-1, 1)$ и $\mu \notin [-1, 1]$. Ненулевой собственный вектор v , соответствующий значению λ , порождает прямую E^s , проходящую через начало координат в $\mathbb{R}^2$, и аналогично получается прямая E^u для собственного вектора w , соответствующего значению μ .

Эти две прямые служат основой для восстановления динамики линейного отображения. На E^s отображение ведет себя как линейное сжатие, а на E^u обратное к нему отображение является сжатием. Орбиты точек, не лежащих ни на одной из этих прямых, находятся на кривых, асимптотических к полученным устойчивым и неустойчивым прямым.

Рассмотрим теперь дифференцируемое отображение $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ с такой неподвижной точкой x_0 , что $Df(x_0)$ является гиперболическим линейным отображением. Замечательно то, что наличие сжимающего и растягивающего направлений для линейного отображения $Df(x_0)$ имеет полную аналогию для самого отображения f , за исключением того, что вместо прямых в этом случае мы получаем кривые W^s и W^u . Каждая из этих кривых инвариантна, т. е. $f(W^s) = W^s$ и $f(W^u) = W^u$, и f является сжатием на W^s , а f^{-1} — сжатием на W^u (см. рис. 9.5.1). Строго говоря, это утверждение не совсем корректно, но оно справедливо в окрестности неподвижной точки. Однако чтобы избежать сложностей, связанных с определением размера этой окрестности, мы докажем результат, который формулируется несколько иначе (теорема 9.5.2). Он содержит лишь ту часть общего утверждения, которая касается сжимающейся кривой. Но, применяя этот результат к обратному отображению, мы получаем также и растягивающуюся кривую для f .

Заметим также, что сжимающаяся прямая E^s для гиперболического линейного отображения — это в точности множество точек, последовательные образы которых образуют ограниченную последовательность, т. е. положительная полуорбита каждой из которых ограничена. Соответственно отображение с гиперболической неподвижной точкой (т. е. такое отображение, для которого дифференциал является гиперболическим линейным отображением) имеет гладкую кривую, проходящую через неподвижную точку и состоящую в точности из тех точек, итерации которых остаются достаточно близко от неподвижной точки. Мы докажем этот результат для отображений в $\mathbb{R}^2$, но нетрудно видеть, что то же

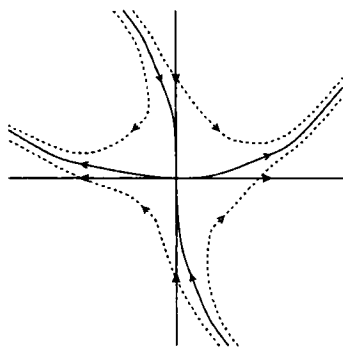


Рис. 9.5.1. Устойчивые и неустойчивые многообразия

самое рассуждение работает и в более высоких размерностях, если по-другому проинтерпретировать обозначения. Как и в § 9.4, мы докажем это утверждение, применяя принцип сжимающих отображений в надлежащем образом построенном «бесконечномерном» метрическом пространстве.

9.5.2. Теорема об устойчивом многообразии. Рассмотрим функцию $\varphi: \{0\} \times \mathbb{R} \rightarrow \mathbb{R} \times \{0\}$ (т. е. φ задает x как функцию от y). Если φ — функция класса C^1 и $|D\varphi| \leq \gamma$, то мы назовем график этой функции $c \subset \mathbb{R}^2$ *вертикальной γ -кривой*. Если φ задает y как функцию от x , мы говорим о *горизонтальной γ -кривой*. Для наших целей полезно следующее утверждение.

Лемма 9.5.1. *Предположим, что $A: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ — линейное отображение, заданное матрицей $\begin{pmatrix} \chi_u & 0 \\ 0 & \chi_s \end{pmatrix}$, где $0 < \chi_s < \chi_u$. Для $\gamma > 0$, $\varepsilon < \frac{\chi_u - \chi_s}{\gamma + 2 + (1/\gamma)}$ и любого C^1 -отображения $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$, для которого $\|Df - A\| \leq \varepsilon$, отображение, обратное к f , сохраняет вертикальные γ -кривые, т. е. прообраз γ -кривой относительно f сам является γ -кривой.*

Доказательство. Мы будем для удобства использовать норму $\|(x, y)\| := |x| + |y|$ на $\mathbb{R}^2$ и писать $f(x, y) = (f_1(x, y), f_2(x, y))$ и $D_1 := \partial/\partial x$, $D_2 := \partial/\partial y$.

Докажем, что для любой γ -кривой, заданной уравнением $x = c(y)$, можно решить уравнение

$$f_1(x, y) = c(f_2(x, y)), \quad \text{или} \quad 0 = f(x, y) := f_1(x, y) - c(f_2(x, y)),$$

где функция $x = g(y)$ задана неявно, причем $|Dg| \leq \gamma$. Так как прообраз нашей γ -кривой непуст, найдется пара чисел (a, b) , для которой $F(a, b) = 0$. Проверим,

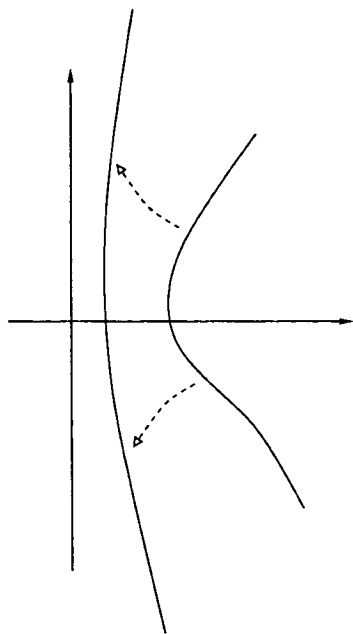


Рис. 9.5.2. Сохранение вертикальных γ -кривых

что $D_1 F \neq 0$. Заметим, что $|D_1 f_2| < \epsilon$, $|D_2 f_1| < \epsilon$ и

$$|D_1 f_1| \geq x_u - |D_1 f_1 - x_u| \geq x_u - \epsilon. \quad (9.5.1)$$

Поэтому

$$|D_1 F| = |D_1 f_1 - Dc \circ f_2 D_1 f_2| \geq x_u - (1 + \gamma)\epsilon > 0.$$

Таким образом, мы можем, по крайней мере локально, решить уравнение для функции $x = g(y)$. Чтобы оценить ее производную, заметим, что

$$|D_2 F| = |D_2 f_1 - Dc \circ f_2 D_2 f_2| \leq \gamma(x_s + \epsilon) - \epsilon = \gamma x_s - (1 + \gamma)\epsilon.$$

Таким образом, $|Dg| < \gamma$ как нам и нужно. Действительно,

$$|Dg| = \left| -\frac{D_2 F}{D_1 F} \right| \leq \frac{\gamma x_s - (1 + \gamma)\epsilon}{x_u - (1 + \gamma)\epsilon} < \gamma.$$

(Чтобы проверить последнее неравенство, достаточно разделить числитель и знаменатель на γ и использовать неравенство $(1 + \gamma)(1 + 1/\gamma)\epsilon < x_u - x_s$.)

Теперь покажем, что функция g определена на $\mathbb{R}$, а не только локально. Из сказанного выше следует, что $|D_2 f_2| \leq x_s + \epsilon$, т. е. отображение f^{-1} растягивает кривые в вертикальном направлении. Для определенности возьмем произвольное значение $y \in \mathbb{R}$. Мы должны показать, что прообраз нашей кривой содержит точку (x, y) . Рассмотрим график функции c на интервале от a до $a + (y - b)$. Ординаты прообраза находятся в пределах от b до $b + (y - b)/(x_s + \epsilon)$ и, в частности, среди них содержится y . Таким образом, мы получаем глобально определенную неявную функцию с необходимой оценкой на производную. $\square$

Заметим, что $\gamma + 2 + (1/\gamma) \geq 4$, так что нам необходимо, чтобы выполнялось неравенство $\epsilon < (x_u - x_s)/4$.

Теперь мы можем доказать результат относительно устойчивых кривых.

Теорема 9.5.2 (об устойчивом многообразии). *Предположим, что $A: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ — линейное отображение, заданное матрицей $\begin{pmatrix} x_u & 0 \\ 0 & x_s \end{pmatrix}$, где $0 < x_s < 1 < x_u$. Тогда существует такое $\epsilon > 0$, что для любого C^r -отображения $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$, для которого $\|Df - A\| \leq \epsilon$, множество точек с ограниченными положительными полуорбитами является графиком функции y класса C^r .*

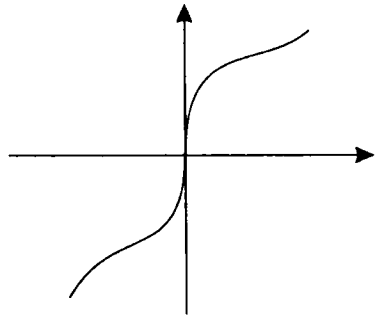


Рис. 9.5.3. Устойчивое многообразие

Доказательство.

Пространство. Мы определим сжимающее отображение на пространстве l^∞ ограниченных последовательностей $(x_n)_{n \in \mathbb{N}_0}$ с нормой $\|(x_n)_{n \in \mathbb{N}_0}\|_\infty := \sup_{n \in \mathbb{N}_0} |x_n|$.

(Оно является полным метрическим пространством согласно теореме А.1.14.)

Отображение. Возьмем такое $\epsilon > 0$, при котором лемма 9.5.1 применима к f^{-1} и A^{-1} , в предположении, что x и y меняются местами, т. е. будем считать, что ϵ настолько мало, что f сохраняет горизонтальные γ -кривые для

некоторого γ . Для данного $y \in \mathbb{R}$ рассмотрим последовательные образы $f^n(L_y)$ прямой $L_y := \{(x, y) : x \in \mathbb{R}\}$ (они являются горизонтальными γ -кривыми класса гладкости C^r). Для любого заданного $x \in \mathbb{R}$ и $n \in \mathbb{N}$ имеется единственное такое $z \in \mathbb{R}$, что $(x, z) \in f^n(L_y)$, так что мы можем сопоставить точке x значение $x' \in \mathbb{R}$, выбирая такое $y' \in \mathbb{R}$, что $(x', y') \in f^{n-1}(L_y)$ и $f(x', y') = (x, y)$. Таким образом, мы задаем отображение $\mathcal{F}_y : l^\infty \rightarrow l^\infty$, полагая $\mathcal{F}_y((x_n)_{n \in \mathbb{N}_0}) = (x'_n)_{n \in \mathbb{N}_0}$ (см. рис. 9.5.4).

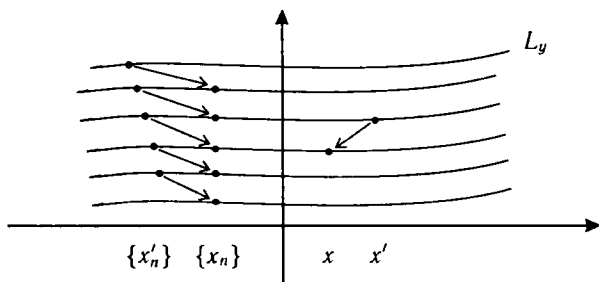


Рис. 9.5.4. Доказательство теоремы 9.5.2

Свойство сжатия. Каждая из этих последовательностей соответствует ограниченной последовательности точек на плоскости (мы просто добавим вторые координаты y_n таким образом, чтобы получить $(x_n, y_n) \in f^n(L_y)$). Отображение $\mathcal{F}_y$ есть действие отображения f на таких последовательностях (первая точка удаляется, а остальные перенумеровываются). Отсюда следует, что неподвижная точка отображения $\mathcal{F}_y$ — последовательность абсцисс орбиты отображения f с исходной ординатой y , и наоборот, каждая ограниченная полуорбита порождает неподвижную точку отображения $\mathcal{F}_y$. Отображение $\mathcal{F}_y$ является сжимающим, потому что f растягивает абсциссы в силу неравенства (9.5.1), т. е. покомпонентные разности между l^∞ -последовательностями делятся на число $\chi_u - \varepsilon > 1$, и поэтому то же самое применимо к разностям между нормами. Следовательно, имеется единственная неподвижная точка $(g(y), y)$, непрерывно зависящая от $y \in \mathbb{R}$.

Таким образом, график непрерывной функции $g(y)$ порождает множество точек с ограниченными положительными полуорбитами.

Тот факт, что l^∞ является нормированным векторным пространством, позволяет говорить о дифференцируемости (см. определение А.2.1), и, поскольку отображение f и кривые $f^n(L_y)$ принадлежат классу C^r , отображение $\mathcal{F} : ((x_n)_{n \in \mathbb{N}_0}, y) \mapsto \mathcal{F}_y((x_n)_{n \in \mathbb{N}_0})$ принадлежит классу C^r (это можно доказать с помощью довольно громоздких выкладок, которые мы здесь не приводим), а следовательно, по теореме 9.2.4 отображение g также принадлежит классу C^r . $\square$

9.5.3. Локализация. Сейчас мы кратко поясним, как утверждение теоремы 9.5.2 связано с приведенным выше описанием в терминах линеаризации. Предположим, что отображение $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ имеет неподвижную точку p , для которой $Df(p)$ является гиперболическим линейным отображением. Заменяя (x, y) на $(x, y) - p$, мы преобразуем координаты так, чтобы неподвижная точка оказа-

лась в начале координат. Далее, хотя $Df(p)$ и является гиперболическим линейным отображением, оно может и не быть диагональным. Если мы произведем замену координат, диагонализировав $Df(p)$, мы получим $Df(0) = A = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$. Если отображение f непрерывно дифференцируемо, тогда по определению $\|Df(q) - A\| < \epsilon$ для точек q , расположенных вблизи начала координат. Теперь можно построить глобально заданное отображение, которое совпадает с данным отображением вблизи неподвижной точки и всюду близко к производной. Тогда, применяя этот результат, мы получим сжимающуюся кривую C , по крайней мере вблизи начала координат, и аналогично можно получить растягивающуюся кривую, рассматривая обратное к f отображение. Нетрудно убедиться, что $\bigcup_{n \in \mathbb{N}} f^{-n}(C)$ является простой кривой, т. е. не имеет самопересечений.

9.5.4. Теорема о гиперболической неподвижной точке. Заметим, что в теореме 9.5.2 начало координат является трансверсальной неподвижной точкой отображения A и, следовательно, сохраняется в силу предложения 9.3.1. Однако эта устойчивость также следует из приведенных здесь рассуждений. Поскольку отображение f близко к A в метрике пространства C^1 , гиперболичность также сохраняется. Итак, мы получаем теорему о гиперболической неподвижной точке.

Теорема 9.5.3. *Если p — гиперболическая точка периода t некоторого C^1 -отображения f , то для всякого отображения g , достаточно близкого к f в метрике пространства C^1 , имеется единственная точка периода t , близкая к p , и она также гиперболическая.*

Следующая «глобальная» версия этого утверждения также полезна.

Теорема 9.5.4. *Предположим, что $0 < \chi_s < 1 < \chi_u$ и $A: \mathbb{R}^n \rightarrow \mathbb{R}^n$ — такое линейное отображение, что $A|_{\mathbb{R}^m \times \{0\}}$ является χ_s -сжатием и $A^{-1}|_{\{0\} \times \mathbb{R}^{n-m}}$ является $1/\chi_u$ -сжатием. Если $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ — такое отображение, что отображение $f := F - A$ ограничено и $\text{Lip}(f) \leq \epsilon := \min\{1 - \chi_s, 1 - \chi_u^{-1}\}$ (см. определение 2.2.1), тогда F имеет единственную такую неподвижную точку p , что $\|p\| < \|F(0)\|/(\epsilon - \text{Lip}(f))$.*

Замечание 9.5.5. Ясно, что p — гиперболическая неподвижная точка отображения F .

Используя лишь методы, изученные к настоящему моменту (устойчивые сечения или принцип сжимающих отображений), можно усилить теорему 9.5.3, доказав локальную сопряженность. Этот результат является простым частным случаем *структурной устойчивости* (см. также п. 10.2.6), замечательный пример которой приведен в теореме 7.4.3.

Теорема 9.5.6 (Гробман—Хартман)¹. *Пусть $U \subset \mathbb{R}^n$ — открытое множество, $f: U \rightarrow \mathbb{R}^n$ — непрерывно дифференцируемое отображение и $O \in U$ —*

¹Прямое доказательство этой теоремы можно найти в книге: А. Каток, Б. Хасселлат. Введение в современную теорию динамических систем; здесь мы получим его с помощью теоремы 9.5.8.

гиперболическая неподвижная точка отображения f . Тогда существуют такие окрестности U_1, U_2, V_1, V_2 точки O и такой гомеоморфизм $h: U_1 \cup U_2 \rightarrow V_1 \cup V_2$, что $f = h^{-1} \circ Df_0 \circ h$ на U_1 .

На самом деле топологический характер отображения f вблизи O определен ориентацией этого отображения на устойчивых и неустойчивых многообразиях и их размерностями, так же как в теореме 7.4.3, в которой отсутствует предположение о близости.

Теорема 9.5.7. *Предположим, что отображения $f: U \rightarrow \mathbb{R}^n$ и $g: V \rightarrow \mathbb{R}^n$ имеют гиперболические неподвижные точки $p \in U$ и $q \in V$ соответственно, причем $\dim E^+(Df_p) = \dim E^+(Dg_q)$, $\dim E^-(Df_p) = \dim E^-(Dg_q)$, $\text{sign det } Df_p|_{E^+(Df_p)} = \text{sign det } Dg_q|_{E^+(Dg_q)}$ и $\text{sign det } Df_p|_{E^-(Df_p)} = \text{sign det } Dg_q|_{E^-(Dg_q)}$. Тогда существуют такие окрестности $U_1 \subset U$ и $V_1 \subset V$ и такой гомеоморфизм $h: U_1 \rightarrow V_1$, что $h \circ f = g \circ h$.*

Следующий результат эквивалентен теореме 9.5.6, в чем можно убедиться с помощью локализации. Мы получаем его как приложение теоремы о гиперболической неподвижной точке (теорема 9.5.4).

Теорема 9.5.8. *Пусть $0 < \chi_s < 1 < \chi_u$ и $A: \mathbb{R}^n \rightarrow \mathbb{R}^n$ — такое линейное отображение, что $A|_{\mathbb{R}^m \times \{0\}}$ является χ_s -сжатием и $A^{-1}|_{\{0\} \times \mathbb{R}^{n-m}}$ является $1/\chi_u$ -сжатием. Предположим, что F — такое отображение, что отображение $f := F - A$ ограничено и $\text{Lip}(f) \leq \varepsilon := \min\{\|A^{-1}\|^{-1}, 1 - \chi_s, 1 - \chi_u^{-1}\}$. Тогда имеется единственный такой гомеоморфизм H пространства $\mathbb{R}^n$, что отображение $h := H - \text{Id}$ ограничено и $H \circ A \circ H^{-1} = F$.*

Доказательство. Условие $\text{Lip}(f) \leq \|A^{-1}\|^{-1}$ гарантирует, что F — гомеоморфизм, потому что $y = F(x) \Leftrightarrow x = A^{-1}(y - f(x))$ и правая часть непрерывна и является сжимающим отображением в точке x , а следовательно, определяет точку x однозначно.

Чтобы добиться большей симметрии при доказательстве, введем отображение $G = A + g$, где функция g ограничена и $\text{Lip}(g) \leq \varepsilon$. Первый шаг заключается в том, чтобы показать, что существует единственное такое непрерывное ограниченное отображение h пространства $\mathbb{R}^n$, что

$$F \circ (\text{Id} + h) = (\text{Id} + h) \circ G, \quad (9.5.2)$$

или, что эквивалентно,

$$A \circ h \circ G^{-1} + f \circ (\text{Id} + h) \circ G^{-1} + A \circ G^{-1} - \text{Id} = h.$$

Для этого рассмотрим пространство E ограниченных непрерывных отображений из $\mathbb{R}^n$ в $\mathbb{R}^n$, которое распадается в пространство E_s ограниченных непрерывных отображений в $\mathbb{R}^m \times \{0\}$ и пространство E_u ограниченных непрерывных отображений в $\{0\} \times \mathbb{R}^{n-m}$. Определим отображения $A: E \rightarrow E$ и $\mathcal{F}: E \rightarrow E$, полагая

$$A(h) := A \circ h \circ G^{-1} \quad \text{и} \quad \mathcal{F}(h) := f \circ (\text{Id} + h) \circ G^{-1} + A \circ G^{-1} - \text{Id}.$$

Тогда $\mathcal{A}$ сохраняет E_s и E_u и при этом $\mathcal{A}|_{E_s}$ является χ_s -сжатием, а $\mathcal{A}^{-1}|_{E_u}$ является χ_u^{-1} -сжатием. Поскольку $\text{Lip}(\mathcal{F}) \leq \text{Lip}(f)$, из теоремы о гиперболической неподвижной точке (теорема 9.5.4) следует, что отображение $\mathcal{A} + \mathcal{F}$ имеет единственную неподвижную точку в E . Отсюда мы получаем искомое отображение h .

Остается показать, что H — гомеоморфизм. Именно здесь нам поможет введенная симметрия. Она позволяет получить такое отображение $\bar{h}$, что

$$G \circ (\text{Id} + \bar{h}) = (\text{Id} + \bar{h}) \circ F.$$

Комбинируя это равенство с соотношением (9.5.2) двумя способами, мы получаем два соотношения

$$G \circ (\text{Id} + \bar{h}) \circ (\text{Id} + h) = (\text{Id} + \bar{h}) \circ (\text{Id} + h) \circ G,$$

$$F \circ (\text{Id} + h) \circ (\text{Id} + \bar{h}) = (\text{Id} + h) \circ (\text{Id} + \bar{h}) \circ F$$

Это соотношения такого же вида, что и равенство (9.5.2), так как его можно получить, если в первой части доказательства положить $F = G$. Поскольку мы уже установили однозначность отображения сопряжения и поскольку Id сопрягает F с F и G с G , мы заключаем, что

$$(\text{Id} + \bar{h}) \circ (\text{Id} + h) = (\text{Id} + h) \circ (\text{Id} + \bar{h}) = \text{Id},$$

а из этого, в частности, следует, что $H = \text{Id} + h$ — гомеоморфизм. □

ГИПЕРБОЛИЧЕСКАЯ ДИНАМИКА

В этой главе описывается общая структура, порождающая те типы поведения, которые мы наблюдали ранее в различных примерах сложной динамики. Сущность хаотического поведения в динамике заключается в комбинации растяжения (для разделения орбит) и изгиба (для обеспечения рекуррентности и совместимости с компактностью). Мы приведем точное определение соответствующих понятий и получим несколько важных общих следствий. Факты и идеи, рассмотренные в этой главе, играют важную роль во многих областях динамики и будут широко использоваться в последующих главах. Опушенные доказательства можно найти в нашей книге «Введение в современную теорию динамических систем» (с теми же самыми обозначениями).

§ 10.1. Гиперболические множества

Чтобы извлечь интересные общие свойства систем из наших примеров, удобно рассматривать обратимые и необратимые отображения по отдельности.

10.1.1. Определение. Гиперболическое множество — это такое инвариантное множество, что в любой его точке x производная имеет такой же вид, как в гиперболической неподвижной точке, т. е. имеются такие дополнительные подпространства E_x^u и E_x^s (растягивающееся, или «неустойчивое», и сжимающееся, или «устойчивое»), что отображение $Df^{-1}(x)$ является x -сжатием (определение 2.2.1) на E_x^u (с образом $E_{f^{-1}(x)}^u$) и отображение $Df|_x$ является x -сжатием на E_x^s для каждого x , а константа $\chi < 1$ не зависит от x . На самом деле достаточно, чтобы отображение было не сжимающим, а почти сжимающим, как в определении 2.6.11, и это упрощает проверку выполнения всех условий определения.

Определение 10.1.1. Пусть U — открытое множество и f — отображение, определенное на U . Пусть Λ — компактное инвариантное множество, т. е. $f(\Lambda) = \Lambda$, и отображение f является обратимым на Λ . Тогда Λ называется *гиперболическим множеством*, если в каждой точке $x \in \Lambda$ существуют такие подпространства E_x^u и E_x^s , что каждый вектор v можно единственным образом записать как $v = v^u + v^s$, где $v^u \in E_x^u$ и $v^s \in E_x^s$, и имеются такие числа $C > 0$ и $\chi < 1$, что $\|Df^{-n}(x)|_{E_x^u}\| \leq C\chi^n$ и $\|Df^n(x)|_{E_x^s}\| \leq C\chi^n$ для всех $x \in \Lambda$ и $n \in \mathbb{N}$.

Гиперболическое линейное отображение тора, определенное в п. 7.1.4, представляет собой пример такой ситуации. Множество Λ — это весь тор, а растягивающееся и сжимающееся подпространства параллельны соответствующим собственным направлениям. Константа χ является наименьшим из двух собственных значений.

Обратимыми примерами являются также троичная подкова из п. 7.3.3 и общая линейная подкова из п. 7.4.4. Инвариантное множество Λ , построенное для этих примеров, является гиперболическим. Растягивающиеся подпространства — горизонтальные прямые, сжимающиеся — вертикальные. Константа χ определяется коэффициентами растяжения и сжатия, которые выбираются для отображения подковы. Например, она равна $1/3$ для троичной подковы.

Мы использовали обратимость для определения направления растяжения. Для необратимых отображений это приводит к трудностям, за исключением случая, когда имеется растяжение во всех направлениях. Поскольку все наши необратимые примеры одномерны, это условие выполнено автоматически.

Определение 10.1.2. Пусть U — открытое множество и f — отображение, определенное на U . Компактное инвариантное множество Λ называется *гиперболическим отталкивающим множеством*, или *репеллером*, если имеется такое число $\chi > 1$, что для любой точки $x \in \Lambda$ и для любого вектора v выполняется неравенство $\|Df^n(x)v\| \geq C\chi^n\|v\|$.

Линейные растягивающие отображения E_m окружности, рассмотренные в п. 7.1.1, обладают этим свойством для $\Lambda = S^1$ (и $\chi = |m|$). Этим же свойством обладают нелинейные растягивающие отображения (см. определение 7.1.7). В силу предположения о компактности из того, что $|f'(x)| > 1$ для всех $x \in S^1$, следует существование такого числа $\chi > 1$, что $|f'(x)| \geq \chi$ для всех $x \in S^1$. Во всех этих случаях можно взять $C = 1$.

При $\lambda > 4$ инвариантное множество $\bigcap_{n \in \mathbb{N}_0} f_\lambda^{-n}([0, 1])$ квадратичных отображений из § 2.5 и п. 7.1.2, определенное в предложении 7.4.4, является гиперболическим репеллером (см. предложение 11.4.1). Если $4 < \lambda \leq 2\sqrt{5}$, то проверка этого утверждения сложнее, чем при доказательстве предложения 7.4.4. Кроме того, здесь мы не можем выбрать $C = 1$, как в определении 10.1.2, если не менять метрику.

Определение 10.1.3. Пусть Λ — гиперболическое множество для отображения f на U . Если имеется такая открытая окрестность V множества Λ , что $\Lambda = \bigcap_{n \in \mathbb{Z}} f^n(\bar{V})$, то множество Λ называется *локально максимальным, изолированным* или *базисным*.

Это предположение вполне естественно, и мы, как правило, будем использовать его в дальнейшем, иногда неявно.

10.1.2. Условие конусов. Приведем еще одно описание гиперболических множеств, более наглядное и зачастую более удобное для проверки. Типичный пример — доказательство существования аттрактора Лоренца (см. теорему 13.3.3).

Определение 10.1.4. Стандартный горизонтальный γ -конус в точке $p \in \mathbb{R}^n$ — это множество вида

$$H_p^\gamma = \{(u, v) \in T_p \mathbb{R}^n : \|v\| \leq \gamma \|u\|\}.$$

Стандартный вертикальный γ -конус в точке p — это множество вида

$$V_p^\gamma = \{(u, v) \in T_p \mathbb{R}^n : \|u\| \leq \gamma \|v\|\}.$$

Конус K в $\mathbb{R}^n$ определяется как образ стандартного конуса под действием обратимого линейного отображения.

Приведем несколько примеров. В размерности $n = 2$ все конусы выглядят одинаково. Горизонтальный конус $|x_2| \leq \gamma |x_1|$ закрашен на рис. 10.1.1 серым цветом. Замыкание его дополнения $|x_1| \leq |x_2|/\gamma$ является вертикальным конусом, трехмерный аналог которого изображен на рис. 10.1.2. В размерности $n = 3$

Рис. 10.1.1. Горизонтальный конус

Рис. 10.1.2. Вертикальный конус

множество, описанное соотношениями $u = x_1$, $v = (x_2, x_3)$, $\sqrt{x_2^2 + x_3^2} \leq \gamma|x_1|$, очевидно, является конусом. Но замыкание дополнения конуса и в этом случае является конусом, так что, полагая $u = (x_2, x_3)$, $v = x_1$, $|x_1| \leq \sqrt{x_2^2 + x_3^2}/\gamma$, мы получим пример конуса, который совсем не похож на рожок для мороженого.

Под *полем конусов* мы подразумеваем отображение, которое каждой точке $p \in \mathbb{R}^n$ ставит в соответствие конус K_p в точке p . Дiffeоморфизм $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ естественно действует на полях конусов согласно правилу

$$(f_*K)_p := Df_{f^{-1}(p)}(K_{f^{-1}(p)}).$$

Теорема 10.1.5. *Компактное f -инвариантное множество Λ гиперболично тогда и только тогда, когда существуют такие числа λ, μ , где $\lambda < 1 < \mu$, что для каждого $x \in \Lambda$ имеются дополнительные подпространства S_x и T_x (вообще говоря, не Df -инвариантные), поле горизонтальных конусов $H_x \supset S_x$ и поле вертикальных конусов $V_x \supset T_x$, ассоциированных с этим разложением, удовлетворяющих условиям*

$$Df_x H_x \subset \text{Int } H_{f(x)}, \quad Df_x^{-1} V_{f(x)} \subset \text{Int } V_x,$$

$$\|Df_x \xi\| \geq \mu \|\xi\| \text{ для } \xi \in H_x \text{ и } \|Df_x^{-1} \xi\| \geq \lambda^{-1} \|\xi\| \text{ для } \xi \in V_{f(x)}.$$

10.1.3. Устойчивые многообразия. Теорема 9.5.2 утверждает, что подпространствам E^u и E^s , растягивающимся и сжимающимся под действием дифференциала, соответствуют кривые, растягивающиеся и сжимающиеся под действием отображения (это доказано с помощью принципа сжимающих отображений). Хотя в теореме 9.5.2 эти кривые были получены только для гиперболической неподвижной точки, используя этот результат во вспомогательном пространстве, мы можем показать, что в случае гиперболического множества такие кривые существуют в каждой точке, вне зависимости от того, неподвижная она или нет. Иными словами, можно изменить доказательство так, что оно будет применимо вдоль орбит всех точек, которые, возможно, и не являются неподвижными.

Мы уже использовали эти инвариантные кривые прежде. Доказывая, что линейные гиперболические отображения на торе являются перемешивающими (см. предложение 7.2.9), мы использовали прямые, параллельные собственным направлениям, которые являются в точности устойчивыми и неустойчивыми кривыми. Общая гиперболическая система изображается точно так же с помощью этих линий, возможно немного «искривленных». Тот же метод, который применялся при доказательстве теоремы 9.5.2, позволяет получить следующий результат.

Теорема 10.1.6 (об устойчивом и неустойчивом многообразиях для гиперболических множеств). *Пусть Λ — такое гиперболическое множество для диффеоморфизма f класса C^1 на множестве U , что Df на Λ растягивает E^u с коэффициентом $\mu > 1$ и сжимает E^s с коэффициентом $\lambda < 1$. Тогда для всякого $x \in \Lambda$ найдется пара таких C^1 -вложенных (см. определение A.2.5) дисков $W^s(x)$, $W^u(x)$, называемых локальным устойчивым многооб-*

разием и локальным неустойчивым многообразием точки x соответственно, что

1) пространство $E^s(x)$ является касательным к $W^s(x)$ в точке x , а $E^u(x)$ — касательным к $W^u(x)$ в точке x ;

2) $f(W^s(x)) \subset W^s(f(x))$, $f^{-1}(W^u(x)) \subset W^u(f^{-1}(x))$;

3) для каждого $\delta > 0$ существует такое число $C(\delta)$, что для $n \in \mathbb{N}$ выполняются неравенства¹

$$\text{dist}(f^n(x), f^n(y)) < C(\delta)(\lambda + \delta)^n \text{dist}(x, y) \text{ для } y \in W^s(x),$$

$$\text{dist}(f^{-n}(x), f^{-n}(y)) < C(\delta)(\mu - \delta)^{-n} \text{dist}(x, y) \text{ для } y \in W^u(x);$$

4) существуют такое число $\beta > 0$ и такое семейство окрестностей O_x , содержащих шар с центром в точке $x \in \Lambda$ радиуса β , что

$$W^s(x) = \{y: f^n(y) \in O_{f^n(x)}, \quad n \in \mathbb{N}\},$$

$$W^u(x) = \{y: f^{-n}(y) \in O_{f^{-n}(x)}, n \in \mathbb{N}\}.$$

Из этого также следует, что глобальные устойчивое и неустойчивое многообразия

$$\tilde{W}^s(x) = \bigcup_{n=0}^{\infty} f^{-n}(W^s(f^n(x))),$$

$$\tilde{W}^u(x) = \bigcup_{n=0}^{\infty} f^n(W^u(f^{-n}(x)))$$

определены независимо от конкретного выбора локальных устойчивого и неустойчивого многообразий и могут быть описаны топологически:

$$\tilde{W}^s(x) = \{y \in U: \text{dist}(f^n(x), f^n(y)) \rightarrow 0, n \rightarrow \infty\},$$

$$\tilde{W}^u(x) = \{y \in U: \text{dist}(f^{-n}(x), f^{-n}(y)) \rightarrow 0, n \rightarrow \infty\}.$$

Легко доказать следующий полезный результат.

Предложение 10.1.7. Обозначим через $W_\epsilon^s(x)$ и $W_\epsilon^u(x)$ шары радиуса ϵ в $\tilde{W}^s(x)$ и $\tilde{W}^u(x)$. Тогда существуют такое $\epsilon > 0$, что для любых $x, y \in \Lambda$ пересечение $W_\epsilon^s(x) \cap W_\epsilon^u(y)$ состоит не более чем из одной точки $[x, y]$, и такое $\delta > 0$, что если $d(x, y) < \delta$ для некоторых $x, y \in \Lambda$, то $W_\epsilon^s(x) \cap W_\epsilon^u(y) \neq \emptyset$.

Из тех же рассуждений, с помощью которых мы доказали, что автоморфизмы тора являются топологически перемешивающими, также следует, что если глобальные устойчивые и неустойчивые многообразия плотны в гиперболическом множестве, то гиперболическое множество является топологически перемешивающим (а следовательно, и хаотическим; см. предложение 7.2.14). Обратное также справедливо.

Предложение 10.1.8. Если Λ — компактное локально максимальное гиперболическое множество для f и $f: \Lambda \rightarrow \Lambda$ — топологически перемешива-

¹Здесь и далее $\text{dist}(a, b)$ обозначает расстояние между точками a и b . — Прим. ред.

ющее отображение, то найдется такое $N \in \mathbb{N}$, что для $x, y \in \Lambda$ и $n \geq N$ выполнено условие $f^n(W^u(x)) \cap W^s(y) \neq \emptyset$.

Этот простой результат имеет замечательные динамические следствия, среди которых теорема 10.2.8. Неустойчивость по начальным условиям — одна из составляющих хаоса, и она является немедленным следствием теоремы 10.1.6. Однако из нее вытекает и более сильное свойство. Неустойчивость означает, что близкие орбиты могут расходиться, но описание локальной гиперболической структуры, приведенное в теореме 10.1.6, показывает, что близкие орбиты *должны* расходиться или в будущем, или в прошлом — локальная картина вблизи гиперболической точки показывает, что все другие точки удаляются от нее либо в будущем из-за ненулевой неустойчивой составляющей, либо в прошлом благодаря устойчивой составляющей, а обычно случается и то, и другое. Это свойство весьма важно.

Определение 10.1.9. Гомеоморфизм (соответственно непрерывное отображение) $f: X \rightarrow X$ называется *разделяющим*, если существует такая константа $\delta > 0$, что из выполнения неравенства $d(f^n(x), f^n(y)) < \delta$ для всех $n \in \mathbb{Z}$ (соответственно $n \in \mathbb{Z}^+$) следует, что $x = y$.

Максимально возможное значение δ иногда называется постоянной разделения для динамической системы. В силу компактности свойство отображения быть разделяющим не зависит от конкретного выбора метрики на X , определяющей данную топологию, и, следовательно, является инвариантом топологического сопряжения. Однако постоянная разделения зависит от выбора метрики.

Следствие 10.1.10. *Ограничение диффеоморфизма на гиперболическое множество является разделяющим.*

Доказательство. Если $\text{dist}(f^n(x), f^n(y)) < \beta$ для $n \in \mathbb{Z}$, тогда $f^n(y) \in O_{f^n(x)}$ для $n \in \mathbb{Z}$ и $y \in W^s(x) \cap W^u(x) = \{x\}$ в силу теоремы 10.1.6, 4). $\square$

В наших примерах гиперболических динамических систем топологическая энтропия также является показателем экспоненциального роста периодических орбит. Это утверждение справедливо для всех гиперболических систем, и одна из причин этого состоит в том, что отображение разделяющее: благодаря свойству разделения, периодические орбиты являются отделенными, так что их можно взять в качестве отделенных множеств при определении энтропии. Мы вернемся к этому в п. 10.2.4.

10.1.4. Принцип сжимающих отображений. Принцип сжимающих отображений (предложение 2.6.10) является одним из краеугольных камней для общей гиперболической теории. Он лежит в основе построения устойчивых и неустойчивых многообразий, вне зависимости от того, используется ли доказательство, приведенное в § 9.5, или какое-либо другое. Уже одно это свидетельствует о важности принципа сжимающих отображений. Однако существуют и другие прямые и косвенные приложения этого принципа, которые используются в гиперболической теории. Например, в гл. 9 он применяется таким образом: некоторый интересующий нас объект является неподвижной точкой отображения, и это отображение сжимающее.

Мы показываем, что отображение является сжимающим, с использованием предположений о сжатии/растяжении из определения 10.1.1. При этом в большинстве случаев используются похожие рассуждения, поэтому доказательство можно сократить, используя теорему 9.5.4 о неподвижной гиперболической точке, которая предоставляет универсальный способ перехода от гиперболичности к сжатию. Таким образом, многочисленные доказательства приводят к построению отображения во вспомогательном пространстве, которое является скорее гиперболическим, а не сжимающим, потому что такое отображение можно построить более непосредственно. Примером этого является доказательство теоремы 10.2.2.

§ 10.2. Структура и рост орбит

10.2.1. Плотность множества периодических точек, лемма о замыкании.

Плотность периодических орбит (см. предложения 7.1.2, 7.1.3, 7.1.10 и следствие 7.4.7), — это не просто общая особенность, характерная для наших примеров, а внутреннее свойство гиперболической динамики.

Плотность множества периодических точек следует из леммы Аносова о замыкании. Эта лемма, тесно связанные с ней теоремы об ε -траекториях и спецификации, а также инвариантные многообразия, кодирование и принцип сжимающих отображений позволяют получить детальную картину структуры орбит и фундаментальные результаты о статистическом поведении и структурной устойчивости.

Мы приводим лемму о замыкании вместе с доказательством, чтобы проиллюстрировать использование теоремы о гиперболической неподвижной точке (теорема 9.5.4) (а следовательно, косвенно и принципа сжимающих отображений) в конкретной ситуации.

Определение 10.2.1. Пусть (X, d) — метрическое пространство, $U \subset M$ — открытое множество и задано отображение $f: U \rightarrow X$. Для $a \in \mathbb{Z} \cup \{-\infty\}$ и $b \in \mathbb{Z} \cup \{\infty\}$ последовательность $\{x_n\}_{a < n < b} \subset U$ называется ε -орбитой или ε -псевдоорбитой отображения f , если $d(x_{n+1}, f(x_n)) < \varepsilon$ при $a < n$ и $n + 1 < b$. Если, кроме того, $-\infty < a < b < \infty$ и $x_{b-1} = x_{a+1}$, тогда эта последовательность называется *периодической ε -орбитой*.

Теорема 10.2.2 (лемма Аносова о замыкании). Пусть Λ — гиперболическое множество для отображения f на U . Тогда имеются такая окрестность $V \supset \Lambda$ и такие константы $C, \varepsilon_0 > 0$, что для $\varepsilon < \varepsilon_0$ и любой периодической ε -орбиты $(x_0, \dots, x_m) \subset V$ найдется точка $y \in U$, удовлетворяющая условиям $f^m(y) = y$ и $\text{dist}(f^k(y), x_k) < C\varepsilon$, $k = 0, \dots, m - 1$.

Замечание 10.2.3. Частным случаем ε -периодической орбиты является конечная орбита $x_0, f(x_0), \dots, f^{m-1}(x_0)$, для которой $\text{dist}(f^m(x_0), x_0) < \varepsilon$. Таким образом, из леммы Аносова о замыкании, в частности, следует, что вблизи любой точки в гиперболическом множестве, орбита которой почти возвращается к ней, имеется периодическая орбита, которая следует близко за «почти возвращающейся» частью. Эту лемму можно переформулировать по-другому, как результат о плотности множества периодических точек.

Определение 10.2.4. Множество $\{x: \forall \epsilon > 0 \text{ имеется периодическая } \epsilon\text{-орбита, содержащая точку } x\}$ называется *цепочно-рекуррентным множеством*.

Согласно лемме Аносова о замыкании множество периодических точек в цепочно-рекуррентном множестве плотно.

Если V — открытая окрестность множества Λ , тогда любая периодическая точка в V содержится в Λ'_V (см. определение 10.1.3). Если окрестность V достаточно мала и множество Λ локально максимально, тогда эти орбиты лежат в Λ . Таким образом, в этом случае из леммы Аносова о замыкании мы получаем периодическую точку $y \in \Lambda$. В частности, из топологической транзитивности следует плотность множества периодических точек, а значит, и хаотичность гиперболического множества.

Доказательство теоремы 10.2.2. Для каждого x_k имеется окрестность V_k , на которой f является малым возмущением гиперболического линейного отображения и имеет вид $f_k(u, v) = (A_k u + \alpha_k(u, v), B_k v + \beta_k(u, v))$, где $\|\alpha_k\|$, $\|\beta_k\|$, $\|D\alpha_k\|$ и $\|D\beta_k\|$ ограничены константой $C_1 \epsilon$ для всех k и некоторого $C_1 > 0$. Мы не предполагаем, что начало координат неподвижно для всех f_k .

Последовательность $(u_k, v_k) \in V_k$, $k = 0, \dots, m-1$, является периодической орбитой тогда и только тогда, когда

$$(u, v) := ((u_0, v_0), (u_1, v_1), \dots, (u_{m-1}, v_{m-1})) = \\ = (f_{m-1}(u_{m-1}, v_{m-1}), f_0(u_0, v_0), \dots, f_{m-2}(u_{m-2}, v_{m-2})) =: F(u, v).$$

Поэтому мы должны найти неподвижную точку отображения F . Мы используем норму $\|(x_0, x_1, \dots, x_{m-1})\| := \max_{0 \leq i \leq m-1} \|x_i\|$. Представим отображение F как «линейное плюс малое»:

$$F(u, v) = L(u, v) + S(u, v),$$

где

$$S((u_0, v_0), (u_1, v_1), \dots, (u_{m-1}, v_{m-1})) := \\ := ((\alpha_{m-1}(u_{m-1}, v_{m-1}), \beta_{m-1}(u_{m-1}, v_{m-1})), \dots \\ \dots, (\alpha_{m-2}(u_{m-2}, v_{m-2}), \beta_{m-2}(u_{m-2}, v_{m-2}))), \\ L((u_0, v_0), (u_1, v_1), \dots, (u_{m-1}, v_{m-1})) := \\ := ((A_{m-1}u_{m-1}, B_{m-1}v_{m-1}), (A_0u_0, B_0v_0), \dots, (A_{m-2}u_{m-2}, B_{m-2}v_{m-2})).$$

Отображение L является гиперболическим: оно растягивает подпространство $((u_0, 0), (u_1, 0), \dots, (u_{m-1}, 0))$ и сжимает подпространство $((0, v_0), (0, v_1), \dots, (0, v_{m-1}))$. Поскольку $\|S(u, v) - S(u', v')\| \leq C_3 \cdot \epsilon \cdot \|(u, v) - (u', v')\|$ для некоторой константы $C_3 = C_3(f, \Lambda) > 0$, мы можем применить теорему 9.5.4 о неподвижной гиперболической точке и получим искомую замкнутую орбиту вблизи псевдоорбиты. $\square$

С помощью леммы Аносова о замыкании можно получить множество периодических точек, о чем будет говориться ниже. Но сначала мы представим

связанные с этой леммой результаты, которые в некотором смысле являются более сильными.

10.2.2. Лемма об ϵ -траекториях. Самый непосредственный способ усиления леммы Аносова о замыкании состоит в том, чтобы опустить условие периодичности и приближать любую псевдоорбиту настоящей орбитой. Это понятие очень важно, поэтому введем для него специальное название.

Определение 10.2.5. Пусть (X, d) — метрическое пространство, $U \subset M$ — открытое множество и задано отображение $f: U \rightarrow X$. Для $a \in \mathbb{Z} \cup \{-\infty\}$ и $b \in \mathbb{Z} \cup \{\infty\}$ говорят, что орбита $\mathcal{O}(x)$ точки $x \in U$ δ -приближает псевдоорбиту $\{x_n\}_{a < n < b} \subset U$, если $d(x_n, f^n(x)) < \delta$ для всех $n \in (a, b)$.

Теорема 10.2.6 (лемма об ϵ -траекториях). Пусть f — диффеоморфизм с компактным гиперболическим множеством Λ . Тогда существует окрестность $U(\Lambda)$ множества Λ , удовлетворяющая следующему условию: для любого $\delta > 0$ имеется такое $\epsilon > 0$, что всякая ϵ -орбита в $U(\Lambda)$ δ -приближается некоторой истинной орбитой отображения f .

Лемму об ϵ -траекториях можно доказать почти точно так же, как и лемму Аносова о замыкании (теорема 10.2.2), рассматривая последовательности отображений пространства $\mathbb{R}^n$, близких к гиперболическим линейным отображениям. Хотя эта лемма не содержит непосредственного утверждения о том, что приближающая орбита для периодической псевдоорбиты периодическая, можно получить это дополнительное утверждение, приближая бесконечную повторяющую себя псевдоорбиту, а затем используя свойство разделения. Таким образом, это в самом деле усиление леммы о замыкании.

Лемма об ϵ -траекториях подтверждает возможность проведения численных экспериментов при вычислении гиперболических множеств. Наличие неустойчивости (и, более того, разделения) наводит на мысль о том, что вычисление отдельной орбиты невозможно, потому что ошибки округления будут увеличены динамикой экспоненциально. Однако лемма об ϵ -траекториях гарантирует, что вычисленная псевдоорбита находится близко к *некоторой* орбите в гиперболическом множестве.

Поскольку цель таких вычислений состоит в том, чтобы получить полезное графическое представление рассматриваемого гиперболического множества, обычно странного аттрактора, имеется существенная проблема. Хотя из леммы об ϵ -траекториях следует, что каждая изображенная орбита представляет подлинную орбиту, нет гарантии, что эти подлинные орбиты типичны. В этом можно убедиться на примере линейного растягивающего отображения E_2 , для которого при повторном удвоении двоичных представлений с помощью компьютера все компьютерные орбиты притягиваются к началу координат и, таким образом, мы пропускаем все типичные орбиты. Итак, даже без учета ошибок округления мы пока не знаем, как убедиться, что получены именно типичные орбиты. Ответ на этот вопрос будет найден с помощью приведенной ниже теории SRB-мер, из которой следует, что почти при любом выборе начальных данных мы получаем типичную орбиту.

10.2.3. Спецификация. Лемму об ϵ -траекториях можно рассматривать как инструмент для создания орбит. При правильном выборе начальных данных неустойчивость по начальным условиям может использоваться для усиления ϵ -отклонения в псевдоорбите к макроскопическим эффектам желательного типа, и орбита, реализующая эти эффекты, находится с помощью леммы об ϵ -траекториях. С этой точки зрения имеется красивое видоизменение леммы, которое позволяет получать орбиты, поведение которых описывается с большой точностью. По существу, оно состоит в том, что для того чтобы попасть в указанное время на реальную орбиту, которая может быть выбрана периодической, можно взять любой конечный набор (конечных) отрезков орбит и сшить их вместе. Объединяя этот замечательный результат со свойством разделения, мы получим удобное средство для определения топологической и статистической структуры орбиты гиперболического множества.

Определение 10.2.7. Пусть $f: X \rightarrow X$ — биекция множества X . *Спецификация* $S = (\tau, P)$ состоит из такого конечного набора $\tau = \{I_1, \dots, I_m\}$ конечных отрезков $I_i = [a_i, b_i] \subset \mathbb{Z}$ и такого отображения $P: T(\tau) := \bigcup_{i=1}^m I_i \rightarrow X$, что для $t_1, t_2 \in I \in \tau$ выполняется неравенство $f^{t_2-t_1}(P(t_1)) = P(t_2)$. Спецификация S называется *n -разделенной*, если $a_{i+1} > b_i + n$ для всех $i \in \{1, \dots, m\}$, и минимальное такое число n называется *разделением* спецификации S . Мы говорим, что спецификация S *параметризует* набор $\{P_I: I \in \tau\}$ отрезков орбит отображения f .

Мы полагаем, что $T(S) := T(\tau)$ и $L(S) := L(\tau) := b_m - a_1$. Пусть (X, d) — метрическое пространство. Тогда мы говорим, что спецификация S ϵ -*приближается* точкой $x \in X$, если $d(f^n(x), P(n)) < \epsilon$ для всех $n \in T(S)$.

Таким образом, спецификация — это параметризованное объединение отрезков $P|_{I_i}$ орбит отображения f .

Пусть (X, d) — метрическое пространство и $f: X \rightarrow X$ — гомеоморфизм. Тогда говорят, что f обладает *свойством спецификации*, если для любого $\epsilon > 0$ существует такое $M = M_\epsilon \in \mathbb{N}$, что любая M -разделенная спецификация S ϵ -приближается некоторым элементом $x \in X$ и, более того, для любого $q \geq M + L(S)$ существует периодическая орбита периода q , ϵ -приближающая S .

Теорема 10.2.8 (о спецификации). Пусть Λ — топологически перемешивающее компактное локально максимальное гиперболическое множество для диффеоморфизма f . Тогда $f|_\Lambda$ обладает свойством спецификации.

Доказательство основано главным образом на использовании предложения 10.1.8. Начнем с первого отрезка орбиты и обозначим его конечную точку через x . Обозначая начальную точку второго отрезка через y , применим предложение 10.1.8 к достаточно маленьким частям инвариантных многообразий. Полученная в результате точка пересечения определяет второе приближение к искомому отрезку орбиты. Требуемое время перехода зависит только от того, насколько точное приближение нам нужно. Поскольку точка принадлежит множеству $W^u(x)$, она расположена чрезвычайно близко к начальному отрезку

орбиты, а находясь в множестве $W^s(y)$, она также хорошо приближает второй отрезок. Присоединяя таким же образом третий отрезок орбиты, мы оставляем точки предыдущего приближения почти неизменными; общее накопление ошибок ограничено некоторой геометрической прогрессией.

Замечание 10.2.9. Этот результат нельзя доказать без использования перемешивания. Действительно, как легко убедиться, из свойства спецификации следует, что отображение $f|_\Lambda$ — топологическое перемешивание. Однако это условие не столь ограничительно, как может показаться. Спектральное разложение (теорема 10.3.6) показывает, что в некотором смысле оно всегда выполнено.

10.2.4. Рост периодических орбит. В то время как плотность служит качественным показателем многочисленности периодических точек, в наших примерах мы получим также количественные оценки. Мы уже показали, что число периодических точек растет экспоненциально (см. предложения 7.1.2, 7.1.3, 7.1.10 и следствие 7.4.7) и с определенной скоростью. Обе эти черты являются общими для гиперболических динамических систем. Скорость роста можно получить с помощью кодирования, и она совпадает с топологической энтропией (см. гл. 8). Это означает, что рост периодических орбит является мерой полной динамической сложности в гиперболических динамических системах.

Предложение 10.2.10. Для разделяющего гомеоморфизма f компактного метрического пространства выполняется неравенство $\rho(f) \leq h_{\text{top}}(f)$.

Доказательство. Если δ_0 — постоянная разделения, то множество $\text{Fix}(f^n)$ является (n, δ_0) -отделенным для всех $n \in \mathbb{N}$, поскольку если $x \neq y \in \text{Fix}(f^n)$ и $\delta := \max\{d(f^i(x), f^i(y)) : 0 \leq i < n\}$, то $d(f^i(x), f^i(y)) \leq \delta$ для $i \in \mathbb{Z}$ и, следовательно, $\delta > \delta_0$. Таким образом, $P_n(f) \leq N(f, \epsilon, n)$ для $\epsilon < \delta_0$, откуда и следует нужное неравенство. $\square$

Используя свойство спецификации, мы можем симитировать отделенные множества множествами периодических орбит и получим обратное неравенство.

Теорема 10.2.11. Для разделяющего гомеоморфизма f компактного метрического пространства со свойством спецификации выполняется равенство $\rho(f|_\Lambda) = h_{\text{top}}(f|_\Lambda)$.

Доказательство. Любой элемент (n, ϵ) -отделенного множества E_n можно $\epsilon/2$ -приблизить периодической точкой периода $n + M_{\epsilon/2}$. Эти точки различны в соответствии с неравенством треугольника в метрике d_n^f . Таким образом, имеется по крайней мере $\text{card}(E_n)$ различных периодических точек периода $n + M_{\epsilon/2}$, и, следовательно, $P_{n+M_{\epsilon/2}}(f') \geq N(f', \epsilon, n)$, а значит, $\rho(f') \geq h_{\text{top}}(f')$. $\square$

Основными примерами разделяющих отображений со спецификацией являются локально максимальные гиперболические множества для диффеоморфизмов, но имеются и другие важные примеры, включая транзитивные топологические цепи Маркова и некоторые более общие классы символических систем типа *софических*.

В предыдущем доказательстве свойство спецификации не было использовано в полной мере. Более аккуратное использование этого свойства приводит к следующему более тонкому результату.

Теорема 10.2.12. Пусть X — компактное метрическое пространство и $f: X \rightarrow X$ — разделяющий гомеоморфизм со свойством спецификации. Тогда существуют такие константы $c_1, c_2 > 0$, что для $n \in \mathbb{N}$ выполняется оценка

$$c_1 e^{nh_{\text{top}}(f)} \leq P_n(f) \leq c_2 e^{nh_{\text{top}}(f)}.$$

Кодирование приводит к более точному результату, потому что порядок роста периодических орбит топологических цепей Маркова равен $\lambda^n +$ (экспоненциальные члены низшего порядка) в силу следствия 7.3.6. (Однако для систем с непрерывным временем кодирование не является столь мощным инструментом.)

10.2.5. Теорема о семействах ϵ -траекторий. В то время как теорема о спецификации является усилением леммы об ϵ -траекториях и используется как точный инструмент для построения индивидуальных орбит указанного вида, лемму об ϵ -траекториях можно усилить в другом направлении, чтобы получить жесткий глобальный контроль над структурой орбит.

Теорема 10.2.13 (о семействах ϵ -траекторий). Пусть M — риманово многообразие, d — естественная метрика, $U \subset M$ — открытое множество, $f: U \rightarrow M$ — диффеоморфизм, и $\Lambda \subset U$ — компактное гиперболическое множество для диффеоморфизма f .

Тогда существуют такая окрестность $U(\Lambda) \supset \Lambda$ и такие константы $\epsilon_0, \delta_0 > 0$, что для всякого $\delta > 0$ найдется число $\epsilon > 0$ со следующим свойством.

Если $f': U(\Lambda) \rightarrow M$ — C^2 -диффеоморфизм, ϵ_0 -близкий к f в топологии пространства C^1 , Y — топологическое пространство, $g: Y \rightarrow Y$ — гомеоморфизм, $\alpha \in C^0(Y, U(\Lambda))$ и $d_{C^0}(\alpha g, f'\alpha) := \sup_{y \in Y} d(\alpha g(y), f'\alpha(y)) < \epsilon$, то суще-

ствует такое отображение $\beta \in C^0(Y, U(\Lambda))$, что $\beta g = f'\beta$ и $d_{C^0}(\alpha, \beta) < \delta$.

Более того, отображение β локально единственно: если $\bar{\beta} g = f'\bar{\beta}$ и $d_{C^0}(\alpha, \bar{\beta}) < \delta_0$, то $\bar{\beta} = \beta$.

Замечание 10.2.14. 1) Локальная максимальность множества Λ не требуется.

2) Чтобы получить лемму об ϵ -траекториях, выберем $Y = \mathbb{Z}$ с дискретной топологией, $f' = f$, $\epsilon_0 = 0$, $g(n) = n + 1$ и заменим $\alpha \in C^0(Y, U(\Lambda))$ на $\{x_n\}_{n \in \mathbb{Z}} \subset U(\Lambda)$ и условие «такое $\beta \in C^0(Y, U(\Lambda))$, что $\beta g = f'\beta$ » на « $\{f^n(x)\}_{n \in \mathbb{Z}} \subset U(\Lambda)$ ». Тогда $d(x_n, f^n(x)) < \delta$ для всех $n \in \mathbb{Z}$.

3) Лемму о замыкании можно получить как другой частный случай этого результата, если взять $f' = f$, $Y = \mathbb{Z}/n\mathbb{Z}$, $g(k) = k + 1 \pmod{n}$.

10.2.6. Устойчивость и классификация. Гладкая динамическая система называется C^1 -структурно устойчивой, если всякое C^1 -возмущение топологически сопряжено с ней.

На первый взгляд кажется, что структурная устойчивость возможна только в том случае, когда есть конечное число (гиперболических) периодических точек. К каждой из них можно применить предложение 9.3.1 для некоторого определенного размера возмущений. Самый маленький из этих размеров сохраняет все эти орбиты. Однако при наличии бесконечного числа периодических точек провести такое рассуждение не удастся, потому что допустимые размеры возмущений могут быть произвольно малыми. Тем не менее, из теоремы 7.4.3 следует, что растягивающие отображения окружности структурно устойчивы. В самом деле, замечательно то, что даже если гиперболические динамические системы имеют бесконечно много периодических точек, тесная взаимозависимость разных орбит все же делает систему устойчивой в целом.

Структурная устойчивость характерна для всех гиперболических динамических систем, и эта особенность послужила одной из первых причин интереса к их изучению. Она позволяет автоматически получать много новых примеров гиперболических динамических систем, а именно возмущений наших примеров. В п. 7.4.4 предположение линейности при построении подковы несущественно. Нелинейные подковы демонстрируют точно такое же поведение. Они естественно возникают во многих приложениях; это описано в гл. 12.

Есть тесная связь между приближением псевдоорбит и структурной устойчивостью. Из структурной устойчивости (с непрерывной зависимостью сопрягающих гомеоморфизмов от возмущений), конечно же, следует, что орбиты возмущения приближаются невозмущенными орбитами. Обратно, орбиты возмущения динамической системы являются ϵ -орбитами для первоначальной системы. Поскольку они приближаются невозмущенными орбитами, соответствие, которое переводит возмущенные орбиты в невозмущенные, приближая их, позволяет получить сопряжение. Таким образом, теорема о семействах ϵ -траекторий может использоваться для доказательства структурной устойчивости гиперболических множеств.

Теорема 10.2.15 (сильная структурная устойчивость гиперболических множеств). Пусть $\Lambda \subset M$ — гиперболическое множество диффеоморфизма $f: U \rightarrow M$. Тогда для любой открытой окрестности $V \subset U$ множества Λ и любого $\delta > 0$ существует такое $\epsilon > 0$, что если $f': U \rightarrow M$ — гомеоморфизм и $d_{C^1}(f|_V, f') < \epsilon$, то найдутся гиперболическое множество $\Lambda' = f'(\Lambda) \subset V$ для f' и гомеоморфизм $h: \Lambda' \rightarrow \Lambda$, удовлетворяющие условиям $d_{C^0}(\text{Id}, h) + d_{C^0}(\text{Id}, h^{-1}) < \delta$ и $h \circ f'|_{\Lambda'} = f|_{\Lambda} \circ h$. Кроме того, гомеоморфизм h единствен при достаточно малом δ .

Доказательство. Мы будем использовать теорему о семействах ϵ -траекторий 10.2.13 три раза. Сначала возьмем $\delta_0 < \delta$ и, применяя теорему при $\epsilon < \delta_0/2$, $Y = \Lambda$, $\alpha = \text{Id}|_{\Lambda}$ (включение) и $g = f$, получим единственное такое отображение $\beta: \Lambda \rightarrow U(\Lambda)$, что $\beta \circ f = f' \circ \beta$. Условие конусов для гиперболичности показывает, что орбиты, остающиеся достаточно близко от гиперболического множества, должны быть гиперболическими, т. е. множество $\Lambda' := \beta(\Lambda)$ гиперболическое.

Чтобы показать, что отображение β является инъективным, применим теорему о семействах ϵ -траекторий по-другому: выбирая $\epsilon < \delta_0/2$, $Y = \Lambda'$, $\alpha' = \text{Id}|_{\Lambda'}$ (вложение), $g = f'$, получим такое отображение h , что $h \circ f' = f \circ h$. Важно иметь в виду, что в теореме о семействах ϵ -траекторий можно использовать f' вместо f , если ϵ выбрано достаточно малым. Мы утверждаем, что $h \circ \beta = \text{Id}$, а следовательно, $h = \beta^{-1}$ — гомеоморфизм.

Применим теперь утверждение о единственности к случаю $f = f'$. Очевидно, что $\alpha \circ f = f \circ \alpha$ и в то же самое время, как сказано выше, $\beta \circ f = f \circ \beta$, где $\bar{\beta} := h \circ \beta$.

Поскольку $d_{C^0}(\alpha, \bar{\beta}) = d_{C^0}(\text{Id}, h \circ \beta) \leq d_{C^0}(\text{Id}, \text{Id} \circ \beta) + d_{C^0}(\text{Id} \circ \beta, h \circ \beta) = d_{C^0}(\text{Id}, \beta) + d(\text{Id}, h) < \delta_0$, из утверждения о единственности в теореме о семействах ϵ -траекторий следует, что $\bar{\beta} = \alpha = \text{Id}|_{\Lambda}$, как и требуется. $\square$

Одним из выдающихся достижений динамики XX в. было доказательство того факта, что гиперболические динамические системы — это *в точности* все C^1 -структурно устойчивые системы, т. е. что C^1 -структурная устойчивость эквивалентна гиперболичности². Соответствующий вопрос для C^2 -структурной устойчивости (которая определяется аналогично) все еще открыт.

Вообще говоря, классификация является весьма отдаленной целью, но для некоторых важных наборов гиперболических динамических систем эта цель была достигнута. Так, например, теорема 7.4.3, показывает, что растягивающее отображение окружности эквивалентно линейному растягивающему отображению той же степени. Приведем еще один важный пример: если гиперболическое множество является тором, тогда отображение на нем эквивалентно линейному отображению тора. Это линейное отображение можно определить из глобального условия, похожего на условие равенства степеней. Степень растягивающего отображения E_m равна m , и аналог степени отображения F_L — матрица L .³

Дальнейшие примеры классификации можно получить для гиперболических аттракторов в динамических системах малой размерности.

§ 10.3. Кодирование и перемешивание

10.3.1. Кодирование. Каждый из наших стандартных гиперболических примеров тесно связан с символической системой (см. п. 7.3.1, предложения 7.4.2, 7.4.4, 7.4.6, следствие 7.4.10). Причиной этого служит одна из важных особенностей гиперболической динамики, а именно возможность моделировать ее топологической цепью Маркова с точностью до некоторых «исправлений», необходимых для того чтобы принять в расчет пересечения на границах разбиения. При этом кодировании используется марковское разбиение гиперболического множества. Это конечное покрытие замкнутыми множествами, которые пересе-

²J. Robbin. A Structural Stability Theorem // Annals of Mathematics (2). 1971. V. 94. P. 447—493; R. C. Robinson. Structural Stability of C^1 Diffeomorphisms // Journal of Differential Equations. 1976. V. 22, no. 1. P. 28—73; R. Mañé. A Proof of the C^1 Stability Conjecture // Publications Mathématiques de l'Institut des Hautes Études Scientifiques. 1988. V. 66. P. 161—210.

³Об этом говорится, например, в книге: А. Каток, Б. Хасселблат. Введение в современную теорию динамических систем.

каются только на границах, обладающее следующим марковским свойством: если множества U, V, W из этого покрытия таковы, что $f(U) \cap V \neq \emptyset \neq f(V) \cap W$, то $f^2(U) \cap W \neq \emptyset$. Это свойство превращает символическую систему, определенную возможными маршрутами в соответствии с этим разбиением, в топологическую цепь Маркова. Марковское свойство является следствием геометрического утверждения о том, что если $f(U) \cap V \neq \emptyset$, то «множество $f(U)$ проходит через V ».

Через Int_Λ и ∂_Λ мы будем обозначать соответственно внутренность и границу множества Λ .

Определение 10.3.1. Пусть Λ — компактное локально максимальное гиперболическое множество. Выберем ϵ, δ и $[x, y]$ точно так же, как в предложении 10.1.7, и положим $\eta = \epsilon$. Тогда множество $R \subset \Lambda$ называется *прямоугольником*, если его диаметр меньше чем $\eta/10$ и из того, что $x, y \in R$, следует, что $[x, y] \subset R$. Прямоугольник R называется *собственным*, если $R = \text{Int}_\Lambda R$. Введем обозначение $W_R^i(x) := W_\eta^i(x) \cap R$ для $x \in R$, $i = u, s$ и положим $\partial^s R := \{x \in R: x \notin \text{Int}_{\Lambda \cap W_\eta^s(x)} W_R^u(x)\}$, $\partial^u R := \{x \in R: x \notin \text{Int}_{\Lambda \cap W_\eta^u(x)} W_R^s(x)\}$.

Марковское разбиение — это конечное покрытие $\mathcal{R} = \{R_0, \dots, R_{m-1}\}$ множества Λ такими собственными прямоугольниками, что

- 1) $\text{Int } R_i \cap \text{Int } R_j = \emptyset$ при $i \neq j$;
- 2) если $x \in \text{Int } R_i$ и $f(x) \in \text{Int } R_j$, то $W_{R_i}^u(f(x)) \subset f(W_{R_i}^u(x))$ и $f(W_{R_i}^s(x)) \subset W_{R_j}^s(f(x))$.

Хотя слово «прямоугольник» вполне естественно, элементы марковского разбиения обычно бывают не столь простыми, как в приведенных примерах. В общем случае, как видно уже на примере автоморфизмов тора в размерности 3 и выше, марковское разбиение имеет сложную геометрию с фрактальными границами.

Отметим следующий результат.

Лемма 10.3.2. Если R — прямоугольник, то $\partial_\Lambda R = \partial^s R \cup \partial^u R$.

Доказательство. Если $x \in \text{Int}_\Lambda R$, то $x \in \text{Int}_{\Lambda \cap W_\eta^u(x)}(R \cap W_\eta^u(x) \cap \Lambda) = \text{Int}_{\Lambda \cap W_\eta^u(x)} W_R^u(x)$, поскольку R представляет собой окрестность точки x в множестве Λ . Таким образом, $\partial^s R \subset \partial_\Lambda R$. Аналогично $\partial^u R \subset \partial_\Lambda R$. Если $x \in (\text{Int}_{\Lambda \cap W_\eta^s(x)} W_R^s(x)) \cap (\text{Int}_{\Lambda \cap W_\eta^u(x)} W_R^u(x))$, то в силу непрерывности отображения $[\cdot, \cdot]$ существует такая окрестность U точки x в Λ , что для всех $y \in U$ выполняются включения $[x, y], [y, x] \in R$, а следовательно, $y' := [[y, x], [x, y]] \in R \cap W_\eta^s(x) \cap W_\eta^u(y) \subset W_\eta^s(x) \cap W_\eta^u(y) \subset \{y\}$, так что $x \in \text{Int}_\Lambda R$. $\square$

Теорема 10.3.3. Для компактного локально максимального гиперболического множества существует марковское разбиение произвольно малого диаметра.

Набросок доказательства. Возьмем малое число $\delta > 0$ и выберем ϵ в соответствии с теоремой 10.2.13. Далее, выберем такое число $\gamma < \epsilon/2$, что $d(f(x), f(y)) < \epsilon/2$ при $d(x, y) < \gamma$, и γ -плотное множество $P := \{p_0, \dots, p_{N-1}\}$ в гиперболическом множестве Λ . Тогда множество $\Omega(P) := \{\omega \in \Omega_N: d(f(p_{\omega_i}), p_{\omega_{i+1}}) < \epsilon\}$ является топологической цепью Маркова. Для произвольной ϵ -орбиты из $\Omega(P)$ имеется единственное отображение $\beta(\omega) \in \Lambda$, которое δ -приближает отображение $\alpha(\omega) := \{p_{\omega_i}\}_{i \in \mathbb{Z}}$. Можно показать, что отображение β сюръективно

и непрерывно. Мы расширим соответствие, заданное скобкой $[\cdot, \cdot]$, на ε -орбиты, полагая

$$[\omega, \omega']_i = \begin{cases} \omega_i, & \text{если } i \geq 0, \\ \omega'_i, & \text{если } i \leq 0, \end{cases}$$

для всех $\omega, \omega' \in \Omega(P)$, $\omega_0 = \omega'_0$. Тогда отображение $[\cdot, \cdot]$ коммутирует с β , т. е. $\beta([\omega, \omega']) \in W_{2\delta}^s(\beta(\omega)) \cap W_{2\delta}^u(\beta(\omega')) = \{[\beta(\omega), \beta(\omega')]\}$.

Множество $R'_i := \{\beta(\omega) : \omega_0 = i\}$ — это прямоугольник, поскольку для $x = \beta(\omega)$, $y = \beta(\omega') \in R'_i$ мы имеем $[\omega, \omega']_0 = i$ и, таким образом, $[x, y] = [\beta(\omega), \beta(\omega')] = \beta([\omega, \omega']) \in R'_i$. Нетрудно получить утверждение 2) из определения 10.3.1. Для получения марковского разбиения нам необходимо построить собственные прямоугольники с попарно непересекающимися внутренностями. Для этого изменим наши прямоугольники.

Для $x \in \Lambda$ обозначим через $\mathcal{R}(x)$ множество прямоугольников из $\mathcal{R}'$, которые содержат точку x , и через $\mathcal{R}^*(x)$ — множество прямоугольников из $\mathcal{R}'$, которые пересекаются с прямоугольником из $\mathcal{R}(x)$. Тогда множество $A := \{x \in \Lambda : W_\eta^s(x) \cap \partial^s R'_i = \emptyset, W_\eta^u(x) \cap \partial^u R'_i = \emptyset \text{ для всех } i\}$ открыто и плотно. Если $R'_i \cap R'_j \neq \emptyset$, то разрежем R'_j на четыре прямоугольника следующим образом:

$$R(i, j, su) := R'_i \cap R'_j,$$

$$R(i, j, 0u) := \{x \in R'_j : W_{R'_i}^s(x) \cap R'_j = \emptyset, W_{R'_i}^u(x) \cap R'_j \neq \emptyset\},$$

$$R(i, j, s0) := \{x \in R'_j : W_{R'_i}^s(x) \cap R'_j \neq \emptyset, W_{R'_i}^u(x) \cap R'_j = \emptyset\},$$

$$R(i, j, 00) := \{x \in R'_j : W_{R'_i}^s(x) \cap R'_j = \emptyset, W_{R'_i}^u(x) \cap R'_j = \emptyset\}$$

и для $x \in A$ положим $R(x) := \bigcap \{\text{Int}_\Lambda R(i, j, q) : x \in R'_i, R'_i \cap R'_j \neq \emptyset, x \in R(i, j, q), q \in \{su, 0u, s0, 00\}\}$. Тогда $\overline{R(x)}$ — это прямоугольники, покрывающие множество $R'_i \cap A$, и $R(x)$ — конечный набор попарно непересекающихся открытых прямоугольников, так что набор

$$\mathcal{R} := \{\overline{R(x)} : x \in A\} = \{R_0, \dots, R_{m-1}\}$$

представляет собой конечное покрытие множества Λ собственными прямоугольниками с попарно непересекающимися внутренностями. Проверив справедливость свойства 2) из определения 10.3.1, убедимся, что это и есть искомое марковское разбиение. $\square$

Как следствие, мы получаем интересующий нас результат о существовании полусопряжения между компактным локально максимальным гиперболическим множеством и топологической цепью Маркова.

Теорема 10.3.4. *Если Λ — компактное локально максимальное гиперболическое множество, $\mathcal{R} = \{R_1, \dots, R_m\}$ — марковское разбиение достаточно малого диаметра и*

$$A_{ij} := \begin{cases} 1, & \text{если } R_i \cap f^{-1}(R_j) \neq \emptyset, \\ 0 & \text{в противном случае,} \end{cases}$$

то $f|_{\Lambda}$ является топологическим фактором топологической цепи Маркова (Ω_A, σ_A) . Полусопряжение $h: \Omega_A \rightarrow \Lambda$ инъективно на множестве $h^{-1}(\Lambda')$, где $\Lambda' := \Lambda \setminus \bigcup_{i \in \mathbb{Z}} f^i(\partial^s \mathcal{R} \cup \partial^u \mathcal{R})$ и $\partial^s \mathcal{R} := \bigcup_{R \in \mathcal{R}} \partial^s R$, $\partial^u \mathcal{R} := \bigcup_{R \in \mathcal{R}} \partial^u R$.

Доказательство. Для $\omega \in \Omega_A$ положим $h(\omega) = \bigcap_{i \in \mathbb{Z}} f^{-i}(R_{\omega_i})$. Используя марковское свойство, покажем, что это пересечение непусто. Для этого проверим свойство конечного пересечения. Пересечение не может содержать более одной точки в силу разделения. Отображение h непрерывно, а также сюръективно, потому что $h(\Omega_A)$ — компактное множество, содержащее Λ' . Ясно, что $h \circ \sigma_A = f \circ h$ и любая точка $x \in \Lambda'$ имеет только один прообраз. $\square$

10.3.2. Топологическое перемешивание и спектральное разложение.

Наши стандартные гиперболические примеры оказались не только топологически транзитивными, но также и топологически перемешивающими; см. предложение 7.2.7 (линейные растягивающие отображения), предложение 7.2.9 (автоморфизмы тора), следствие 7.4.7 (подковы), теорему 7.4.3 (нелинейные растягивающие отображения) и предложение 7.3.4 (квадратичные отображения посредством кодирования).

Как следует из предложения 7.3.12, транзитивная топологическая цепь Маркова является топологически перемешивающей. Однако топологическая цепь Маркова может быть топологически транзитивна и без транзитивной матрицы перехода. Проблема в том, что могут иметься ограничения на моменты времени, в которые образы одного множества пересекаются. Однако топологически транзитивная топологическая цепь Маркова всегда переставляет лишь конечное число частей с топологически перемешивающим отображением возвращения. Приведем краткое пояснение этого обстоятельства. Предположим, что A — это $(m \times m)$ -матрица из нулей и единиц, в каждом столбце и каждой строке которой имеется по крайней мере одна единица. Если $i \in \{0, \dots, m-1\}$, тогда $\Omega_{A,i} = \{\omega \in \Omega_A : \omega_0 = i\} \neq \emptyset$. Если найдется элемент $\omega \in \Omega_A$, который содержит символ i по крайней мере дважды, тогда символ i называется *существенным*. Два существенных символа i и j эквивалентны, если существуют такие элементы $\omega, \omega' \in \Omega_A$, $k_1 < k_2$, $l_1 < l_2$, что $\omega_{k_1} = \omega'_{l_2} = i$, $\omega_{k_2} = \omega'_{l_1} = j$. Это отношение является отношением эквивалентности. Если отображение σ_A имеет плотную положительную полуорбиту, тогда все символы существенны и эквивалентны. Пусть N — наибольший общий делитель длин циклов (последовательностей, начало и конец которых совпадают). отождествим два символа, если они связаны путем, длина которого кратна N . Пусть $\Lambda_1, \dots, \Lambda_N = \Lambda_0$ — классы эквивалентности. Покажите, что ограничение отображения $(\sigma_A)^N$ на каждое множество Λ_i является топологически перемешивающим.

Аналогично можно разложить гиперболические множества с помощью кодирования, даже если они не топологически транзитивны. Укажем, как получить такое разложение непосредственно. Квадратичные отображения и подковы являются топологически перемешивающими, так сказать, по счастливой случайности. Но для растягивающих отображений и линейных отображений на торе это не

случайно. Это происходит потому, что не существует нетривиального разбиения окружности или тора на компактные множества. По той же причине любое связанное гиперболическое множество является топологически перемешивающим.

Прежде чем искать разложение на перемешивающие части, внесем одну поправку. Гиперболическое множество может содержать притягивающую и отталкивающую точки вместе с гетероклинической орбитой (см. определение 2.3.4). Это свойство несовместимо с транзитивностью, поэтому мы должны отказаться от таких точек.

Определение 10.3.5. Точка $x \in X$ называется *неблуждающей* относительно отображения $f: X \rightarrow X$, если для любого открытого множества $U \ni x$ найдется такое $N > 0$, что $f^N(U) \cap U \neq \emptyset$. Множество всех неблуждающих точек отображения f обозначается через $NW(f)$ ⁴.

Итак, мы можем получить спектральное разложение неблуждающего множества компактного локально максимального гиперболического множества на конечное число компонент, на которых соответствующая итерация отображения f является топологически перемешивающей.

Теорема 10.3.6 (спектральное разложение). Пусть M — риманово многообразие, $U \subset M$ — открытое множество, $f: U \rightarrow M$ — диффеоморфизм и $\Lambda \subset U$ — компактное локально максимальное гиперболическое множество для диффеоморфизма f . Тогда существуют такие непересекающиеся замкнутые множества $\Lambda_1, \dots, \Lambda_m$ и такая перестановка σ на множестве $\{1, \dots, m\}$, что $NW(f|_\Lambda) = \bigcup_{i=1}^m \Lambda_i$, $f(\Lambda_i) = \Lambda_{\sigma(i)}$ и если $\sigma^k(i) = i$, то отображение $f^k|_{\Lambda_i}$ является топологически перемешивающим.

При доказательстве используется следующее отношение эквивалентности на множестве $\text{Per}(f|_\Lambda)$ периодических точек из Λ : $x \sim y$ тогда и только тогда, когда $W^u(x) \cap W^s(y) \neq \emptyset$ и $W^s(x) \cap W^u(y) \neq \emptyset$, причем оба пересечения трансверсальны по крайней мере в одной точке. Множества Λ_i — это замыкания классов эквивалентности.

§ 10.4. Статистические свойства

В гиперболической динамике важно взаимодействие между топологическими и статистическими свойствами. В §§ 7.5 и 7.6 выделяются свойства, которые удобно использовать при изучении статистических свойств гиперболических систем. Это отсутствие строгой эргодичности (п. 7.5.1), равномерное распределение/эргодичность (предложение 7.5.1, теорема 7.5.6, п. 7.5.4) и перемешивание (предложения 7.6.6, 7.6.7).

10.4.1. Растягивающие отображения; трудности, возникающие при использовании прямых методов. Нелинейные растягивающие отображения ок-

⁴Обозначение $NW(f)$ связано с английским термином «nonwandering point» (неблуждающая точка). — Прим. ред.

ружности S^1 служат моделью для изучения некоторых общих особенностей статистического анализа.

Как следует из теоремы 7.5.6, для удваивающего отображения E_2 на S^1 средние Биркгофа $B_n(\varphi)(x)$ непрерывной функции φ сходятся к $\int_{S^1} \varphi(t) dt$ почти всюду. Хотелось бы получить аналогичный результат для произвольных растягивающих отображений (хотя бы для тех, которые имеют степень 2), хотя они не сохраняют длину. Мы должны искать равномерное распределение с весом, как предлагается в п. 7.6.4, где орбиты имеют искаженное, но согласованное асимптотическое распределение. В идеале должна существовать такая непрерывная функция $g: [0, 1] \rightarrow [0, 1]$, что «длина с весом» $\int_I g(x) dx$ интервалов I является инвариантной относительно растягивающего отображения (см. также п. А.3.2). Казалось бы, задать точную функцию просто, поскольку нелинейное растягивающее отображение степени 2 топологически сопряжено с E_2 (см. теорему 7.4.3). Если h — сопряжение, тогда $h \circ f = E_2 \circ h$, так что плотность g должна определяться из условия $\int_a^b g(x) dx = h(b) - h(a)$. Однако применить этот подход не удастся, потому что сопряжение h напоминает функции распределения, показанные на рис. 7.6.1. Оно не дифференцируемо, в отличие от выражения $\int_a^b g(x) dx$ для непрерывной функции g . Так что мы не можем получить функцию плотности таким способом. Это довольно тонкий вопрос, но он связан со следующим элементарным замечанием: сопряжение не может быть дифференцируемым, если производная отображения f в неподвижной точке не равна 2, т. е. не совпадает с производной отображения E_2 .

Хотя построение инвариантной плотности не тривиально, такая плотность все же существует, и можно поставить вопрос о равномерном распределении относительно этой функции. Для нелинейных растягивающих отображений, особенно степени 2, можно доказать, что распределение равномерно (в принципе), с помощью вероятностных рассуждений, как в доказательстве предложения 7.6.1.

Это довольно трудный вопрос, хотя, решив его, мы можем получить гораздо более сильные статистические свойства, чем просто равномерное распределение. Удобнее всего использовать эргодическую теорему Биркгофа⁵. Эта теорема позволяет доказать, что распределение равномерно, с помощью понятия эргодичности, определенного в рамках теории меры. Мы не будем рассматривать теорию меры в этой книге, но отметим, что с ее помощью проверить эргодичность в данном контексте несложно.

Итак, растягивающие отображения имеют инвариантную плотность и почти все точки равномерно распределены с этой плотностью.

10.4.2. Изобилие инвариантных мер. Чтобы понять, почему строгая эргодичность отсутствует в гиперболических динамических системах, полезно описать это явление в терминах чисто качественных свойств. Это можно сделать с использованием инвариантных мер.

⁵ А. Каток, Б. Хасселблат. Введение в современную теорию динамических систем, теорема 4.1.2.

Определение 10.4.1. Пусть X — компактное метрическое пространство и отображение $f: X \rightarrow X$ непрерывно. *Инвариантный интеграл* для f — это действительнозначное линейное отображение (линейный функционал) $\mathcal{I}: C(X) \rightarrow \mathbb{R}$ на пространстве непрерывных функций на X , отличное от нуля, которое является непрерывным в топологии равномерной сходимости (т.е. если $\phi_n \rightarrow \phi$ равномерно, то $\mathcal{I}(\phi_n) \rightarrow \mathcal{I}(\phi)$) и f -инвариантным (т.е. $\mathcal{I}(\phi \circ f) = \mathcal{I}(\phi)$ для любой функции $\phi \in C(X)$). Если $\mathcal{I}(1) = 1$, то инвариантный интеграл также называется *инвариантной мерой* или инвариантной вероятностью.

Простой пример описан в п. 11.4.3.2 и А.3.2. Он задается соотношением $\mathcal{I}(\phi) := \int \phi \rho dx$, где ρ — некоторая подходящая функция. Меры, которые можно представить таким образом, называются *абсолютно непрерывными*.

Предложение 10.4.2. *Преобразование f строго эргодично тогда и только тогда, когда оно имеет в точности одну инвариантную меру.*

В терминах этого результата нарушение строгой эргодичности, о котором говорится в п. 7.5.1, можно переформулировать как наличие нескольких инвариантных мер. Действительно, гиперболические динамические системы имеют много инвариантных мер. Это очевидно, потому что имеется много периодических точек (см. предложения 7.1.2, 7.1.3, 7.1.10 и следствие 7.4.7) и каждая периодическая орбита $\mathcal{O}(x) = \{x, f(x), \dots, f^{n-1}(x)\}$ порождает инвариантную меру $\mathcal{I}$, определенную равенством $\mathcal{I}(\phi) = \sum_{i=0}^{n-1} \phi(f^i(x))/n$. Конечно же, эти меры не являются абсолютно непрерывными.

Есть различные методы, позволяющие получить из этих инвариантных мер новые. Самый простой из них — *выпуклая комбинация*: если $\mathcal{I}$ и $\mathcal{J}$ — инвариантные меры и $a, b \geq 0$, $a + b = 1$, то $a\mathcal{I} + b\mathcal{J}$ — инвариантная мера. Другими словами, пространство инвариантных мер выпукло (см. определение 2.2.13). В частности, конечные выпуклые комбинации «периодических мер» являются инвариантными мерами.

Хотя это не совсем очевидно, пространство инвариантных мер также компактно⁶. Поэтому каждая последовательность инвариантных мер $(\mathcal{I}_i)_{i \in \mathbb{N}}$ имеет сходящуюся подпоследовательность. Ее предел определяет новую инвариантную меру. Применяя эти предельные процессы к множеству периодических мер, мы получим большое многообразие инвариантных мер. Теорема о спецификации дает конкретный способ построения периодических орбит и, таким образом, используется для построения инвариантных мер с определенными свойствами, таких как меры Гиббса.

10.4.3. Равномерное распределение. Рассматривая лемму об ε -траекториях (см. п. 10.2.2), мы отметили, что из этой леммы не следует, что вычисленные орбиты являются типичными. Естественный пример удваивающего отображения в то же время весьма специфичен: двоичные рациональные числа, поведение которых мы описываем, образуют множество меры нуль, причем его орбиты не распределены равномерно на отрезке.

⁶Это следует из теоремы Алаоглу в функциональном анализе.

1. *Абсолютно непрерывные инвариантные меры.* Рассматривая удваивающее отображение в п. 7.6.4, мы получили пример равномерного асимптотического распределения со сложной функцией распределения. Это расширенное понятие равномерного распределения подходит для гиперболических динамических систем в общей ситуации; краткое описание соответствующей структуры см. в § А.3. Новый ряд примеров возникает из квадратичного семейства, которое имеет абсолютно непрерывную инвариантную меру для многих значений параметра (см. § 11.4). Это приводит к *равномерному распределению с весом*, которое несколько сильнее, чем равномерное асимптотическое распределение, потому что имеется подходящая функция плотности, описывающая распределение. Здесь следует вновь отметить, что это явление в комбинации с отсутствием строгой эргодичности (что отражает довольно разнородную орбитальную структуру) является одним из аспектов очень сложного асимптотического поведения. Мы уже не раз встречались с равномерным распределением; см. предложение 7.5.1, теорему 7.5.6, п. 7.5.4, предложения 7.6.6 и 7.6.7.

2. *Абсолютно непрерывные меры на аттракторах.* Глава 13 посвящена (странным) аттракторам. При их изучении чрезвычайно важны вопросы, связанные с равномерным распределением. Сформулируем эти вопросы более аккуратно. Орбиты должны быть в некотором смысле равномерно распределены на аттракторе, и, так как аттрактор зачастую представляет собой множество меры нуль в фазовом пространстве, это требует существенного уточнения понятия равномерного распределения.

Рассмотрим в качестве модели соленоид, изображенный на рис. 13.2.1. Под действием рассматриваемого отображения точки из аттрактора расходятся, а точки, расположенные вне аттрактора, двигаются к нему. Не учитывая точки, расположенные вне аттрактора, мы можем считать, что равномерное распределение есть равномерное распределение орбиты на аттракторе. Его естественно определить, рассматривая меры *на аттракторе*, т.е. меры, по которым можно интегрировать непрерывные функции, определенные только на аттракторе. Плотность (для равномерного распределения с весом) должна определяться только на аттракторе, т.е. мы хотим получить абсолютно непрерывную меру на аттракторе (см. § А.3). Чтобы получить равномерное распределение, нужно также добиться, чтобы мера была эргодической.

Проблема состоит в том, что мы ищем функционал вида $\mathcal{I}(\phi) := \int \phi g dx$, не имея строгого определения интеграла. Дело в том, что весь аттрактор является множеством меры нуль. Но, рассматривая соленоид, можно придать смысл понятию интеграла, потому что аттрактор локально представляет собой накопление гладких кривых (неустойчивых многообразий) и мы можем определить интеграл по любому участку таких кривых, хотя он и является множеством меры нуль в $\mathbb{R}^3$. Чтобы получить точное определение, необходимо описать, как переходить на отдельное неустойчивое многообразие. Поскольку эти многообразия являются множествами меры нуль даже в пределах аттрактора, требуется сделать нормализацию. Для этого возьмем отрезок I в некотором многообразии и предположим, что I_n — объединение устойчивых дисков радиуса $1/n$, центр которых

лежит в I , т. е. трубка радиуса $1/n$ вокруг множества I . Искомая мера должна обладать следующим свойством: найдется такая функция g на аттракторе, что для любого отрезка I и любой непрерывной функции ϕ выполняется равенство $\lim_{n \rightarrow \infty} \mathcal{I}(\phi\chi_{I_n})/\mathcal{I}(\chi_{I_n}) = \int_I \phi g dx$, где dx обозначает длину дуги на I .

Это объясняет, что мы подразумеваем под эргодической инвариантной мерой на аттракторе, абсолютно непрерывной на неустойчивых многообразиях, и почему естественно пытаться ее найти. Мы выяснили, что наличие такой меры позволяет получить равномерное распределение (с весом) почти всех орбит, которые лежат на аттракторе. (Мы можем говорить о «почти всех» орбитах, потому что множества меры нуль на гладких кривых имеют смысл.)

3. Меры Синяя—Рюэля—Боуэна. Мера, обладающая указанными выше свойствами, называется мерой Синяя—Рюэля—Боуэна (или SRB-мерой). Пока что описание таких мер представляет серьезные трудности. Странные аттракторы не описываются аналитически. Их можно получить как компьютерные картинки, если двигаться вдоль случайной орбиты. Но так как аттрактор представляет собой множество меры нуль, орбита, построенная с помощью компьютера, может не лежать на аттракторе. Предшествующие рассуждения ничего не говорят об этом случае. Необходимо связать асимптотику орбиты, которая приближается к аттрактору, с асимптотикой орбит на аттракторе. Это вновь можно проиллюстрировать с помощью соленоида. Точка x , расположенная вблизи аттрактора, находится в устойчивом многообразии некоторой точки p , лежащей на аттракторе, потому что (в данном случае) каждое сечение полнотория, определяемое фиксированным значением угла φ , является устойчивым многообразием. Поэтому движение точки x под действием отображения разлагается на движение к аттрактору и движение точки p , которая определяет движение устойчивого сечения. Это означает, что, как только итерации точки x подходят достаточно близко к аттрактору, они становятся практически неотличимыми от соответствующих итераций точки p . Если они равномерно распределены на аттракторе, тогда орбита точки x отражает плотность SRB-меры.

Мы хотим показать, что почти каждая близлежащая орбита отображения соленоида в конечном счете будет иметь описанное выше равномерное распределение. Для этого нужно исключить из рассмотрения все точки x , находящиеся на устойчивых многообразиях точек p из множества меры нуль «плохих» точек для SRB-меры. По определению множества меры нуль (на участке кривой из аттрактора) множество таких точек p можно покрыть счетным числом интервалов, сумма длин которых сколь угодно мала. Объединение устойчивых сечений, проходящих через все точки этих «опасных» интервалов, есть счетное множество «долек» полнотория, которые по форме близки к прямоугольникам⁷, и объем в $\mathbb{R}^3$ каждого из них совпадает с длиной базового отрезка на соленоиде с точностью до множителя, равного площади устойчивой части. Поэтому сумма объемов этих долек сколь угодно мала и «плохие» точки из этой окрестности образуют множество меры нуль в $\mathbb{R}^3$.

⁷Считая, что такая аппроксимация закона, мы обходим сложный технический вопрос, который можно разрешить лишь с использованием абсолютной непрерывности устойчивых слоев.

Один из важных результатов теории гиперболических динамических систем состоит в том, что гиперболический аттрактор имеет меру Синая—Рюэлля—Боуэна. Как было показано выше, из этого следует, что все точки в области притяжения аттрактора, за исключением множества меры нуль, распределены равномерно с соответствующей плотностью. Это дает нам гарантию, что изображения «странных аттракторов», полученные с помощью численных методов, соответствуют реальному аттрактору. Мера Синая—Рюэлля—Боуэна также называется «физически наблюдаемой мерой».

§ 10.5. Неравномерно гиперболические динамические системы

В гиперболической динамике имеются замечательные утверждения о системах с очень сложной структурой орбит. Предположение о равномерной гиперболическости позволяет довольно легко получить даже очень сильные результаты.

Как и в наших примерах, сложность структуры орбит в системах, возникающих в реальной жизни, в значительной степени проистекает из комбинации растяжения и изгиба. Однако, как мы увидим в последующих главах, изгибы не всегда столь же аккуратны. Чтобы расширить применимость гиперболической теории, в нее были включены системы, в которых каждая точка подчинена гиперболическому действию производной, но коэффициенты сжатия и растяжения меняются в зависимости от точки, не будучи равномерно отделенными от 1. Теория неравномерно гиперболических динамических систем более тесно связана с теорией меры, чем описанная нами в общих чертах исходная теория. Поэтому в этой книге мы рассматриваем ее только на примерах. Обратим внимание на несколько основных особенностей.

Может случиться, что инвариантные направления E^u и E^s не определены на множестве меры нуль. Кроме того, они могут не зависеть непрерывно от точки, и может не существовать нижней границы для угла между ними. Устойчивые и неустойчивые многообразия существуют, но, так как в доказательстве этого утверждения используются локальные величины и нет равномерности в размере окрестностей, размеры листов не равномерны, и углы между ними трудно контролировать. В отличие от равномерно гиперболической картины (как для линейных автоморфизмов тора, где вдоль неустойчивой кривой строятся устойчивые кривые, образующие равномерный пучок листов), хорошей иллюстрацией неравномерной ситуации может служить «забор», состоящий из отрезков, пересекающих канторово множество (но не множество меры нуль) на неустойчивой кривой, при этом в «дырах» содержатся подобные «заборы» с более короткими и более изогнутыми частями, и т. д.

Поскольку исследование неравномерно гиперболических систем основано на теории меры, возникает большой интерес к стохастическому поведению. Это играет определенную роль при изучении квадратичных отображений и аттракторов. Для последних всегда важно установить существование (и единственность) меры Синая—Рюэлля—Боуэна, чтобы подтвердить правильность картин, полученных численными методами.

КВАДРАТИЧНЫЕ ОТОБРАЖЕНИЯ

§ 11.1. Предварительные замечания

11.1.1. Простое и сложное поведение в квадратичном семействе. Квадратичное или логистическое семейство отображений

$$f_{\lambda}(x) = \lambda x(1 - x)$$

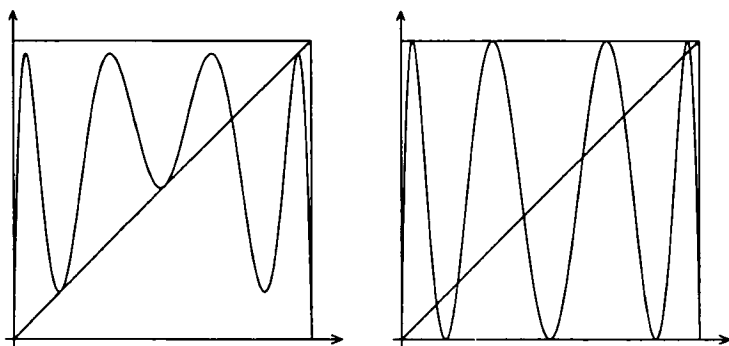
на отрезке $[0, 1]$ впервые было рассмотрено в п. 1.2.9, где мы изучали модели популяций. В п. 7.1.2 и 7.4.3 наблюдалось более сложное поведение для некоторых значений параметра. Действительно, для целого ряда значений параметра это семейство демонстрирует изумительное многообразие динамических свойств. В данной главе мы попытаемся объяснить это замечательное явление.

В § 2.5 при изучении квадратичного семейства мы получили примеры монотонных отображений отрезка, которые имеют довольно простое асимптотическое поведение, допускающее точное описание в очень простых терминах. Такое поведение встречается при $0 \leq \lambda \leq 3$. Подобному анализу поддаются и отображения с несколько большими значениями параметра. В следующем параграфе мы в общих чертах опишем первый шаг этого анализа (не приводя полного доказательства) и объясним, что происходит после следующей существенной бифуркации (каскад *удвоений периодов*). В течение этого процесса постепенного усложнения поведение все еще можно считать простым в том смысле, что все асимптотические режимы периодические, хотя и с возрастающим числом орбит различных периодов. Однако такое поведение наблюдается только до $\lambda_{\infty} \approx 3,58$.

Для больших значений λ топологическая энтропия отображения f_{λ} положительна, и сложность глобального поведения растет. В частности, топологическая энтропия возрастает с ростом λ , и появляются точки с периодами, отличными от степеней числа 2. Для $\lambda = 1 + \sqrt{8}$ впервые появляется орбита периода 3 (см. рис. 11.1.1). К этому моменту имеются периодические точки всех периодов (см. предложение 11.3.8). Однако по мере увеличения значения λ степень однородности в асимптотическом поведении меняется совершенно нерегулярно.

Соответственно такое сосуществование большого количества периодических орбит не противоречит тому, что для большинства (но не для всех) начальных условий асимптотическое поведение довольно просто. При $3,8284 \approx 1 + \sqrt{8} \leq \lambda \leq 3,841499008$ ¹ итерации случайно выбранного начального условия демонстри-

¹Явное выражение для этого значения параметра в радикалах весьма громоздко: $\lambda = 1 + \sqrt{52800 - 3900z + 285z^2 + 15\sqrt{201}z(20 + z)/120}$, где $z = \sqrt[3]{460 + 60\sqrt{201}}$. Это значение найдено Шарон Чуба и Эндрю Шерером. Далее мы будем иметь дело с более простыми значениями параметра.

Рис. 11.1.1. Отображения $f_{1+\sqrt{8}}^3$ и f_4^3

руют асимптотическое периодическое поведение с периодом 3 (см. рис. 11.2.3). С другой стороны, при других значениях параметра, например при $\lambda = 4$, асимптотическое поведение равномерно распределено (с плотностью), так же, как для линейного растягивающего отображения (см. п. 11.4.3).

Для того чтобы описать возможные типы поведения и установить, как они появляются при изменении параметра, были проведены серьезные исследования квадратичного семейства и подобных классов одномерных отображений. Некоторые из структурных особенностей отражают общие темы теории динамических систем (гиперболичность, структурная устойчивость, марковские разбиения, существование абсолютно непрерывных инвариантных мер), в то время как другие специфичны только для одномерной динамики (упорядочение Шарковского для орбит (теорема 11.3.9), теория инвариантов перемешивания). При исследовании используются как методы общей дифференциальной динамики, например теоремы о неподвижной точке, основанные на гиперболичности, так и методы, применимые только для размерности один (теорема о промежуточном значении, оценки искажения), и определенные свойства аналитических функций или даже многочленов. Все это делает квадратичное семейство отличным «полигоном» для доказательств с использованием различных подходов к динамике и модельным примером для демонстрации и преодоления трудностей, которые возникают даже в просто формулируемых задачах.

Благодаря многим по-настоящему глубоким работам основные черты структуры квадратичного семейства (иногда называемого также *действительным квадратичным семейством*) в настоящий момент хорошо поняты.

11.1.2. Притягивающие периодические орбиты. Начиная с предложения 11.2.1 мы описываем первые бифуркации в квадратичном семействе, которые приводят к появлению разнородных асимптотик. Хотя мы можем обнаружить и классифицировать периодические орбиты с помощью явных вычислений, наши результаты можно усилить, если использовать некоторые топологические факты. Таким образом, в этом пункте мы рассмотрим топологические утверждения, ис-

пользующиеся при доказательстве существования притягивающих периодических точек².

Напомним, что m -периодическая точка x_0 непрерывного отображения f отрезка I в себя называется *притягивающей*, если существует такое $\epsilon > 0$, что из неравенства $|x - x_0| < \epsilon$ следует, что $|f^n(x) - f^n(x_0)| \rightarrow 0$ при $n \rightarrow \infty$. Это равносильно тому, что точка x_0 является притягивающей неподвижной точкой для f^m (см. определение 2.2.22). Периодическая орбита $\{x_0, f(x_0), \dots, f^{m-1}(x_0)\}$ называется *притягивающей*, если одна из ее точек (а следовательно, и все) притягивающая.

Пусть $f: I \rightarrow I$ — дифференцируемое отображение и x_0 — m -периодическая точка. Если $|(f^m)'(x_0)| < 1$, то точка x_0 является притягивающей (см. предложение 2.2.17), а если $|(f^m)'(x_0)| > 1$, то не является. Если $|(f^m)'(x_0)| = 1$, она может быть притягивающей, а может и не быть.

Определение 11.1.1. Область притяжения притягивающей периодической орбиты $\mathcal{O} = \{x_0, f(x_0), \dots, f^{m-1}(x_0)\}$ — это множество таких точек x , что $|f^n(x) - f^{n+k}(x_0)| \rightarrow 0$ при $n \rightarrow \infty$ для некоторого k .

Область притяжения орбиты $\mathcal{O}$ состоит из точек, асимптотическое поведение которых повторяет поведение одной из точек орбиты. Такая точка единственна, так как различные точки орбиты не могут приближаться друг к другу. Таким образом, можно говорить о *фазе* для каждой точки из области притяжения.

Определение 11.1.2. Область непосредственного притяжения периодической точки x_0 — это максимальный такой интервал J , содержащий точку x_0 , что $|f^n(x) - f^n(x_0)| \rightarrow 0$ при $n \rightarrow \infty$ для любой точки $x \in J$. Область непосредственного притяжения периодической орбиты — это объединение областей непосредственного притяжения всех точек орбиты.

Замечание 11.1.3. Область притяжения и область непосредственного притяжения — открытые множества.

Лемма 11.1.4. Пусть $I = [a, b]$ и отображение $f: I \rightarrow I$ вогнуто и дважды дифференцируемо, причем $f(a) = f(b) = a$. Тогда область непосредственного притяжения любой притягивающей периодической орбиты содержит критическую точку.

Доказательство. Если $f'(a) \leq 1$, тогда в силу вогнутости $f'(x) < 1$ для $x > a$ и, следовательно, $f(x) < x$ для $x > a$, так что отрезок I лежит в области непосредственного притяжения точки a . В этом случае утверждение очевидно. Таким образом, мы можем предположить, что $f'(a) > 1$.

Пусть интервал J — область непосредственного притяжения для m -периодической точки x_0 с концами $c < d$. Тогда $f^m(J)$ является интервалом по теореме о промежуточном значении и содержит точку x_0 . Интервал $J \cup f^m(J)$ принадлежит области непосредственного притяжения точки x_0 , так что $f^m(J) \subset J$.

²Первоначальные удвоения были изучены Мирбергом в ряде работ 1950—1960 гг.; см., например: P.J. Myrberg, Iteration der reellen Polynome zweiten Grades. II // Annales Academiæ Scientiarum Fennicæ Mathematica Ser. A I. 1959. V. 268.

Если, вопреки утверждению, ни один из интервалов $f^i(J)$ ($0 \leq i \leq m-1$) не содержит критическую точку отображения f , то отображение f^m монотонно на J . Получим отсюда противоречие.

Рассмотрим сначала случай, когда один из образов концевых точек является концом отрезка J . Поскольку $f(b) = a$, мы можем без потери общности предполагать, что это a . Заменяя точку x_0 ее итерацией, мы можем предполагать, что a — конец интервала J , т. е. $c = a$.

Поскольку a является неподвижной точкой, из условия $f'(a) > 1$ следует, что $(f^m)'(a) > 1$, и, таким образом, производная $(f^m)'$ положительна и убывает на $[a, d]$ благодаря наличию притягивающей неподвижной точки x_0 . (Мы используем правило дифференцирования сложной функции и то обстоятельство, что отображение f^m монотонно на $[a, d]$.) Таким образом, $0 < (f^m)'(d) < (f^m)'(x_0) < 1$, так что отображение f^m является сжимающим на отрезке $[x_0, d]$ и $d < b$, поскольку отрезок $[a, d]$ не содержит критическую точку отображения f . Поэтому отображение f является сжимающим на отрезке $[x_0, d + \epsilon]$ для достаточно малого $\epsilon > 0$, что противоречит максимальности интервала J .

Следовательно, ни один из образов точек c или d не является концом отрезка J . Тогда из условия $f^m(J) \subset J$ следует, что $f^m(J) = J$, поскольку ни одна из точек c или d не лежит в области притяжения. Следовательно, c и d являются периодическими точками периода m или $2m$. Поскольку производная отображения f^m (и, следовательно, f^{2m}) имеет постоянный знак на J и отображение f вогнуто, производные отображений f^m и f^{2m} монотонны на J согласно правилу дифференцирования сложной функции (они являются произведениями монотонных функций с постоянными знаками). Среднее для $(f^m)'$ (и для $(f^{2m})'$) по интервалу J равно ± 1 , потому что $f^m(J) = J$, так что значение в одном из концов по модулю меньше 1, и по меньшей мере одна из точек c или d является притягивающей для f . Это невозможно, поскольку в этом случае область непосредственного притяжения этой точки пересекалась бы с J , а J является областью непосредственного притяжения точки x_0 . $\square$

Предложение 11.1.5. Пусть $I = [a, b]$ и отображение $f: I \rightarrow I$ вогнуто и дважды дифференцируемо, причем $f(a) = f(b) = a$. Тогда f имеет не более одной притягивающей периодической орбиты.

Доказательство. Области непосредственного притяжения различных периодических орбит не пересекаются. Поскольку отображение f имеет одну критическую точку, предложение следует из леммы 11.1.4. $\square$

Следствие 11.1.6. Квадратичное отображение f_λ имеет не более одной притягивающей периодической орбиты.

Можно задать вопрос, что происходит, если имеется периодическая точка вне области притяжения. Позже мы увидим, что поведение на остающемся множестве R , которое является инвариантным и в положительном, и в отрицательном направлениях, может быть весьма сложным. Это множество называется *универсальным отталкивающим множеством* (или *репеллером*³); см. кни-

³От англ. *repel* — отталкивать. — Прим. ред.

гу А. Катка и Б. Хасселблата «Введение в современную теорию динамических систем».

Следующее важное предложение справедливо и в более общей ситуации, но мы сформулируем его для квадратичного семейства.

Теорема 11.1.7. *Если квадратичное отображение f_λ имеет притягивающую периодическую точку, тогда универсальное отталкивающее множество является нигде не плотным множеством меры нуль.*

Таким образом, и в топологическом, и в вероятностном смысле большинство точек притягиваются к единственной периодической орбите.

Внутренняя структура ограничения квадратичного отображения на универсальное отталкивающее множество хорошо описывается цепями Маркова (см. определение 7.3.2, п. 7.3.7).

Теорема 11.1.8. *Ограничение квадратичного отображения с притягивающей периодической точкой на универсальное отталкивающее множество топологически сопряжено односторонней топологической цепи Маркова (см. определение 7.3.2).*

Обе теоремы следуют из *гиперболичности* универсального отталкивающего множества (см. определение 10.1.2). Этим объясняется следующая терминология.

Определение 11.1.9. Квадратичные отображения с притягивающей периодической точкой называются *гиперболическими*.

Замечание 11.1.10. Поскольку существование притягивающих периодических точек — это условие, которое выполняется на открытом множестве, множество значений параметра, для которых квадратичное отображение гиперболично, является открытым.

В свете приведенных рассуждений становятся естественными и интересными следующие вопросы.

Насколько велико множество значений параметра λ , для которых квадратичное отображение f_λ является гиперболическим, т. е. имеет притягивающую периодическую орбиту? В частности, является ли оно плотным? Является ли его дополнение множеством меры нуль?

На оба вопроса получены ответы: отрицательный ответ на последний вопрос получен Михаилом Якобсоном в начале 80-х гг., другой вопрос решен положительно Грегоржем Швентеком и его соавторами в конце 90-х гг. Их работы являются высшими достижениями в одномерной динамике в соответствующие периоды. В результате этих исследований установлены две главные парадигмы поведения в одномерной динамике, которые мы будем называть «гиперболической» и «стохастической».

Перед тем как перейти к подробному описанию всей картины, изучим изменение поведения после первой критической бифуркации при $\lambda = 3$. В этой ситуации мы можем наблюдать последовательность удвоений периодов.

§ 11.2. Эволюция простого поведения после первой бифуркации

11.2.1. Первое удвоение периода. Для следующего интервала значений параметра после $\lambda > 3$ орбиты асимптотически стремятся к периодическим орбитам, а не к неподвижной точке.

Предложение 11.2.1. Для $3 < \lambda \leq 1 + \sqrt{6}$ все орбиты отображения $f_\lambda(x) = \lambda x(1-x)$ на отрезке $[0, 1]$, кроме точек 0 и x_λ и их прообразов, асимптотически стремятся к единственной периодической орбите периода два.

Набросок доказательства. В отличие от в предыдущего случая (см. предложение 2.5.2) мы не будем доказывать утверждение полностью, а просто вычислим периодические точки и определим их устойчивость. Мы могли бы исследовать полную динамику тем же способом, как и прежде, но рассуждение в этом случае было бы несколько длиннее. Чтобы показать, что на самом деле имеется орбита периода 2, мы будем искать неподвижные точки отображения f_λ^2 , решая уравнение четвертой степени

$$x = f_\lambda^2(x) = f_\lambda(f_\lambda(x)) = \lambda(\lambda x(1-x))(1-\lambda x(1-x)).$$

Сначала отбросим уже известные решения (неподвижные точки), разделив разность $f_\lambda^2(x) - x$ на $f_\lambda(x) - x$:

$$\frac{f_\lambda^2(x) - x}{f_\lambda(x) - x} = \frac{\lambda^2 x(1-x)(1-\lambda x(1-x)) - x}{\lambda x(1-x) - x} = -(\lambda x)^2 + (\lambda + 1)\lambda x - (\lambda + 1).$$

Это квадратичная функция переменной λx с корнями

$$\lambda x = \frac{-(\lambda + 1) \pm \sqrt{(\lambda + 1)^2 - 4(\lambda + 1)}}{-2},$$

так что мы получаем две точки периода два:

$$x_{\lambda,\pm} = \frac{\lambda + 1 \pm \sqrt{(\lambda + 1)(\lambda - 3)}}{2\lambda}. \quad (11.2.1)$$

Укажем несколько особенностей этого явного описания. Для $\lambda < 3$ действительные решения отсутствуют и $x_{3,\pm} = 2/3 = x_3$. Таким образом, эти две периодические точки появляются по мере роста λ при $\lambda > 3$, отделяясь от ненулевой неподвижной точки $x_\lambda = (\lambda - 1)/\lambda$. Расстояние от x_λ равно

$$\begin{aligned} x_{\lambda,\pm} - x_\lambda &= \frac{\lambda + 1 \pm \sqrt{(\lambda + 1)(\lambda - 3)} - 2(\lambda - 1)}{2\lambda} = \\ &= \frac{3 - \lambda \pm \sqrt{(\lambda + 1)(\lambda - 3)}}{2\lambda} \approx \frac{3 - \lambda}{6} \pm \frac{\sqrt{4(\lambda - 3)}}{6} \approx \frac{\sqrt{\lambda - 3}}{3}. \end{aligned} \quad (11.2.2)$$

Поэтому на бифуркационной диаграмме возникают кривые, похожие на параболу (см. рис. 11.2.3.)

Это качественное изменение структуры орбит является бифуркацией, но она отличается от бифуркаций, рассмотренных в п. 2.3.2. В этом случае происходит переход от ситуации, когда имеются две неподвижные точки, а все остальные

точки асимптотически стремятся к одной из них, к ситуации, когда имеются две неподвижные точки и орбита периода 2.

Если $\lambda > 3$, то $|f'_\lambda(x_\lambda)| = \lambda|1 - 2x_\lambda| = |2 - \lambda| > 1$, так что ненулевая неподвижная точка становится отталкивающей в точности после значения параметра, при котором появляются периодические точки. Это явление называется *бифуркацией удвоения периода*. Мы можем интуитивно понять необходимость таких переходов, если заметим, что для $\lambda < 3$ все точки вблизи неподвижной точки притягиваются к ней (поочередно то с одной стороны, то с другой). Если λ немного больше 3, то близлежащие точки отталкиваются, в то время как удаленные все еще приближаются к неподвижной точке. Следовательно, с каждой стороны от неподвижной точки должна существовать точка, разделяющая эти два типа поведения. Эти две разделяющие точки образуют притягивающую периодическую орбиту.

Для $\lambda < 1 + \sqrt{6}$ периодическая орбита является притягивающей. Чтобы убедиться в этом, вычислим производную отображения f_λ в точках $x_{\lambda,\pm}$. Используя равенство $f'_\lambda(x) = \lambda(1 - 2x)$, где значения $x_{\lambda,\pm}$ определяются из (11.2.1), получаем

$$f'_\lambda(x_{\lambda,\pm}) = \lambda \left(1 - \frac{\lambda + 1 \pm \sqrt{(\lambda + 1)(\lambda - 3)}}{\lambda} \right) = -1 \pm \sqrt{(\lambda + 1)(\lambda - 3)},$$

и значение $(f_\lambda^2)'(x_{\lambda,\pm})$ равно произведению этих двух чисел: $(f_\lambda^2)'(x_{\lambda,\pm}) = 1 - (\lambda + 1)(\lambda - 3)$. Эта величина равна 1 при $\lambda = 3$. Она убывает до -1 при возрастании λ до положительного корня $\lambda = 1 + \sqrt{6}$ уравнения $(\lambda + 1)(\lambda - 3) = 2$. Поэтому орбита периода два является притягивающей для $3 < \lambda < 1 + \sqrt{6}$.

Для $\lambda = 1 + \sqrt{6}$ все орбиты, не являющиеся прообразами неподвижных точек, все еще сходятся к орбите периода два $x_{1+\sqrt{6},\pm} = \frac{\sqrt{2} + \sqrt{3} \pm 1}{\sqrt{2} + 2\sqrt{3}}$, но теперь с субэкспоненциальной скоростью; см. рис. 11.2.1 или рис. 11.2.2. $\square$

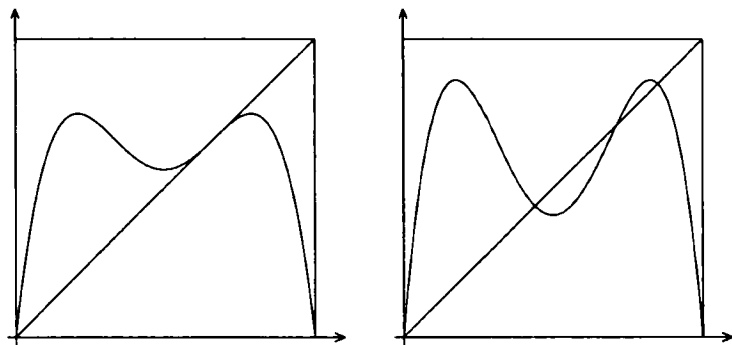
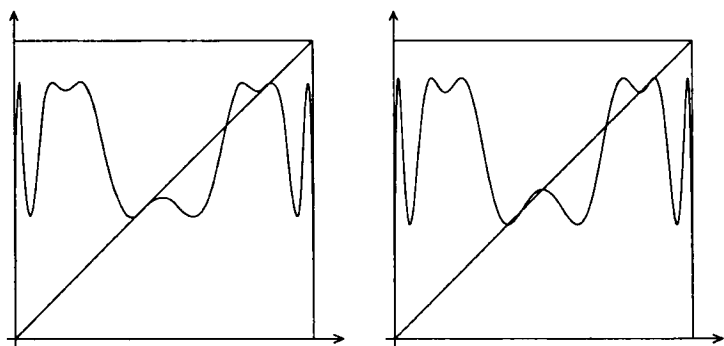


Рис. 11.2.1. Отображения f_3^2 и $f_{1+\sqrt{6}}^2$

Опишем структуру областей притяжения для $3 < \lambda \leq 1 + \sqrt{6}$. Неустойчивая неподвижная точка x_λ имеет один прообраз (если не считать саму эту точку)

$$x_{\lambda,1} := 1 - x_\lambda = 1/\lambda,$$

Рис. 11.2.2. Отображения $f_{1+\sqrt{6}}^4$ и $f_{1.05+\sqrt{6}}^4$

который в свою очередь имеет два прообраза

$$x_{\lambda,2}^{\pm} := 1/2 \pm \sqrt{\lambda^2 - 4}/2\lambda.$$

Проверка с использованием порядка точки показывает, что области непосредственного притяжения точек $x_{\lambda,\pm}$ имеют вид

$$(x_{\lambda}, x_{\lambda,2}^{+}) \quad \text{и} \quad I_1(\lambda) := (x_{\lambda,1}, x_{\lambda}) \quad (11.2.3)$$

соответственно.

Дальнейшие прообразы неустойчивой неподвижной точки x_{λ} образуют две последовательности, сходящиеся к концам отрезка 0 и 1. Эти прообразы разделяют область притяжения устойчивой орбиты периода два на счетное число непересекающихся интервалов. Для отображения f_{λ}^2 область притяжения каждой из двух устойчивых неподвижных точек состоит из счетного числа непересекающихся интервалов, расположенных поочередно то по одну, то по другую сторону от области непосредственного притяжения.

Основное наблюдение, позволяющее проводить дальнейший анализ, заключается в том, что ограничение отображения f_{λ}^2 при $3 < \lambda < 1 + \sqrt{6}$ на область непосредственного притяжения любой из двух его притягивающих неподвижных точек качественно выглядит точно так же, как и первоначальное отображение f_{λ} на всем отрезке $[0, 1]$ при $\lambda < 3$. Оно имеет единственную критическую точку, отталкивающую неподвижную точку, совпадающую с одним из концов отрезка, прообраз которой совпадает с другим концом, и притягивающую неподвижную точку с отрицательной производной.

По мере роста параметра притягивающая неподвижная точка становится отталкивающей при бифуркации удвоения периода. После бифуркации можно ограничить отображение $f_{\lambda}^4 = (f_{\lambda}^2)^2$ на отрезок $I_2(\lambda)$ между прообразом неустойчивой неподвижной точки и самой точкой (он соответствует отрезку $I_1(\lambda)$ для $f_{\lambda}^2|_{I_1(\lambda)}$). Эту процедуру можно продолжать далее.

Здесь мы впервые встречаемся с идеей *ренормализации*.

11.2.2. Каскад удвоений периода. Действительно, описанную процедуру можно применять к степеням отображения. Описанное выше качественное по-

добие позволяет определять отрезки $I_n(\lambda)$ индуктивно, с помощью отображений $f_\lambda^{2^n-1}$. При этом мы будем наблюдать бифуркацию удвоения периода устойчивой неподвижной точки данного отображения на таком отрезке. Это явление не тривиально, но, тем не менее, оно наблюдается не только для квадратичного семейства. Перенормировка возможна прежде всего по той причине, что производная Шварца

$$Sf := \frac{f'''}{f'} + \frac{3}{2} \left(\frac{f''}{f'} \right)^2 \quad (11.2.4)$$

отрицательна. Это свойство сохраняется при взятии композиции отображений. Оно предотвращает появление почти плоских частей вдали от критических точек. Аналитически его можно использовать, чтобы контролировать искажения, вызываемые отображением.

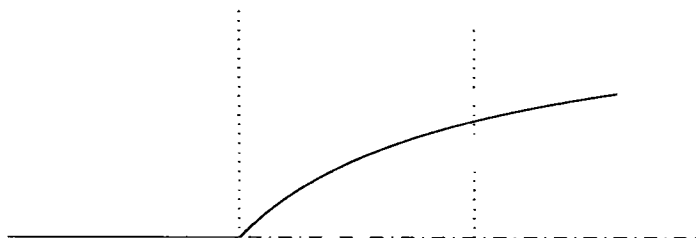
Прежде чем перейти к обсуждению этих общих вопросов, мы опишем результаты, полученные в данном частном случае. Следующая теорема дает качественную картину структуры орбит, возникающую при этом каскаде удвоений периода.

Теорема 11.2.2⁴. *Существует такая монотонная последовательность значений параметра $\lambda_1 = 3, \lambda_2 = 1 + \sqrt{6}, \lambda_3, \dots$, что для $\lambda_n < \lambda \leq \lambda_{n+1}$ квадратичное отображение f_λ имеет одну притягивающую периодическую орбиту $\mathcal{O}_n(\lambda)$ периода 2^n , две отталкивающие неподвижные точки 0 и x_λ и по одной отталкивающей периодической орбите $\mathcal{O}_k(\lambda)$ периода 2^k для каждого $k = 1, 2, \dots, n-1$. Область притяжения орбиты $\mathcal{O}(\lambda)$ является плотным множеством и состоит из всех точек, отличных от точек из этих периодических орбит и их прообразов. При $\lambda = \lambda_n$ орбита $\mathcal{O}_n(\lambda)$ подвергается бифуркации удвоения периода.*

Во время удвоения периода точки возникающей притягивающей орбиты периода 2^{n+1} появляются парами вокруг точек орбиты периода 2^n , так что порядок точек в различных периодических орбитах легко описывается: 11, 1212, 14241424, 1848284818482848 и т. д.

Любопытно, что критические значения параметра, с которыми мы до сих пор встречались, можно описать в общем виде: при $\lambda = 1 + \sqrt{0}$ появляется отличная от нуля неподвижная точка, при $\lambda = 1 + \sqrt{1}$ меняется «ориентация» (производная становится отрицательной), при $\lambda = 1 + \sqrt{4}$ появляется орбита периода два, и при $\lambda = 1 + \sqrt{6}$ появляется орбита периода четыре. При $\lambda = 1 + \sqrt{8}$ впервые появляется орбита периода три и, тем самым, орбиты всех периодов (см. предложение 11.3.8). Наконец, при $\lambda = 1 + \sqrt{9}$ мы приходим к максимальной сложности (см. п. 7.1.2). Кроме того, производная в устойчивой орбите периода два меняет знак, когда значение $(f_\lambda^2)'(x_{\lambda,\pm}) = 1 - (\lambda + 1)(\lambda - 3)$ равно нулю, т. е. $(\lambda + 1)(\lambda - 3) = 1$, или $\lambda = 1 + \sqrt{5}$. Эти специальные значения параметра $1 + \sqrt{n}$ для $n = 0, 1, 4, 5, 6, 8, 9$ отмечены вертикальными точечными линиями на рис. 11.2.3. Неустойчивые точки отмечены пунктиром. Позже могут появиться дополнительные точки данного периода. Интересный пример —

⁴ W. de Melo, S. van Strien. One-dimensional dynamics. Berlin: Springer-Verlag, 1993.

Рис. 11.2.3. Неподвижные и периодические точки как функции параметра λ

появление второй орбиты периода 4 для $\lambda = 1 + \sqrt{4 + \sqrt[3]{108}} = 1 + \sqrt{4 + 3\sqrt[3]{4}}$, отмеченной небольшой галочкой на оси x на рис. 11.2.3⁵. Для того чтобы проверить, что при $\lambda = 1 + \sqrt{8}$ появится орбита периода 3, мы используем следующее утверждение: если $g_\alpha(x) := \alpha - x^2$ и $h_\lambda = \lambda\left(x - \frac{1}{2}\right)$, тогда

$h_\lambda(f_\lambda(x)) = g_\alpha(h_\lambda(x))$, где $\alpha = \frac{\lambda^2}{4} - \frac{\lambda}{2}$ (см. упражнение 2.5.2). Для $\lambda = 1 + \sqrt{n}$ это означает, что $\alpha = (n-1)/4$, так что нам нужно доказать, что орбита периода 3 появляется для отображения $g_\alpha(x) = \alpha - x^2$ для $\alpha = 7/4$. Это действительно так, потому что многочлен $g_{7/4}^3(x) - x$ имеет пять корней, три из которых являются также корнями многочлена $(g_{7/4}^3)'(x) - 1$: проверьте, что $64 \cdot (g_{7/4}^3(x) - x) = (g_{7/4}(x) - x)(1 - 18x - 4x^2 + 8x^3)^2$ и $64 \cdot ((g_{7/4}^3)'(x) - 1) = -(8 - 3x - 22x^2 + 4x^3 + 8x^4)(1 - 18x - 4x^2 + 8x^3)$.

Само появление бесконечного числа удвоений периода характерно для гораздо более общей ситуации, чем для квадратичного семейства (см. п. 11.3.2). Дальнейшее развитие этого общего подхода основано на общем принципе, известном как *теория инвариантов перемешивания*⁶. Монотонность, т.е. отсутствие скачков с периода 2^k назад к периоду 2^{k-1} по мере увеличения параметра, является тонким результатом, при доказательстве которого в общем случае используется отрицательность производной Шварца (11.2.4). А также верно следующее утверждение: весь каскад удвоений периода появляется только один раз, и при каждом значении параметра происходит только одна бифуркация. Хотя эти особенности не ограничены квадратичным семейством, они зависят от более специальных свойств рассматриваемого семейства.

§ 11.3. Сложные динамические явления

11.3.1. Универсальность Фейгенбаума. Каскад удвоений периода был известен еще в начале 1960-х гг. На конференции в 1975 г. Стивен Смейл предположил, что накопление удвоений с ростом параметра подчинено некоторым закономерностям. Затем Митчелл Фейгенбаум на основе численных экспери-

⁵Этот результат получен Аном Нгуеном.

⁶Его описание и ссылки на литературу можно найти в книге: А. Каток, Б. Хасселблат. Введение в современную теорию динамических систем.

ментов пришел к выводу, что в квадратичном семействе имеется регулярная закономерность, с которой происходят последовательные бифуркации удвоения периода.⁷ Точнее говоря, расстояние между последовательными бифуркациями убывает приблизительно со скоростью геометрической прогрессии. Как следует из численных экспериментов и полученного позже доказательства, существует предел $\delta := \lim_{n \rightarrow \infty} \frac{\lambda_n - \lambda_{n-1}}{\lambda_{n+1} - \lambda_n}$. Фейгенбаум с большой точностью вычислил значение $\delta \approx 4,6992016090$.

С этого интересного наблюдения мы и начнем изучение роста сложности. Еще более интересно, что каскады удвоения периодов появляются для всех близких однопараметрических семейств унимодальных отображений и что последовательность параметров бифуркации представляет собой почти геометрическую прогрессию, как и в квадратичном случае, причем, что особенно удивительно, *с тем же самым предельным отношением*. Иными словами, определенное численное значение, полученное для квадратичного семейства, не связано непосредственно с конкретным видом этих отображений, а является внутренней особенностью, зависящей от общей формы отображения. Приведем краткое описание результатов Фейгенбаума, которые впоследствии были строго доказаны.⁸

Рассмотрим отображение f_{λ_∞} в конце каскада удвоений периода. Ограничение его квадрата на отрезок $[1 - x_{\lambda_\infty}, x_{\lambda_\infty}]$ имеет те же свойства, что и само отображение f_{λ_∞} : оно оставляет неподвижным один из концов отрезка (который является образом другого конца), является унимодальным и находится в конце каскада удвоений периода. Оказывается, отображение f и это ограничение топологически сопряжены (см. определение 7.3.3), т. е. они совпадают с точностью до непрерывной замены переменной. Фейгенбаум задался вопросом, связаны ли они более близко, а именно линейной заменой переменной. Хотя оказалось, что это не так, он все же смог получить полезные результаты. Для упрощения обозначений сделаем такую замену переменной, чтобы квадратичное семейство имело вид $f(x) = \alpha - x^2$ на симметричном отрезке (см. упражнение 2.5.2). Тогда имеется одна положительная неподвижная точка и 0 является критической точкой. Фейгенбаум показал, что существует единственное четное аналитическое отображение g , которое линейно связано с ограничением своего квадрата. Это означает, что имеется единственное такое вещественное число α (приблизительно равное $-2,5$), что $\alpha g^2(x/\alpha) = g(x)$. Можно получить отображение g численно, находя приближенные значения коэффициентов в его разложении в степенной ряд. Оказывается, $g(x) \approx 1 - 1,52763x^2 + 0,10481x^4 - 0,0267057x^6 + \dots$

Как получить это замечательное автомодельное отображение? Оно является единственной неподвижной точкой оператора ренормализации, определенного равенством

$$\mathcal{R}f(x) = \alpha f^2(x\alpha),$$

⁷ M. J. Feigenbaum. Quantitative universality for a class of nonlinear transformations // Journal of Statistical Physics. 1978. V. 19, no. 1. P. 25—52.

⁸ O. Lanford III. A computer-assisted proof of the Feigenbaum conjectures // Bull. Amer. Math. Soc. (N. S.). 1982. V. 6, no. 3. P. 427—434.

где x находится между положительной неподвижной точкой x_f отображения f и точкой $-x_f$. Мы предполагаем, что f — четная функция, $f(0) = 1$ и $\alpha = 1/f^2(0) = 1/f(1)$. Нас не будет беспокоить, что пространство таких функций бесконечномерно.

Изучая этот оператор, мы получим результат о скорости появления бифуркаций удвоения периода. Автомоделное отображение g является гиперболической неподвижной точкой оператора $\mathcal{R}$. Кроме того, дифференциал оператора $\mathcal{R}$ в точке g имеет одномерное собственное подпространство с собственным значением $\delta \approx 4,69920166$, и имеется дополнительное инвариантное подпространство, на котором он является сжимающим. Из существования гиперболической неподвижной точки следует (см. § 9.5), что для пространства, сжимающегося по действию дифференциала оператора $\mathcal{R}$, имеется поверхность, состоящая из функций, стремящихся к g при итерациях оператора $\mathcal{R}$, причем она является устойчивым многообразием. Это многообразие можно описать как множество отображений, для которых орбита критической точки упорядочена так же, как для отображения g . Поверхность S представляет собой предел поверхностей S_n , состоящих из функций, для которых критическая точка является 2^n -периодической. Заметим, что $\mathcal{R}(S_{n+1}) = S_n$. Из этого следует, что расстояние от S_n до S приблизительно равно $1/\delta^n$ для больших значений n , потому что δ — это скорость, с которой поверхности S_n «отодвигаются» от S при действии оператора $\mathcal{R}$ (см. рис. 11.3.1).

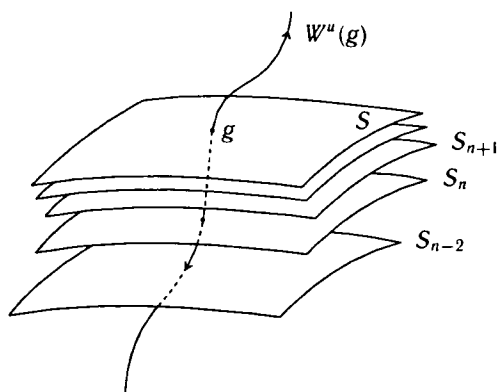


Рис. 11.3.1. Оператор ренормализации

Метод, описанный в § 9.5, позволяет получить стандартное однопараметрическое семейство функций, содержащее g : для любого однопараметрического семейства функций f_λ , которое пересекает S вблизи от g , определим $\hat{\mathcal{R}}(f_\lambda) := \mathcal{R}(f_{\lambda/\delta})$. Этот оператор устраняет растяжение, в δ раз в одномерном направлении, возникающее при действии оператора $\mathcal{R}$. Таким образом, для этого измененного оператора существует неподвижная точка, а значит, имеется $\mathcal{R}$ -инвариантное однопараметрическое семейство, содержащее функцию g .

Основная идея этого последнего наблюдения состоит в том, что с помощью такой ренормализации мы можем сравнить расстояния между последовательными бифуркациями удвоения периода в однопараметрическом семействе. По-

сколько мы только что описали предельное поведение как универсальное, любой каскад удвоений периода происходит с асимптотическим отношением $1/8$ между последовательными удвоениями.

11.3.2. Появление степеней числа 2. Проверка утверждений, сформулированных в предыдущем пункте, требует трудоемкого анализа. Однако, чтобы показать, что некоторые важные особенности являются универсальными, мы будем использовать только теорему о промежуточном значении. Покажем, что порядок появления степеней числа 2 в квадратичном семействе является единственно возможным для любого семейства.

Определение 11.3.1. Рассмотрим непрерывное отображение $f: I \rightarrow I$ отрезка I . Мы говорим, что отрезок $J \subset I$ *накрывает* (или f -накрывает) отрезок $K \subset I$ (под действием f), если $K \subset f(J)$. Обозначим это отношение $J \rightarrow K$. Рассмотрим конечную систему отрезков, принадлежащих отрезку I . Граф, вершины которого — это в точности отрезки из данной системы, а стрелки определены отношением f -накрытия, называется ассоциированным *марковским графом*.

Лемма 11.3.2. Если J, K — отрезки и $J \rightarrow K$, то существует такой отрезок $L \subset J$, что $f(L) = K$.

Доказательство. Пусть $K = [a, b]$ и $c := \max f^{-1}(\{a\})$. Возьмем $L = [c, d]$, где $d := \min((c, \infty) \cap f^{-1}(\{b\}))$, если этот минимум определен. В противном случае выберем $L = [c', d']$, где $c' := \max((-\infty, c) \cap f^{-1}(\{b\}))$ и $d' := \min((c', \infty) \cap f^{-1}(\{a\}))$. $\square$

Таким образом, если $J \rightarrow K$, то имеются такие отрезки $L_1, \dots, L_k \subset J$ с попарно непересекающимися внутренностями, что $f(L_i) = K$. Иногда мы будем писать $J \rightrightarrows K$ с k стрелками, где k — максимальное число таких отрезков L_i . Эти отрезки L_i называются *полными компонентами*, связанными с отношением $J \rightarrow K$. Заметим, что прообраз отрезка K в J может содержать бесконечно много отрезков, несмотря на то что имеется только конечное число полных компонент в силу компактности.

Следующие две леммы описывают связь между этим отношением накрытия и периодическими точками.

Лемма 11.3.3. Если $J \rightarrow J$, то отображение f имеет неподвижную точку $x \in J$.

Доказательство. Если $J = [a, b]$, то, поскольку $J \rightarrow J$, имеются такие точки $c, d \in J$, что $f(c) = a \leq c$ и $f(d) = b \geq d$. По теореме о промежуточном значении функция $f(x) - x$ обращается в нуль на отрезке J . $\square$

Лемма 11.3.4. Если $I_0 \rightarrow I_1 \rightarrow I_2 \rightarrow \dots \rightarrow I_n$, то множество $\bigcap_{i=0}^n f^{-i}(I_i)$ содержит такой отрезок Δ_n , что $f^n(\Delta_n) = I_n$.

Доказательство. Положим $\Delta_0 = I_{n-1}$ и определим по индукции полную компоненту $\Delta_i \subset I_{n-i}$, связанную с отношением $I_{n-i} \rightarrow \Delta_{i-1}$. $\square$

На основании леммы 11.3.3 мы получаем отсюда такое следствие.

Следствие 11.3.5. *Если $l_0 \rightarrow l_1 \rightarrow l_2 \rightarrow \dots \rightarrow l_{n-1} \rightarrow l_0$, то существует такая точка $x \in \text{Fix } f^n$, что $f^i(x) \in l_i$ для $0 \leq i < n$.*

Лемма 11.3.6 (Бартон—Бернс). *Если отображение f имеет периодическую точку, не являющуюся неподвижной, то оно имеет такую точку периода 2.*

Доказательство. Докажем, что если точка $x \neq f^2(x)$ имеет период p , то для отображения f существует точка периода меньше чем p , которая не является неподвижной.

Рассмотрим отрезки $l_1, \dots, l_{p-1}$, концы которого являются последовательными итерациями точки x . Для каждого $i \in \{1, \dots, p-1\}$ найдется такое $j \neq i$, что $l_i \rightarrow l_j$, так как концы отрезка не могут быть 2-периодическими точками. Таким образом, имеется нетривиальный цикл в марковском графе, который проходит через $k \leq p-1$ отрезков. Согласно следствию 11.3.5 существует k -периодическая точка, которая не является неподвижной. $\square$

Отсюда следует, что степени числа 2 появляются в возрастающем порядке.

Теорема 11.3.7. *Если $f: I \rightarrow I$ — непрерывное отображение отрезка с периодической точкой периода 2^n , то f имеет периодические точки всех периодов 2^m для $m \leq n$.*

Доказательство. Если $m = 0$, применим лемму 11.3.3, в противном случае применим лемму 11.3.6 к отображениям $f, f^2, \dots, f^{2^{n-2}}$. $\square$

11.3.3. Другие соотношения, из которых следует периодичность. Есть еще один красивый результат, имеющий также и исторический интерес, потому что в статье, в которой он был опубликован, впервые появился термин «хаос»⁹.

Предложение 11.3.8. *Если $f: I \rightarrow I$ — непрерывное отображение отрезка с периодической точкой периода 3, то f имеет периодические точки всех периодов.*

Доказательство. Обозначим точки из орбиты периода 3 через $\{x_1 < x_2 < x_3\}$.

Рассмотрим отрезки $l_1 = [x_1, x_2]$ и $l_2 = [x_2, x_3]$. Предположим, что $f(x_2) = x_3$. Тогда $f^2(x_2) = x_1$ и, следовательно, отрезок l_2 f -накрывает отрезки l_1 и l_2 , а l_1 накрывает l_2 . Если $f(x_2) = x_1$, то, переобозначив отрезки l_1 и l_2 , мы придем к тому же выводу. Таким образом, марковский граф, ассоциированный с l_1, l_2 , содержит

⁹Т.-Y. Li, J. A. Yorke. Period Three Implies Chaos // American Mathematical Monthly. 1975. V. 82, no. 10. P. 985—992. Понятие хаоса использовалось и раньше, и даже в более строгой форме, но это не обязательно был детерминистский хаос; см.: N. Wiener. The Homogeneous Chaos // American Journal of Mathematics. 1938. V. 60, no. 4. P. 897—936, а также N. Wiener, A. Wintner. The Discrete Chaos // American Journal of Mathematics. 1943. V. 65, no. 2. P. 279—298. Специалистов по естественным наукам скорее могли познакомиться с этим термином через работы Роберта Мэя: R. M. May. Biological Populations with Non-Overlapping Generations: Stable Point, Stable Cycles and Chaos // Science. 1974. V. 186. P. 645—647 или R. M. May. Simple Mathematical Models with Very Complicated Dynamics // Nature. 1976. V. 261. P. 459—467, в которых автор выражает благодарность Йорку за то, что он ввел этот термин.

граф



(11.3.1)

и для любого $n \in \mathbb{N}$ мы имеем цикл $I_1 \rightarrow I_2 \rightarrow I_2 \rightarrow \dots \rightarrow I_2 \rightarrow I_1$ (с $n - 1$ появлениями отрезка I_2), а значит, в силу следствия 11.3.5 существует периодическая точка периода n , которая, очевидно, не может иметь какой-либо меньший период. $\square$

Этот результат является частным случаем более общего утверждения, которое было доказано ранее.

Теорема 11.3.9 (Шарковский)¹⁰. Если $I \subset \mathbb{R}$ — отрезок, $f: I \rightarrow I$ — непрерывное отображение с периодической точкой минимального периода p и $q < p$, где отношение « $\triangleleft$ » определено следующим образом:

$$1 \triangleleft 2 \triangleleft 2^2 \triangleleft 2^3 \triangleleft \dots \triangleleft 2^m \triangleleft \dots \triangleleft 2^k (2n - 1) \triangleleft \dots \\ \dots \triangleleft 2^k \cdot 3 \triangleleft \dots \triangleleft 2 \cdot 3 \triangleleft \dots \triangleleft 2n - 1 \triangleleft \dots \triangleleft 9 \triangleleft 7 \triangleleft 5 \triangleleft 3,$$

то отображение f имеет периодическую точку минимального периода q .

Существо доказательства теоремы Шарковского проявляется в двух частных случаях, разобранных выше (теорема 11.3.7 и предложение 11.3.8). Можно развивать эту тему и далее: хотя из наличия любого периода, отличного от 3, вовсе не следует, что имеются все другие периоды, можно, например, показать, что если отображение $f: [0, 1] \rightarrow [0, 1]$ имеет такую периодическую орбиту $\{x_1 < x_2 < x_3 < x_4\}$, что $f(x_i) = x_{i+1}$ для $i < 4$ и $f(x_4) = x_1$, то f имеет периодические точки всех периодов. Это приводит к вопросу о том, для каких конфигураций периодических орбит существуют другие периодические орбиты.

11.3.4. Аттрактор Фейгенбаума—Мизюревича. Отображение f_{λ_∞} в конце каскада удвоения периодов должно иметь большой набор свойств автомодельности. При обсуждении универсальности Фейгенбаума мы видели, что оно в некотором смысле инвариантно относительно ренормализации. Теперь установим в явном виде некоторые важные свойства автомодельности (самоподобия) динамики.

Мы будем использовать два подхода. Сначала кратко опишем расположение периодических точек отображения f_{λ_∞} относительно друг друга. Затем опишем динамику на важном инвариантном множестве, которое не содержит периодических точек.

1. Комбинаторика периодических точек. Комбинаторная модель периодических точек была уже описана после теоремы 11.2.2: 11, 1212, 14241424,

¹⁰ А. Н. Шарковский. Сосуществование циклов непрерывного отображения прямой в себя // Украинский математический журнал. 1964. Т. 16, № 1. С. 61—71; Циклы и структура непрерывного отображения // Украинский математический журнал. 1965. Т. 17, № 3. С. 104—111.

1848284818482848 и т. д. Кроме того, можно определить динамическое упорядочение каждой из этих орбит. Например, орбита периода 4 имеет левую и правую половины, и эти две половины меняются местами под действием отображения f_{λ_∞} , потому что они связаны с двумя точками орбиты периода 2. Аналогично левые и правые половины орбиты любой точки периода 2^n меняются местами под действием отображения f_{λ_∞} . Свойства автомодельности начинают проявляться, например, когда левая половина орбиты периода 8 отображается в себя под действием отображения $f_{\lambda_\infty}^2$ или отображается в правую половину под действием отображения f_{λ_∞} . Поскольку два набора, соответствующие двум левым точкам орбиты периода 4, связаны с этими точками, эти два набора для орбиты периода 8 отображаются по-разному. Точнее говоря, если мы обозначим точки через $x_1, \dots, x_8$ в соответствии с их порядком на отрезке, то $f_{\lambda_\infty}(\{x_1, x_2\}) = \{x_i, x_{i+1}\}$, где $i = 5$ или $i = 7$, в зависимости от орбиты периода 4.

Поэтому орбиты периодов 2^n для различных n переплетены друг с другом таким образом, что при отображении разбиваются на два набора, причем за движением по таким орбитам можно следить рекурсивно, используя комбинаторику орбиты половинного периода. Это один из путей возникновения автомодельности.

2. Аттрактор Фейгенбаума—Мизюревича. Автомодельная динамика для отображения f_{λ_∞} была введена абстрактно в упражнении 4.1.15 и может быть описана в более знакомых терминах следующим образом.

Определение 11.3.10. Пусть Ω_2^R — пространство односторонних 0-1-последовательностей, определенных в п. 7.3.4. Отображение $A: \Omega_2^R \rightarrow \Omega_2^R$, задаваемое равенством

$$(A\omega)_i = \begin{cases} 1 - \omega_i, & \text{если } \omega_j = 1 \text{ для всех } j < i, \\ \omega_i & \text{в противном случае,} \end{cases}$$

называется **двоичным одометром**.

Теорема 11.3.11. *Отображение f_{λ_∞} имеет замкнутое инвариантное множество, которое состоит из изолированных отталкивающих периодических орбит с периодами 2^n для $n = 0, 1, 2, \dots$ (по две орбиты для $n = 0$ и по одной для других периодов) и канторова множества S , динамика на котором топологически сопряжена с двоичным одометром.*

Множество S — это в точности множество предельных точек для множества периодических точек. Кроме того, S — это множеством ω -предельных точек (см. определение 4.3.18) любой точки, которая не является прообразом периодической точки.

Набросок доказательства. Приведенный выше комбинаторный анализ периодических точек можно эквивалентным образом описать как анализ отрезков, определенных следующим условием: орбита периода 2^{n+1} , не содержащая точек периода 2^n , f_{λ_∞} -покрывает орбиту, содержащую точку периода 2^{n-1} . Мы используем такое описание для построения множества S .

Сначала рассмотрим отрезок I , концы которого являются точками периода 2. Обозначим через S^0 полную компоненту для $I \rightarrow I$. Имеются два отрезка, концы которых лежат в орбите периода 4, содержащие точку периода 2. Обозначим левый из них I_1 , а правый I_2 . Пусть S^1 — объединение одной полной компоненты для $I_1 \rightarrow I_2 \rightarrow I_1$ и одной полной компоненты для $I_2 \rightarrow I_1 \rightarrow I_2$. Заметим, что расстояние от множества S^1 до неподвижной точки положительно. Множество S^2 получается аналогично как объединение полных компонент, соответствующих циклам длины 4, ассоциированным с четырьмя отрезками, которые получены подобным образом из орбит периода 8. Таким образом, множество S^2 отделено от точек периода один и два. Аналогично мы получаем множества S^n , состоящие из 2^{n-1} отрезков, отделенных от точек периода до 2^{n-1} включительно. Затем обозначим через S границу множества $\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} S^n$. Это означает, что S — множество непериодических точек в замыкании множества периодических точек. Обозначим через y_n правый конец крайнего левого отрезка из множества S^n и положим $S^n = \{x \in S: x \leq y_n\}$.

Определим отображение $h: S \rightarrow \Omega_2^R$ равенствами $h(S^n) = \{\omega \in \Omega_2^R: \omega_1 = \dots = \omega_n = 1\}$ и $h \circ f_{\lambda_\infty} = A \circ h$. Тогда отображение h непрерывно и сюръективно, т. е. одометр A является фактором ограничения $f_{\lambda_\infty}|_S$. В этом смысле динамику одометра A на Ω_2^R можно получить из динамики отображения f_{λ_∞} на S . $\square$

В свете теоремы 11.3.11 естественно обратить внимание на внутреннюю динамику двоичного одометра.

Предложение 11.3.12. *Двоичный одометр строго эргодичен.*

Доказательство. Фазовое пространство двоичного одометра A имеет структуру вложенных множеств. На n -м шаге имеется 2^n непересекающихся канторовых подмножеств, называемых *цилиндрами ранга n* , которые одновременно замкнуты и открыты и которые циклически переставляются при отображении. Любая функция, постоянная на каждом цилиндре некоторого ранга, непрерывна, а каждую непрерывную функцию можно представить как равномерный предел функций, постоянных на цилиндрах некоторых рангов. Для функции ϕ , которая является постоянной на цилиндрах ранга n , справедливо равенство

$$\text{const} = \frac{1}{2^n} \sum_{i=0}^{2^n-1} \phi \circ A^i = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^N \phi \circ A^i.$$

Равномерно приближая произвольную непрерывную функцию функциями, постоянными на цилиндрах, мы получим равномерную сходимость средних. $\square$

Таким образом, одометр имеет некоторые общие черты с иррациональными поворотами окружности. В некотором смысле он даже проще (малые участки возвращаются точно, а не приблизительно, как малые интервалы на окружности), но строгая эргодичность здесь менее «совершенна», потому что четные итерации не строго эргодичны.

Следующий результат показывает, что в некотором смысле двоичный одометр — единственная модель с относительно простым непериодическим рекуррентным поведением, которое возникает для отображений отрезка ¹¹.

Теорема 11.3.13. *Предположим, что f — непрерывное отображение отрезка в себя с нулевой топологической энтропией и S — замкнутое f -инвариантное множество без периодических точек, но с плотной орбитой (т. е. топологически транзитивное, см. определение 4.1.3). Тогда ограничение отображения f на S топологически сопряжено с двоичным одометром.*

§ 11.4. Гиперболическое и стохастическое поведение

Теперь мы будем изучать динамику квадратичных отображений не с точки зрения комбинаторного анализа орбит, а с точки зрения их устойчивости и неустойчивости.

11.4.1. Гиперболическое отталкивающее канторово множество. Предложение 7.4.4 справедливо для всех $\lambda > 4$; мы предположили, что $\lambda > 2 + \sqrt{5}$, лишь для того, чтобы упростить доказательство в нашей ситуации. Для других значений доказательство более сложно и основано на отрицательности производной Шварца (11.2.4). Таким образом, справедливо следующее утверждение.

Предложение 11.4.1. *Существует такой гомеоморфизм $h: \Omega_2^{\mathbb{R}} \rightarrow \Lambda := \bigcap_{n \in \mathbb{N}_0} f^{-n}([0, 1])$, что $h \circ \sigma^{\mathbb{R}} = f \circ h$, где $f: \mathbb{R} \rightarrow \mathbb{R}$, $x \rightarrow \lambda x(1 - x)$, $\lambda > 4$.*

На самом деле из доказательства предложения 7.4.4 следует, что множество Λ является гиперболическим отталкивающим множеством (см. определение 10.1.2).

Эту полезную модель удобно использовать при изучении возможного поведения квадратичных отображений. Для значений параметра $\lambda < \lambda_{\infty}$ асимптотика очевидна: орбиты притягиваются к орбите периода 2^n с самым большим возможным значением n . В предыдущем параграфе динамика отображения $f_{\lambda_{\infty}}$ была описана в терминах периодических точек (которые все являются отталкивающими) и автомодельного инвариантного множества S .

Остается изучить значения λ между λ_{∞} и 4. Это требует сложного анализа, выходящего за рамки этой книги, однако мы можем здесь описать результаты этого анализа. Для начала опишем один из двух главных типов поведения.

Первым шагом является общий анализ роста сложности.

Теорема 11.4.2 ¹². *Топологическая энтропия отображения f_{λ} не убывает, равна нулю для $\lambda \leq \lambda_{\infty}$ и положительна для $\lambda > \lambda_{\infty}$.*

Отсюда следует такой результат ¹³.

¹¹ Доказательство содержится в книге: А. Каток, Б. Хасселблат. Введение в современную теорию динамических систем (см. доказательство теоремы 15.4.2).

¹² W. de Melo, S. van Strien. One-dimensional dynamics. Berlin: Springer-Verlag, 1993.

¹³ Из положительности топологической энтропии следует существование (одномерной) подковы для итерации, откуда мы и получаем утверждение следствия; см. следствие 15.2.4 в книге: А. Каток, Б. Хасселблат. Указ. соч.

Следствие 11.4.3. *Отображение f_λ для $\lambda > \lambda_\infty$ имеет бесконечно много периодических орбит с периодами, отличными от степеней числа 2.*

11.4.2. Притягивающая периодическая орбита и отталкивающее марковское множество. Напомним (см. п. 11.1.2), что для гиперболического квадратичного отображения возвращение происходит только на периодических орбитах и инвариантном канторовом множестве, которое может быть пусто и является множеством меры 0 (см. определение 7.5.3). Имеется только одна притягивающая периодическая орбита, и для $\lambda < 4$ любое гиперболическое отображение имеет одну такую орбиту. Всякая орбита, находящаяся вне канторова множества, стремится к этой орбите. Канторово множество является гиперболическим отталкивающим множеством (см. определение 10.1.2) и непусто для $\lambda > \lambda_\infty$.

В этой гиперболической ситуации непосредственное численное исследование показало бы только наличие притягивающей периодической орбиты. Хотя динамика на канторовом множестве сложна, мы не смогли бы обнаружить это с помощью компьютера, потому что канторово множество имеет меру нуль. Даже если начальное значение, выбранное для численного эксперимента, принадлежало канторову множеству, за счет погрешности округления вычисленная орбита выйдет из этого множества и будет стремиться к периодической орбите.

Теорема 11.4.4 (Грачик—Швентек)¹⁴. *Множество значений $\lambda \in [0, 4]$, для которых квадратичное отображение f_λ является гиперболическим, открыто и плотно.*

Трудной частью этой теоремы является доказательство плотности, так как утверждение об открытости следует почти непосредственно из определения гиперболичности. Это утверждение иллюстрирует необходимость строгого математического подхода при доказательстве и показывает, что нельзя ограничиться только численным экспериментом. Действительно, данное множество значений параметра не очень заметно на бифуркационной диаграмме (см. рис. 11.2.3), на которой видны лишь несколько «окон» (светлые вертикальные полосы). Это происходит по двум причинам. За исключением самых малых периодов, эти «окна» очень узкие. К тому же, помимо того что орбиты достаточно больших периодов содержат достаточное количество точек для перехода в затемненную область, компьютеру требуется очень много итераций для приближения к этой орбите.

Тем не менее, бифуркационная диаграмма наводит на мысль, что имеется другое примечательное множество параметров.

11.4.3. Стохастическое поведение. Существует сложное поведение, сильно отличающееся от рассмотренного выше. Для него характерно, что сложная динамика сосуществует с равномерным распределением на отрезке. Несмотря на необычайную общую сложность картины поведения, многие явления возникают в довольно исключительных ситуациях. Все же имеется один тип неравномерно гиперболического поведения, который особенно важен, потому что он появля-

¹⁴J. Graczyk, G. Świątek. Generic Hyperbolicity in the Logistic Family // Annals of Mathematics (2). 1997. V. 146, no. 1. P. 1—52.

ется на множестве значений параметра ненулевой меры, а также в силу своей внутренней структуры.

1. *Тентообразное отображение.* Чтобы определить этот тип поведения, рассмотрим сначала модельную ситуацию, в которой $g: [0, 1] \rightarrow [0, 1]$ — это тентообразное отображение

$$g(x) = \begin{cases} 2x, & 0 \leq x \leq 1/2, \\ 2 - 2x, & 1/2 \leq x \leq 1. \end{cases}$$

Ясно (см. п. 7.1.1), что это отображение сохраняет меру в следующем смысле: для любого отрезка $I \subset [0, 1]$ мера множества $g^{-1}(I)$ совпадает с мерой самого отрезка I . (В данном случае мера объединения непересекающихся отрезков определяется как сумма их длин, как в определении 4.2.2.) За исключением того обстоятельства, что на правой половине отрезка ориентация изменяется, тентообразное отображение выглядит в точности как растягивающее отображение. Действительно, можно повторить рассуждения, доказывающие равномерное распределение орбит для удваивающего отображения (см. теорему 7.5.6), и убедиться в существовании равномерного распределения орбит для тентообразного отображения. Это приводит к значительной динамической сложности, но, в отличие от гиперболического поведения, которое обсуждалось выше, эта сложность равномерно распределена по всему отрезку.

2. *Отображение Чебышёва—фон Неймана—Улама.* Предыдущее наблюдение применимо к квадратичному семейству, потому что тентообразное отображение g топологически сопряжено с квадратичным отображением $f_4: x \mapsto 4x(1-x)$ посредством функции $h(x) = \sin^2(\pi x/2)$. Отображение f_4 представляет собой *второй полином Чебышёва*. П. Л. Чебышёв знал об этой сопряженности, но вытекающие из нее динамические результаты были рассмотрены только Джоном фон Нейманом и Станиславом Уламом. По этой причине f_4 иногда называется *отображением фон Неймана—Улама*.

Отображение f_4 описывается простой алгебраической формулой (без использования абсолютных значений, дробных частей и других математических фокусов), и это отображение демонстрирует настоящее «стохастическое» поведение. Проверка сопряженности этого отображения с тентообразным отображением сводится к доказательству тригонометрического тождества. Это не значит, что f_4 сохраняет меру в том же самом смысле, но, поскольку искажение при замене координат (отображении h) невелико, имеется такая положительная функция «плотности» $\rho: [0, 1] \rightarrow \mathbb{R}$, определяемая в данном случае равенством $\rho(x) = 1/\sqrt{x(1-x)}$, что f_4 сохраняет меру в смысле определения 4.2.2, если взять «длины» интервалов равными $\ell_\rho(I) := \int_I \rho(x) dx$. Это первый нетривиальный пример *равномерного распределения с плотностью* или *стохастичности*. Отсюда следует, что если мы знаем, в какой части отрезка мы находимся, то мы получаем (посредством тентообразного отображения, которое совпадает с E_2) такую же статистическую сложность, как в опытах с подбрасыванием монет (см. п. 7.5.1). С одной стороны, отсюда следуют различные формы закона

больших чисел. С другой стороны, это означает, что, как описано в п. 7.5.1, существуют точки, не имеющие долгосрочных средних¹⁵.

Для целей классификации квадратичных отображений мы будем считать плотность неотрицательной и не будем требовать, чтобы множество нулей имело меру 0.

3. *Механизмы стохастичности.* Наличие критической точки $1/2$, в которой производная квадратичного отображения равна нулю, является очевидной преградой для локального растяжения, необходимого для существования гиперболических отталкивающих множеств и соответствующего стохастического поведения. В гиперболических случаях эта точка просто поглощается областью притяжения и, таким образом, безопасна для отталкивающего множества, которое отделено от нее и от ее прообразов. Для отображения Чебышёва—фон Неймана—Улама критическая точка отображается в нуль, который является растягивающей неподвижной точкой. Таким образом, точки, близкие к критической точке (даже рекуррентные), имеют достаточно времени, чтобы восстановить локальное растяжение перед возвращением в область с малыми производными. Обобщая и усвершенствуя этот механизм, мы можем получать все большие множества параметров со стохастическим поведением.

4. *Отображения Рюэлля.* Наиболее прямое обобщение ситуации Чебышёва—фон Неймана—Улама относится к случаю, когда *итерация критической точки — отталкивающая периодическая точка*. Множество таких параметров счетно, и инвариантная плотность — кусочно гладкая функция.

5. *Отображения Мизюревича.* Предыдущие рассуждения были основаны на том, что критическая точка не рекуррентна. С учетом этого наблюдения Мизюревич впервые построил несчетное (хотя все еще нулевой меры) множество значений параметра, при которых возникает стохастическое поведение. Условие Мизюревича состоит в том, что *критическая точка $1/2$ не рекуррентна*, т. е. ее ω -предельное множество не содержит саму эту точку. В нашем случае ω -предельное множество критической точки является гиперболическим отталкивающим множеством и все еще имеется инвариантная плотность, гладкая на счетном множестве отрезков, но, возможно, имеющая разрывы на канторовом множестве. Доказательство стохастичности в ситуации Мизюревича очень напоминает случай Чебышёва—фон Неймана—Улама: любая рекуррентная точка, подходящая близко к критической, имеет достаточно времени для восстановления растяжения во время блуждания около отталкивающего ω -предельного множества.

6. *Отображения Якобсона—Колле—Экмана.* Для получения множества параметров со стохастическим поведением ненулевой меры необходимо рассмотреть ситуации, когда критическая точка возвращается произвольно близко к самой себе. Модель стохастичности в такой ситуации была описана Колле

¹⁵ Это также отметил Эдвард Лоренц, который использовал квадратичное отображение как модель для описания погодных явлений и заключил, что «климат» формируется не автоматически, где под климатом подразумевается долгосрочное среднее данных о погоде.

и Экманом, а Якобсон в своей замечательной работе 1980 г.¹⁶ показал, что стохастическое поведение появляется на множестве ненулевой меры и является преобладающим среди значений параметра, близких к 4 и к значениям Мизюревича и Рюэлля.

Механизм Якобсона—Колле—Экмана позволяет критической точке возвращаться близко к самой себе, но достаточно редко, так, чтобы *между* возвращениями точки, близкие к возвращению (но не настолько, чтобы их производная на предыдущем маршруте была близка к нулю), восстанавливают достаточное растяжение. Сложная индуктивная процедура показывает, что для достаточного количества точек растяжение во всех масштабах времени преобладает над случайными сжатиями, вызванными возвращениями.

Наиболее важное наблюдение, сделанное Якобсоном, — *метод исключения параметра*, позволяющий контролировать значения параметра, для которых слишком близкое случайное возвращение критической точки приводит к возникновению притягивающих периодических орбит и, таким образом, навсегда запрещает сжатие. В основе этого метода лежит оценка, показывающая, что, если это происходит после очень многих итераций, скорость перемещения соответствующей итерации критической точки настолько высока, что «опасная зона», в которой может появиться притягивающая периодическая точка при данном возвращении, проходит очень быстро.

В случае Якобсона—Колле—Экмана инвариантная плотность в пределе имеет большое количество разрывов, но главные качественные особенности, которые мы наблюдали для тентообразного отображения, по-прежнему преобладают.

¹⁶М. V. Jakobson. Absolutely continuous invariant measures for one-parameter families of one-dimensional maps // Communications in Mathematical Physics. 1981. V. 81, no. 1. P. 39—88.

ГОМОКЛИНИЧЕСКАЯ КАРТИНА

В первой части книги мы изучали сложную динамику главным образом на конкретных примерах. Есть две важные причины для такого подхода. Во-первых, оказывается, это позволяет легко и естественно ввести важные понятия. Во-вторых, диапазон явлений, приводящих к хаотическому поведению, почти полностью исчерпывается явлениями, возникающими в наших примерах. Теперь мы вернемся к изучению подковы и объясним, почему этот, казалось бы, весьма частный пример является важным механизмом, вызывающим хаотическую динамику для некоторых орбит. А именно, мы покажем, что подковы довольно часто встречаются в реальных системах, и объясним, как они используются при решении фундаментальных вопросов динамики. Сначала мы опишем общий механизм возникновения подков, а затем расскажем о различных путях их появления в конкретных задачах.

§ 12.1. Нелинейные подковы

При обсуждении подковы в п. 7.4.4 было удобно ввести предположение о линейности, но, как мы заметили в п. 10.2.6, делать это совсем не обязательно. Теперь мы рассмотрим нелинейные подковы.

Хотя подковы легко определить и полезно использовать в произвольной размерности, по ряду причин мы ограничимся рассмотрением лишь плоской (двумерной) ситуации. В плоском случае легче получить наглядную геометрическую иллюстрацию рассуждений, и поэтому он играет главную роль в развитии теории динамических систем. Наконец, именно в двумерном случае топологическая энтропия динамической системы полностью определяется по присутствующим в ней подковам.

В определении мы используем координатные проекции $\pi_1: (x, y) \mapsto x$ и $\pi_2: (x, y) \mapsto y$, определенные на $\mathbb{R}^2$.

Определение 12.1.1. Пусть множество $U \subset \mathbb{R}^2$ открыто. Тогда прямоугольник $\Delta = D_1 \times D_2 \subset U \subset \mathbb{R}^2$ (где D_1 и D_2 — отрезки) называется *подковой* для диффеоморфизма $f: U \rightarrow \mathbb{R}^2$, если множество $\Delta \cap f(\Delta)$ содержит по крайней мере две такие связные компоненты Δ_0 и Δ_1 , что (в обозначениях $\Delta_i = f(\Delta'_i)$, $i = 1, 2$, и $\Delta' = \Delta_0 \cup \Delta_1$) выполнены следующие свойства:

- 1) $\pi_2(\Delta'_i) = D_2$,
- 2) ограничение проекции $\pi_1|_{f(\Delta'_i \cap (D_1 \times \pi_2(z)))}$ является взаимно однозначным отображением на D_1 для любого $z \in \Delta'_i$,
- 3) $\pi_2(\Delta') \subset \text{int} D_2$, $\pi_1(f^{-1}(\Delta')) \subset \text{int} D_1$,

4) отображение $D(f|_{f^{-1}(\Delta')})$ сохраняет и растягивает горизонтальное семейство конусов на $f^{-1}(\Delta')$,

5) отображение $D(f^{-1}|_{\Delta'})$ сохраняет и растягивает вертикальное семейство конусов на Δ' (см. определение 10.1.4).

На рис. 12.1.1 компоненты Δ_0 и Δ_1 изображены двумя горизонтальными прямоугольниками серого цвета, которые составляют множество $\Delta \cap f(\Delta)$. Их прообразы — это вертикальные прямоугольники, которые проходят через Δ сверху вниз. Это и есть утверждение первого пункта. Второе требование состоит в том, что эти прямоугольники должны сами проходить через Δ горизонтально (при этом не слишком изгибаясь). Третье требование заключается в том, что эти полосы остаются на некотором расстоянии от верхнего и нижнего участков границы, а их прообразы находятся далеко от сторон. Другими словами, в любом направлении имеется свободное пространство. Это свойство позволяет получать устойчивость при возмущениях. Последние два условия — горизонтальное растяжение и вертикальное сжатие, а также требование, чтобы почти горизонтальные линии оставались почти горизонтальными, а почти вертикальные линии — почти вертикальными.

Требования, вводимые в определении 12.1.1, не слишком жесткие и весьма логичные: подкова выглядит примерно так, как показано на рис. 12.1.1, но может быть несколько искаженной и иметь переменное растяжение и сжатие. В § 12.3 мы увидим, как естественно возникают такие подковы. В силу структурной устойчивости (см. п. 10.2.6) динамика на линейной подкове и на ее незначительном возмущении идентичны с точностью до гомеоморфной замены координат, при которой не происходит сильного смещения точек. Из этого, в частности, следует то, что наши результаты о структуре орбит линейных подков также справедливы и для их возмущений (кодирование, см. предложение 7.4.6, рост периодических орбит, плотность и транзитивность, см. следствие 7.4.7).

§ 12.2. Гомоклинические точки

Теперь мы опишем «сценарий» возникновения подков и рассмотрим некоторые из его отличительных особенностей, которые объясняют, почему подковы встречаются так часто.

Для начала рассмотрим линейное гиперболическое отображение $A: \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $A(x, y) = (2x, y/2)$ (см. рис. 12.2.1). Ось y состоит из точек, стремящихся к началу

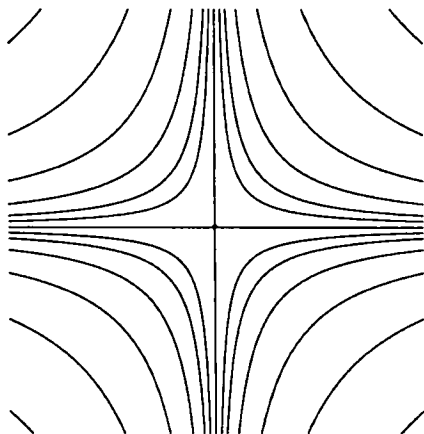


Рис. 12.2.1. Гиперболическая неподвижная точка

координат при положительных итерациях, а ось x состоит из точек, стремящихся к началу координат при отрицательных итерациях, в то время как все другие точки двигаются по гиперболам $xy = \text{const}$.

Теперь рассмотрим нелинейный диффеоморфизм f , который близок к A в некоторой окрестности начала координат. По теореме об устойчивом многообразии (теорема 9.5.2) имеется *устойчивая* кривая (заменяющая ось y для A), состоящая из точек, стремящихся к 0, и *неустойчивая* кривая, состоящая из точек, стремящихся к 0 при итерациях обратного отображения. Это описание означает, что обе кривые инвариантны. Кроме того, ни одна из них не может иметь самопересечения. Если бы устойчивая кривая имела самопересечения, образы получающихся петель приближались бы к 0, а это означало бы, что около нуля множество точек, стремящихся к 0, не является кривой вопреки утверждению теоремы 9.5.2.

Предположим, что вдали от начала координат нелинейные компоненты таковы, что кривые имеют достаточную кривизну для того, чтобы они могли пересекаться в точке p и угол между ними был ненулевым (см. рис. 12.2.2). Такая точка p называется *трансверсальной гомоклинической точкой* для неподвижной точки 0. Эта точка гомоклиническая в смысле определения 2.3.4, и данная ситуация соответствует разрыву гомоклинической петли в п. 6.2.2. (Сравните рис. 6.2.2 и рис. 4.3.3 с рис. 12.2.2.)

Естественно задать вопрос, откуда мы знаем, что такая ситуация возможна. Чтобы ответить на этот вопрос, мы можем немного изменить описание, сначала рассматривая отображение A лишь в окрестности точки 0, а затем рас-

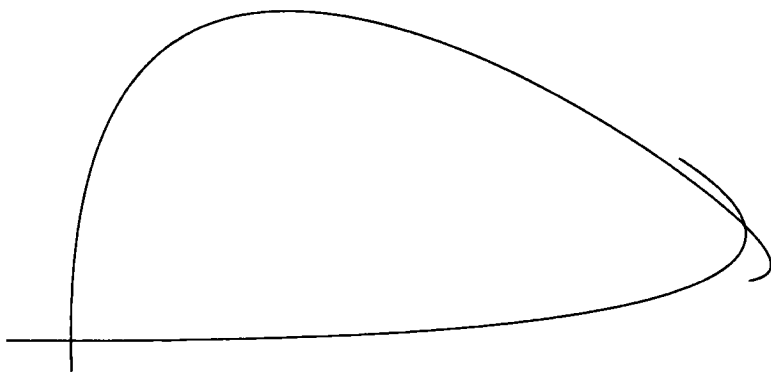


Рис. 12.2.2. Трансверсальная гомоклиническая точка

ширяя определение таким способом, чтобы образы/прообразы горизонтальных и вертикальных отрезков, проходящих через 0, удовлетворяли нужным условиям. С другой стороны, имеется следующее явное отображение плоскости (которое сохраняет площадь) с необходимым свойством:

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 3(x + (x - y)^2) \\ \frac{1}{3}(y + (x - y)^2) \end{pmatrix}.$$

Это *отображение Кремоны*¹ обратимо. Действительно, обратное отображение также имеет простой вид:

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \frac{x}{3} - \left(3y - \frac{x}{3}\right)^2 \\ 3y - \left(3y - \frac{x}{3}\right)^2 \end{pmatrix}.$$

Наконец, как мы уже отмечали, эта картина естественно возникает во всевозможных задачах из реальной жизни. Они часто появляются несколько другим образом: в упрощенной задаче две кривые изгибаются так, что образуют гладкую петлю в первом квадранте, и, когда мы производим возмущение, чтобы приблизить эту задачу к реальной, получается ненулевой угол пересечения. Примером могут служить деформации эллиптического бильярда и возмущения отображения математического маятника за единицу времени (см. п. 6.2.2).

Сложность поведения, которая возникает при этом, начинает проявляться вдоль орбиты $\mathcal{O}(p)$ гомоклинической точки p . Поскольку точка p лежит на устойчивой и на неустойчивой кривых, там же находится и любая точка ее орбиты. Поэтому все эти точки также гомоклинические. Поскольку отображение f — диффеоморфизм, углы, получающиеся при пересечении, всегда отличны от нуля. Эти точки накапливаются вблизи точки 0 по обеим кривым.

Любые две смежные точки орбиты $\mathcal{O}(p)$ определяют петлю с устойчивой и неустойчивой частями границы. Поскольку эти петли накапливаются около 0 (сверху) при положительном направлении времени, они сжаты вертикально

¹Об этом отображении нам сообщил Алекс Дрот.

и растянуты горизонтально. Поскольку неустойчивая кривая не может самопересекаться, эти все более длинные лепестки идут вдоль неустойчивой кривой все ближе друг к другу, как показано на рис. 12.2.3. В отрицательном времени

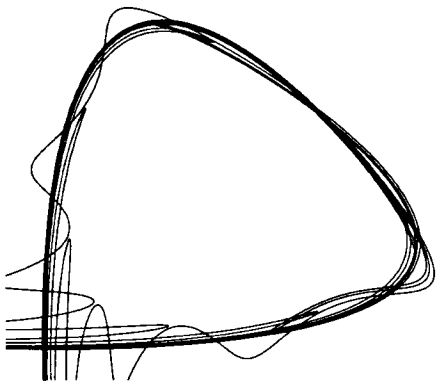


Рис. 12.2.3. Гомоклиническая картина

происходит дополнительное накопление с вертикальным растяжением. Отсюда мы получаем полную картину, изображенную на рис. 12.2.3. Из рисунка видно, что при этом возникает «второе поколение» гомоклинических точек, и мы рекурсивно приходим ко все большей сложности.

Гетероклинические картины, возникающие из трансверсальных гетероклинических точек, также сложны и встречаются очень часто. Например, они появляются в областях неустойчивости закручивающих отображений (см. следствие 14.3.3). Однако трансверсальные гетероклинические точки могут появляться и без существенной рекуррентности. Это происходит, например, в случае градиентного потока функции высоты на торе (поверхности в форме бублика), расположенном почти (но не совсем) вертикально². В этом случае не возникает заметной сложности. Гетероклинические картины появляются в случае гетероклинической *петли*. Так как картина очень похожа на гомоклиническую, мы не будем здесь отдельно заниматься гетероклиническими картинами.

§ 12.3. Возникновение подков

Теперь мы покажем, что растяжение и изгиб инвариантных кривых приводит к растяжению и изгибу прямоугольника в подкову.

Теорема 12.3.1 (Биркгоф—Смейл). Пусть $U \subset \mathbb{R}^2$ — окрестность точки 0 , $f: U \rightarrow \mathbb{R}^2$ — вложение (см. определение A.2.5), для которого 0 является гиперболической неподвижной точкой с трансверсальной гомоклинической точкой p . Тогда в любой малой окрестности точки 0 для некоторой итерации отображения f существует подкова.

Замечание 12.3.2. Это действительно замечательный результат, потому что он означает, что на любую гиперболическую неподвижную точку можно воздей-

²См.: А. Каток, Б. Хасселблат. Введение в современную теорию динамических систем, § 1.6.

ствовать «издалека» таким способом, что динамика вблизи нее будет обладать такой сложностью, как для подковы.

Доказательство. Появление подковы изображено на рис. 12.3.1. Если взять небольшой прямоугольник в первом квадранте с двумя сторонами вне инвариантных кривых, тогда при растяжении и изгибе инвариантных кривых прямоугольник

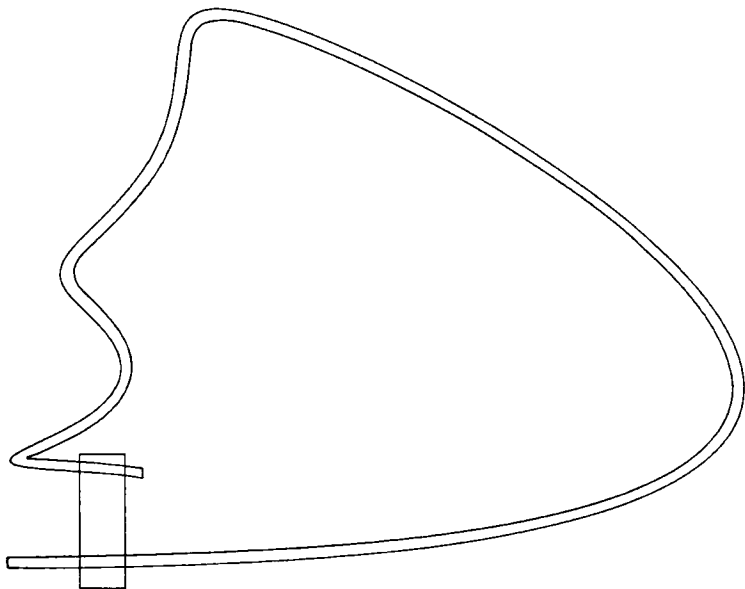


Рис. 12.3.1. Получение подковы

будет растягиваться и после нескольких итераций возникнут самопересечения. Рисунок показывает, что геометрическая интерпретация правильна и итерации отображения f приводят к необходимому горизонтальному растяжению и вертикальному сжатию.

Чтобы упростить обозначения, мы предположим, что вблизи нуля инвариантные кривые совпадают с осями координат (этого можно достичь заменой координат). Затем рассмотрим прямоугольник $\Delta = D_1 \times D_2$, где D_1 и D_2 — маленькие отрезки, для которых 0 является внутренней точкой. На рис. 12.3.1 мы выбираем их так, чтобы левый конец отрезка D_1 и нижний конец D_2 были расположены как можно ближе к 0.

Внутри отрезка D_1 имеется прообраз $p' = f^{-n}(p)$ и близлежащая часть устойчивой кривой, проходящая через точку q' почти вертикально. Все близлежащие части неустойчивых петель близки к горизонтальным. С одной стороны, для любого $n \in \mathbb{N}$ начало координат содержится в одной компоненте Δ_0 образа $f^n(\Delta) \cap \Delta$. С другой стороны, выбирая число n достаточно большим, мы получим вторую компоненту Δ_1 , удовлетворяющую первым трем условиям определения 12.1.1.

Выполнение условий растяжения и сжатия из определения 12.1.1 проверить несложно, потому что большинство итераций отображения f , необходимых для получения правильной геометрической картины, располагаются вблизи 0 и поэтому растягиваются и сжимаются почти так же, как и линейная часть. Если необходимо, можно немного уменьшить отрезки D_1 и D_2 , чтобы увеличить число итераций f . Все дополнительные итерации вносят равномерное растяжение и сжатие. $\square$

В задачах, возникающих в конкретных ситуациях, требуется относительно немного итераций (порядка 10), однако уже эти немногие итерации приводят к сильному растяжению и сжатию. Действительно, на рис. 12.3.1 мы видим очень большое растяжение, и если бы мы сделали точное изображение, то удлинённый прямоугольник выглядел бы как отдельная кривая благодаря соответствующей скорости сжатия в устойчивом направлении. На рис. 12.3.2 изображена другая



Рис. 12.3.2. Подкова, полученная из гомоклинического пересечения

подкова для отображения Кремоны, более пропорциональная, чем подкова, полученная при исследовании окрестности нуля.

§ 12.4. Применение подков

12.4.1. От гомоклинической картины до подков. Гомоклиническая картина впервые была рассмотрена Анри Пуанкаре в его работе, посвященной задаче трех тел, и, изучая ее, он предположил, что в солнечной системе может присутствовать сложная динамика. Позднее он описал эту ситуацию так:

Если попытаться представить себе фигуру, образованную этими двумя кривыми и их бесчисленными пересечениями, каждое из которых соответствует двойно-асимптотическому решению, то эти пересечения образуют нечто вроде решетки, ткани, сети с бесконечно тесными петлями; ни одна из двух кривых никогда не должна пересечь самое себя, но она должна навиваться на самое себя очень сложным образом, чтобы пересечь бесконечно много раз все петли сети.

Поражаешься сложности этой фигуры, которую я даже не пытаюсь изобразить. Ничто не является более подходящим, чтобы дать нам представление о сложности задачи трех тел, и, вообще, всех задач динамики, в которых нет однозначного интеграла.³

Несколькими десятилетиями позже Джордж Дэвид Биркгоф доказал, что в этой ситуации вблизи неподвижной точки имеется много периодических точек. Во время Второй мировой войны Мэри Люси Картрайт и Джон Эденсор Литлвуд на основе результатов Пуанкаре проанализировали работу радарной схемы, настроенной на параметры вне нормального операционного диапазона, для которой было характерно беспорядочное мигание контрольного света, т.е. явление, известное под названием релаксационных колебаний⁴. Примерно в то же время Норман Левинсон, изучая уравнение Ван дер Пола (для вакуумных труб), нашел бесконечно много периодических решений (сосуществующих со структурной устойчивостью)⁵. Примерно в 1960 г. он указал на эту особенность Стивену Смейлу, который получил из этой работы геометрическую картину подковы, изображенную на рис. 12.1.1, и в 1960-х гг. доказал теорему 12.3.1⁶. Хотя подковы встречаются и в высоких размерностях, все важные наблюдения, упомянутые выше, были сделаны для двумерных систем.

Следует подчеркнуть, что подковы, полученные из гомоклинической картины, являются множествами меры нуль, так что сложность в такой динамической системе может быть ограничена множеством такого вида, которым мы обычно пренебрегаем. Другими словами, из наличия гомоклинической картины не следует, что сложное поведение орбит в данной системе возникает повсеместно. Их важность заключается в возможности такой сложности, а также во влиянии на поведение других орбит на больших, хотя и конечных промежутках времени.

12.4.2. Широкое распространение подков в приложениях. Мы встречаемся с подковами во многих приложениях. Многочисленные динамические системы, которые возникают непосредственно из естественнонаучных задач, имеют трансверсальную гомоклиническую точку и, следовательно, содержат подкову.

³См. главу 33, § 396 книги: *J. H. Poincaré. Les méthodes nouvelles de la mécanique céleste.* Paris, 1892—1899. Имеется перевод: *А. Пуанкаре. Новые методы небесной механики* // Избранные труды. М.: Наука, 1972.

⁴*M. L. Cartwright. Forced Oscillations in Nonlinear Systems* // Contributions to the Theory of Nonlinear Oscillations. Princeton, NJ: Princeton University Press, 1950. (Annals of Mathematics Studies, v. 20). P. 149—241; *M. L. Cartwright, J. E. Littlewood. On Non-Linear Differential Equations of the Second Order. I. The Equation $\ddot{y} - k(1 - y^2)y + y = b\lambda k \cos(\lambda t + a)$, k large* // Journal of the London Mathematical Society. 1945. V. 20. P. 180—189; *J. E. Littlewood. On Non-Linear Differential Equations of the Second Order. IV. The General Equation $\ddot{y} + k f(y)\dot{y} + g(y) = b k p(\phi)$, $\phi = t + \alpha$* // Acta Mathematica. 1957. V. 98. P. 1—110.

⁵*N. Levinson. A Second Order Differential Equation with Singular Solutions* // Annals of Mathematics (2). 1949. V. 50. P. 127—153.

⁶*С. Смейл. Структурно устойчивый диффеоморфизм с бесконечным числом периодических точек* // Труды международного симпозиума по нелинейным колебаниям. Т. 2. Киев: Издат. АН УССР, 1963. С. 365—366; *S. Smale. Diffeomorphisms with Many Periodic Points* // Differential and Combinatorial Topology: A Symposium in Honor of Marston Morse / Edited by Stewart S. Cairns. Princeton, NJ: Princeton University Press, 1965. P. 63—80. Имеется перевод: *С. Смейл. Диффеоморфизмы со многими периодическими точками* // Математика. 1967. Т. 11, № 4. С. 88—106.

Кроме того, для любой динамической системы с гомоклиническими пересечениями все достаточно малые возмущения также имеют гомоклинические пересечения и, следовательно, также содержат подкову. Из этого, конечно же, следует, что подковы очень важны для приложений, потому что, если различие между моделью и действительной ситуацией достаточно мало, присутствие подковы в модели соответствует наличию подковы в реальной ситуации. Такая широта применения характерна для всех структурно устойчивых явлений. Устойчивость трансверсальных гомоклинических точек не только используется при изучении возмущений для отдельных примеров, но также показывает, что среди всевозможных динамических систем весьма часто встречаются системы с гомоклиническими пересечениями.

Отметим еще раз, что в таких ситуациях справедливы сильные результаты о сложности орбит динамических систем. Например, система в таком случае имеет положительную топологическую энтропию и экспоненциальный рост периодических точек с их периодами, и она содержит топологическую цепь Маркова в качестве подсистемы. Конечно же, сложность, возникающая из-за наличия подковы, может быть довольно ограниченной и иногда наблюдается лишь на множестве меры нуль. Однако даже возможность экспоненциальной сложности орбит интересна сама по себе (и из нее следует, что система не интегрируема). Кроме того, хотя подкова является множеством меры нуль, некоторые другие орбиты будут совпадать с орбитами в подкове в течение достаточно долгого периода времени, чтобы приобрести чувствительность к начальным условиям и другие особенности сложного поведения на конечных интервалах времени. В квадратичном семействе положительная топологическая энтропия свойственна (одномерным) подковам, но часто наблюдается только на множестве меры нуль, так как периодические орбиты притягивают все другие точки. Другой крайний случай — гиперболический автоморфизм тора, в котором подков чрезвычайно много, но любое их счетное множество есть множество меры нуль, в то время как экспоненциальная сложность орбит заполняет тор полностью.

Приведем несколько примеров возникновения гомоклинической картины. Гомоклинические пересечения, а следовательно, и подковы естественно присутствуют в закручивающих отображениях (см. определение 14.2.1), которые возникают при изучении различных динамических задач типа бильярда, небесной механики и разработки ускорителей элементарных частиц. При попытках решить задачу Штермера о движении заряженных частиц (из солнечного ветра) в магнитном поле Земли мы также встречаемся с гомоклиническими пересечениями⁷. Они служат причиной «неразрешимости задачи Штермера». Имеются также модели явлений переноса в механике сплошных сред, при изучении которых используются гетероклинические картины, возникающие как пересечение устойчивой и неустойчивой кривых двух гиперболических точек. В этих моделях букет лепестков, находящихся далеко от гомоклиники (как, например, вне левого края на рис. 12.2.3), очень похож на турбулентный волновой след, фотографии ко-

⁷A. J. Dragt, J. M. Finn. Insolubility of Trapped Particle Motion in a Magnetic Dipole Field // J. Geophys. Res. 1976. V. 81. P. 2327—2340.

торого получены при проведении опытов⁸. Это важный пример того, как связь переходных орбит с орбитами на подкове приводит к существенным эффектам.

Еще один важный пример, в котором подкова используется для постановки фундаментального математического и физического вопроса, возникает из небесной механики. До 1960-х гг. астрономы предполагали, что без внешних воздействий в задаче n тел (описывающей, скажем, нашу солнечную систему) не может быть захвата постороннего тела. А именно, считалось само собой разумеющимся, что имеется симметрия между долгосрочным поведением в положительном и отрицательном времени. В соответствии с этим предположением тело не могло иметь неограниченной в отрицательном времени орбиты без наличия неограниченной в положительном времени орбиты. Другими словами, солнечная система не могла «захватывать» другое астрономическое тело (если все тела рассматривать как точечные массы). И обратно, считалось, что если Земля всегда была в солнечной системе, тогда ее нельзя просто так удалить из системы без применения специальных внешних сил. При изучении задачи трех тел Алексеев рассеял этот миф, описав подкову и показав, что ее присутствие создает возможность захвата⁹. Это пример того, насколько принципиальна сама возможность сложной динамики, даже если она ограничена множеством меры нуль, т. е. ее последствия могут иметь вероятность нуль.

Не так давно появился результат Книпера и Вайсса, в котором гомоклинические пересечения также используются для того, чтобы придать сложность орбите с помощью подковы. Этот результат состоит в том, что геодезический поток на эллипсоиде (который является вполне интегрируемым и, следовательно, имеет очень небольшую сложность орбит) приобретает гомоклинические пересечения (а следовательно, и подкову, положительную топологическую энтропию, экспоненциальный рост замкнутых геодезических и т. п.) для некоторых произвольно малых деформаций эллипсоида¹⁰. Здесь различие между сложностью на множестве меры нуль и сложностью типичных орбит привело к вопросам, которые остаются открытыми: можно ли построить такие возмущения, которые не только имеют положительную топологическую энтропию (т. е. максимальная возможная сложность орбиты описывается экспоненциальным поведением), но также имеют положительную *метрическую* энтропию, или энтропию *в смысле теории меры*, которая измеряет *среднюю* сложность орбиты.

12.4.3. Хаос на плоскости возникает из подков. В конце этого параграфа мы приведем наиболее поразительный результат. Если динамическая система на

⁸ V. Rom-Kedar, A. Leonard, S. Wiggins. An Analytical Study of Transport, Mixing and Chaos in an Unsteady Vortical Flow // Journal of Fluid Mechanics. 1990. V. 214. P. 347—394.

⁹ В. М. Алексеев. О возможности захвата в задаче трех тел при отрицательном значении константы полной энергии // УМН. 1969. Т. 24, № 1(145). С. 185—186.

¹⁰ G. Knepper, H. Weiss. A Surface with Positive Curvature and Positive Topological Entropy // Journal of Differential Geometry 1994. V. 39, no. 2. P. 229—249; см. также: G. P. Paternain. Real Analytic Convex Surfaces with Positive Topological Entropy and Rigid Body Dynamics // Manuscripta Mathematica. 1993. V. 78, no. 4. P. 397—402 и V. J. Donnay. Transverse Homoclinic Connections for Geodesic Flows // Hamiltonian dynamical systems (Cincinnati, OH, 1992). New York: Springer, 1995. (IMA Volumes in Mathematics and its Applications, v. 63). P. 115—125.

плоскости имеет положительную топологическую энтропию, вся эта энтропия возникает вследствие подков¹¹.

Теорема 12.4.1. *Для диффеоморфизма поверхности топологическая энтропия может быть сколь угодно хорошо приближена ограничениями на подкову.*

В этом смысле для образования хаоса на плоскости достаточно только подков. Это позволяет хорошо понять динамику сложных отображений в размерности 2 с экспоненциально сложной структурой орбит.

Этот результат помогает управлять поведением энтропии при возмущениях.

Теорема 12.4.2. *Энтропия как функция на пространстве диффеоморфизмов поверхности полунепрерывна снизу в C^1 -топологии (см. п. А.2.2).*

Доказательство. Эквивалентная формулировка такова: для любого диффеоморфизма поверхности f и любого $\epsilon > 0$ найдется такое $\delta > 0$, что если $d(f, g) < \delta$ (в C^1 -метрике), то $h_{\text{top}}(g) \geq h_{\text{top}}(f) - \epsilon$.

Для доказательства выберем произвольный такой диффеоморфизм f . Если $h_{\text{top}}(f) = 0$, то утверждение очевидно. В противном случае имеется такая подкова Λ отображения f , что $h_{\text{top}}(f|_{\Lambda}) \geq h_{\text{top}}(f) - \epsilon$. Возьмем такое $\delta > 0$, что (с учетом структурной устойчивости подков) любой диффеоморфизм g , удовлетворяющий неравенству $d(f, g) < \delta$, содержит подкову Λ' , которая топологически сопряжена с Λ . Ее энтропия тогда совпадает с энтропией подковы Λ , так что $h_{\text{top}}(g) \geq h_{\text{top}}(g|_{\Lambda'}) = h_{\text{top}}(f|_{\Lambda}) \geq h_{\text{top}}(f) - \epsilon$. $\square$

Аналогичный результат также справедлив для непрерывных отображений отрезка¹², но упражнение 8.3.7 показывает, что в других ситуациях это может быть и не так. Полунепрерывность сверху можно получить без ограничений на размерность, но вопрос оказывается гораздо более тонким в том, что касается требуемой гладкости: результат справедлив для C^∞ -отображений¹³, но доказательство основано на сложной оценке¹⁴, которая не выполняется для конечно дифференцируемых отображений. Для таких отображений можно только получить оценку того, как может нарушаться полунепрерывность сверху.

§ 12.5. Обнаружение гомоклинических пересечений: метод Пуанкаре—Мельникова

Как мы видели в предыдущих параграфах, трансверсальные гомоклинические точки и возникающая из них гомоклиническая картина представляют большой интерес для математики и естественнонаучных приложений. Поэтому важно

¹¹A. Katok. Nonuniform Hyperbolicity and Structure of Smooth Dynamical Systems // Proceedings of the International Congress of Mathematicians (Warsaw, 1983). Warsaw: PWN, 1984. P. 1245—1254. См. также приложение к книге: А. Каток, Б. Хасселблат. Введение в современную теорию динамических систем.

¹²M. Misiurewicz. Horseshoes for Continuous Mappings of an Interval // Dynamical systems (Bressanone, 1978). Naples: Liguori, 1980. P. 125—135.

¹³S. Newhouse. Continuity Properties of Entropy // Annals of Mathematics (2). 1989. V. 129. P. 215—235.

¹⁴Y. Yomdin. Volume Growth and Entropy // Israel Journal of Mathematics. 1987. V. 57. P. 285—300.

иметь метод, с помощью которого можно обнаружить трансверсальные гомоклинические точки. Численные эксперименты не могут дать определенного результата, если угол в гомоклинической точке мал, поскольку вся гомоклиническая динамика сжата в узкую окрестность локальных устойчивых и неустойчивых многообразий гиперболической точки и ее не всегда можно обнаружить с помощью компьютерных картинок. Кроме того, численные эксперименты не дают доказательства существования данного явления.

Чтобы объяснить суть метода Пуанкаре—Мельникова, рассмотрим поток φ^t на плоскости, определенный дифференциальным уравнением $\dot{x} = f(x)$, имеющим гиперболическое седло в точке 0 и гомоклиническую петлю $\Gamma := \{\varphi^t(q) : t \in \mathbb{R}\} \cup \{0\}$. Метод Пуанкаре—Мельникова определяет появление трансверсальных гомоклинических точек при возмущениях¹⁵. А именно, произведем возмущение потока φ_ϵ^t , которое определяется уравнением $\dot{x} = f(x) + \epsilon g_t(x)$, где для простоты мы полагаем, что функция g_t периодична по t и $g(0) = 0$. В качестве примеров приведем деформации эллиптического билиарда и упомянутую выше работу Книпера и Вайсса.

Нам необходимо изучить устойчивую кривую W_ϵ^s точки 0 для φ_ϵ^t и соответствующую неустойчивую кривую W_ϵ^u . Орбиты на них имеют вид $\varphi^t(q) + \epsilon q^s(t) + O(\epsilon^2)$ ($t \geq 0$) и $\varphi^t(q) + \epsilon q^u(t) + O(\epsilon^2)$ ($t \leq 0$) соответственно, где

$$\dot{q}^s(t) = Df(\varphi^t(q))q^s(t) + g_t(\varphi^t(q)) \quad \text{для } t \geq 0,$$

$$\dot{q}^u(t) = Df(\varphi^t(q))q^u(t) + g_t(\varphi^t(q)) \quad \text{для } t \leq 0$$

(линеаризованный поток). До порядка ϵ^2 расстояние $d(q)$ между W_ϵ^s и W_ϵ^u вблизи q определяется как проекция вектора $\epsilon(q^u(0) - q^s(0))$ на нормаль Jf , так что $d(q) = \epsilon(q^u(0) - q^s(0))Jf(q)/\|Jf(q)\|$. Поскольку мы можем определить Q^u и Q^s на основании линеаризованного потока, эту разность тоже можно получить с помощью линеаризованного потока. *Функция Мельникова* определена на Γ и имеет вид

$$M(q) = \int_{-\infty}^{\infty} g_t(\varphi^t(q))Jf(\varphi^t(q)) dt.$$

Заметим, что этот интеграл вычисляется только с использованием информации о невозмущенном потоке и слагаемом, описывающем возмущение векторного поля.

Теорема 12.5.1. *Если функция Мельникова имеет простой нуль вдоль Γ , тогда для всех достаточно малых $\epsilon \neq 0$ имеется трансверсальная гомоклиническая точка φ_ϵ^t . Если $M \neq 0$ всюду, кроме точки 0, то $W_\epsilon^s \cap W_\epsilon^u = \{0\}$.*

В некоторых важных ситуациях расщепление экспоненциально мало. Это вызывает трудности, которые были замечены еще двадцать лет назад¹⁶ и над преодолением которых с того времени продолжается работа¹⁷.

¹⁵В. К. Мельников. Об устойчивости центра при периодических по времени возмущениях // Труды Московского Математического общества. 1963. Т. 12. С. 3—52.

¹⁶J. A. Sanders. A Note on the Validity of Melnikov's Method. Report 139, Wiskundig Seminarium, Vrije Universiteit Amsterdam.

¹⁷См., например: V. G. Gelfreich. A Proof of the Exponentially Small Transversality of the Separatrices for the Standard Map // Communications in Mathematical Physics. 1999. V. 201, no. 1. P. 155—216.

§ 12.6. Гомоклинические касания

Может случиться, что в однопараметрическом семействе f_ϵ диффеоморфизмов пространства $\mathbb{R}^2$ с гомоклиническими пересечениями, ассоциированными с гиперболической неподвижной точкой p , для $\epsilon = 0$ «язык» устойчивого многообразия касается неустойчивого многообразия в точке q . Это явление называется *гомоклиническим касанием*. Возможно, что q — «первичное» пересечение и ни для отрицательных, ни для положительных значений ϵ нет никаких гомоклинических пересечений вовсе. Возможно также, что q является точкой пересечения более высокого порядка и возникает из трансверсальной гомоклинической точки, которая сохраняется для всех малых ϵ . В любом случае эта ситуация является источником возникновения дополнительной динамической сложности. Заметим, что мы здесь изучаем гиперболичность, но не в равномерно гиперболическом контексте, характерном для предшествующих параграфов, потому что переходное гомоклиническое касание приводит к глобальной бифуркации, которая несовместима со структурной устойчивостью гиперболических множеств. Таким образом, мы получаем новые методы, но при этом в них существенно используется гиперболичность.

Мы предполагаем, что гомоклиническое касание находится в общем положении, а именно что имеется локальная система координат вблизи точки касания q , в которой $W^s(p)$ представляется осью x , а $W^u(p)$ определяется уравнением $y = \epsilon - x^2$.

Источник дополнительной динамической сложности — появление новых подков.

Предложение 12.6.1. *Если однопараметрическое семейство f_ϵ диффеоморфизмов плоскости $\mathbb{R}^2$ сжимает объем в точке p (т.е. $\|Df_\epsilon(p)\| < 1$) и имеет гомоклиническое касание общего положения для $\epsilon = 0$, то для $\epsilon > 0$ появляется дополнительная подкова.*

Мы можем наблюдать этот эффект на рис. 12.6.1, на котором изображена гомоклиническая бифуркация с требуемым прямоугольником и его образом.

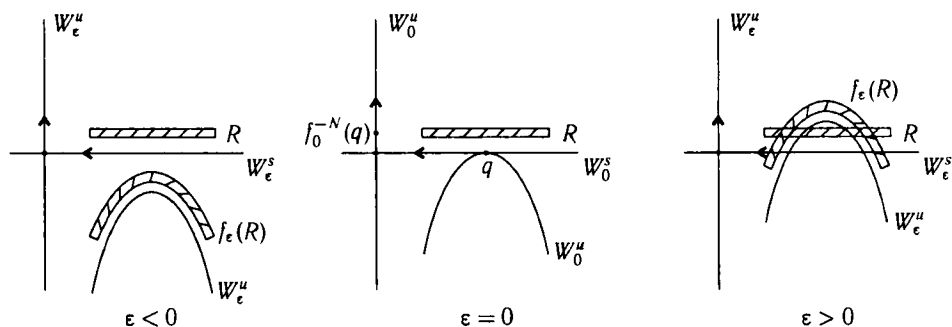


Рис. 12.6.1. Гомоклинические касания

Это явление позволяет получить большее количество подков и тем самым приводит к дополнительной динамической сложности того же самого типа, что

присутствует в любой гомоклинической картине. Однако есть также эффекты, которые имеют в целом иную природу. Они появляются как раз перед бифуркацией.

Предложение 12.6.2 (явление Ньюхауса)¹⁸. *Если однопараметрическое семейство f_ϵ диффеоморфизмов плоскости $\mathbb{R}^2$ сжимает объем в точке p (т.е. $\|Df_\epsilon(p)\| < 1$) и имеет гомоклиническое касание общего положения для $\epsilon = 0$, то для остаточного множества малых значений $\epsilon < 0$ имеется бесконечно много притягивающих периодических орбит (стоков).*

Остаточное множество — пересечение счетного числа открытых и плотных множеств (согласно лемме А.1.15 из этого следует, что оно плотно).

Теперь понятно, что каждое гомоклиническое касание общего положения является источником динамической сложности, а также и существенного изменения в динамике. Аналогично тому, что для гомоклинических пересечений характерна иерархия, т.е. петли («языки») все более высоких порядков (а на самом деле именно по этой причине), гомоклиническое касание никогда не встречается отдельно. Если рассмотреть гомоклинические пересечения петель более высокого порядка, можно заметить, что многие малые возмущения также должны иметь касания.

Предложение 12.6.3. *Если однопараметрическое семейство f_ϵ диффеоморфизмов плоскости $\mathbb{R}^2$ имеет гомоклиническое касание общего положения для $\epsilon = 0$, то существует такая последовательность $\epsilon_n \rightarrow 0$, что f_{ϵ_n} имеет гомоклиническое касание для всех $n \in \mathbb{N}$ в точках $q_n \rightarrow q$.*

Нужно иметь в виду, что большинство этих касаний получены из «языков» высоких порядков и соответственно для этих касаний характерны эффекты, локализованные в малых областях и связанные с высокими итерациями диффеоморфизма. Поэтому сложность добавляется в малых пространственных масштабах и больших масштабах времени. Если учесть, что образование подковы увеличивает энтропию, тогда ясно, что в силу непрерывности энтропии (которая следует из теоремы 12.4.2 и упомянутого выше результата о полунепрерывности сверху) большинство этих изменений должны быть именно такого типа, чтобы ограничить накопление сложности орбитальной структуры. Таким образом, в этой сложной последовательности гомоклинических касательных бифуркаций маленькие подковы для высоких итераций вносят некоторый вклад в сложность орбитальной структуры и энтропию. И в то же время для каждого значения параметра сложность орбитальной структуры в основном можно понять в терминах тех же самых общих свойств, которые мы описали для всех равномерно гиперболических динамических систем.

¹⁸S. Newhouse. Diffeomorphisms with Infinitely Many Sinks // Topology. 1974. V. 13. P. 9—18.

СТРАННЫЕ АТТРАКТОРЫ

Странные аттракторы — это один из самых популярных разделов динамики. Это аттракторы со сложной геометрической структурой, не похожие на простые кривые или поверхности. Прежде чем приступить к изучению странных аттракторов, полезно рассмотреть некоторые геометрически простые аттракторы. От них мы перейдем к изучению странных аттракторов начиная с важного модельного примера.

Характерной особенностью странных аттракторов является гиперболичность того или иного рода. Однако при изучении наиболее известных примеров возникают трудности, поскольку гиперболичность этих аттракторов имеет более слабую форму, чем равномерная гиперболичность, рассмотренная в гл. 10. Она позволяет получить сложное поведение, но ее все же не достаточно, чтобы непосредственно применять результаты из гл. 10.

Динамическое поведение странных аттракторов часто бывает схоже с неравномерно гиперболическим поведением, которое можно наблюдать для стохастических квадратичных отображений (случай Якобсона—Колле—Экмана; см. п. 11.4.3.6). Более того, квадратичные отображения часто используются как модели, а также служат в качестве первого шага при построении известных странных аттракторов с помощью возмущений. Однако мы рассмотрим только аттрактор Лоренца, при изучении которого возникают трудности иного рода, а именно «равномерная гиперболичность с особенностями». Доказательство существования аттрактора Лоренца — один из самых ярких примеров строгих доказательств с использованием компьютеров в непрерывной математике.

§ 13.1. Знакомые аттракторы

С некоторыми аттракторами мы уже встречались. Самые простые из них — это притягивающие неподвижные точки, которые были формально определены в гл. 2 (см. определение 2.2.22). Имеется эквивалентное определение притягивающей неподвижной точки, которое более удобно для обобщений.

Определение 13.1.1. Неподвижная точка p отображения $f: X \rightarrow X$ метрического пространства называется *притягивающей неподвижной точкой*, если имеется такая окрестность U точки p , что $f(U) \subset U$ и $\bigcap_{n \in \mathbb{N}} f^n(U) = \{p\}$.

Другой уже известный аттрактор — это предельный цикл, рассмотренный в п. 2.4.3. Мы определили его как периодическую точку p , имеющую окрестность, каждая точка которой асимптотически стремится к $\mathcal{O}(p)$. Аналогично определению 13.1.1 мы можем дать следующее определение.

Определение 13.1.2. Периодическая точка p называется *притягивающей периодической точкой* (или *предельным циклом*) потока ϕ^t , если имеется такая окрестность U орбиты $\mathcal{O}(p)$, что $\phi^t(U) \subset U$ для $t \geq 0$ и $\bigcap_{t \geq 0} \phi^t(U) = \mathcal{O}(p)$.

Привлекательная особенность этих определений заключается в том, что они не нуждаются в понятии «асимптотического стремления к чему-либо», которое является несколько более тонким, когда предел асимптотик отличен от точки.

Предположим, что p — притягивающая периодическая точка потока ϕ^t , и рассмотрим отображение $f(x) = \phi^1(x)$. Тогда орбита точки p при действии потока представляет собой множество, которое притягивает точки при повторном применении отображения f . Чтобы сделать наши рассуждения более строгими, изменим определение 13.1.2. Мы можем заменить $\mathcal{O}(p)$ любым множеством и определить, является ли оно притягивающим.

Определение 13.1.3. Предположим, что $f: X \rightarrow X$ — некоторое отображение. Компактное множество $A \subset X$ называется *аттрактором* для f , если имеется такая окрестность U множества A , что $f(U) \subset U$ и $\bigcap_{n \in \mathbb{N}} f^n(U) = A$. Обычно также требуется, чтобы множество A не имело собственных подмножеств, обладающих тем же свойством.

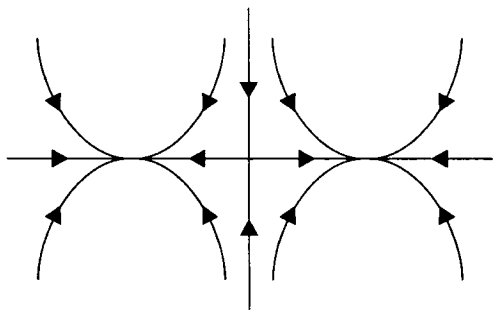


Рис. 13.1.1. Не аттрактор

Требование, чтобы множество A являлось «наименьшим», т. е. не имело собственных притягивающих подмножеств, можно проиллюстрировать с помощью следующего простого примера. На рис. 13.1.1 изображена динамическая система, для которой множество $A = \{(x, y) \in \mathbb{R}^2: x \in [0, 1]\}$ обладает всеми свойствами, перечисленными в определении 13.1.3, за исключением того, что оно содержит аттракторы $(-1, 0)$ и $(1, 0)$. Поэтому мы не считаем множество A аттрактором.

Более существенный, но все еще простой пример аттрактора можно построить следующим образом. Рассмотрим отображение окружности f_1 с притягивающей неподвижной точкой, например такое, как на рис. 4.3.3. Предположим, что это точка 0 . Если f_2 — любое отображение тора $\mathbb{T}^2$, например параллельный перенос или гиперболический автоморфизм, то отображение $f: \mathbb{T}^3 \rightarrow \mathbb{T}^3$, $f(x, y, z) = (f_1(x), f_2(y, z))$, имеет в качестве аттрактора двумерный тор $\{0\} \times \mathbb{T}^2$.

Таким образом, притягивающие неподвижные точки, предельные циклы, притягивающие периодические орбиты и притягивающие инвариантные поверхности

являются примерами аттракторов. Однако имеется множество других примеров. Те аттракторы, которые не имеют сравнительно простой структуры, получили название странных.

§ 13.2. Соленоид

Странные аттракторы часто встречаются в системах, возникающих из приложений динамики. Такие аттракторы иногда трудно исследовать, и поэтому полезно сначала рассмотреть некоторую модельную ситуацию.

Мы опишем один класс примеров, предложенный Смейлом в 1960-х гг. Он называется *аттрактором Смейла* или *соленоидом*.

Рассмотрим трехмерный тор $M := S^1 \times D^2$, где D^2 — единичный диск в $\mathbb{R}^2$. Он напоминает бублик. На нем определим такие координаты (φ, x, y) , что $\varphi \in S^1$ и $(x, y) \in D^2$, т. е. $x^2 + y^2 \leq 1$. Используя эти координаты, определим отображение, которое удваивает тор, при этом в 5 раз уменьшая его толщину.

Предложение 13.2.1. Отображение

$$f: M \rightarrow M := S^1 \times D^2, \quad f(\varphi, x, y) = \left(2\varphi, \frac{1}{5}x + \frac{1}{2}\cos\varphi, \frac{1}{5}y + \frac{1}{2}\sin\varphi\right),$$

определено корректно и является инъективным.

Доказательство. Отображение определено корректно, т. е. $f(M) \subset M$, поскольку

$$\begin{aligned} \left(\frac{1}{5}x + \frac{1}{2}\cos\varphi\right)^2 + \left(\frac{1}{5}y + \frac{1}{2}\sin\varphi\right)^2 &= \\ &= \frac{1}{25}(x^2 + y^2) + \frac{1}{5}(x\cos\varphi + y\sin\varphi) + \frac{1}{4}(\cos^2\varphi + \sin^2\varphi) \leq \frac{1}{25} + \frac{2}{5} + \frac{1}{4} < 1. \end{aligned}$$

Следовательно, образ $f(M)$ содержится во внутреннейности множества M .

Инъективность отображения f не удивительна, поскольку происходит сильное уменьшение толщины. Предположим, что $f(\varphi_1, x_1, y_1) = f(\varphi_2, x_2, y_2)$. Тогда

$$\begin{aligned} 2\varphi_1 &= 2\varphi_2 \pmod{2\pi}, \\ \frac{1}{5}x_1 + \frac{1}{2}\cos\varphi_1 &= \frac{1}{5}x_2 + \frac{1}{2}\cos\varphi_2, \\ \frac{1}{5}y_1 + \frac{1}{2}\sin\varphi_1 &= \frac{1}{5}y_2 + \frac{1}{2}\sin\varphi_2. \end{aligned}$$

Если $\varphi_1 = \varphi_2$, тригонометрические функции взаимно уничтожаются, так что $x_1 = x_2$ и $y_1 = y_2$. Если $\varphi_1 = \varphi_2 + \pi$, то

$$\begin{aligned} \frac{1}{5}x_1 + \frac{1}{2}\cos\varphi_1 &= \frac{1}{5}x_2 - \frac{1}{2}\cos\varphi_1, \\ \frac{1}{5}y_1 + \frac{1}{2}\sin\varphi_1 &= \frac{1}{5}y_2 - \frac{1}{2}\sin\varphi_1, \end{aligned}$$

или

$$\frac{1}{5}(x_2 - x_1) = \cos\varphi_1 \quad \text{и} \quad \frac{1}{5}(y_2 - y_1) = \sin\varphi_1,$$

откуда следует, что

$$(x_2 - x_1)^2 + (y_2 - y_1)^2 = 25.$$

Так как левая часть неравенства не превосходит 8, это невозможно. $\square$

Можно использовать и более умеренное сжатие (меньше чем в 5 раз). То же самое рассуждение, но с использованием неравенства $x \cos \varphi + y \sin \varphi \leq x + y \leq \sqrt{2}$, показывает, что отображение

$$f: M \rightarrow M := S^1 \times D^2, \quad f(\varphi, x, y) = \left(2\varphi, \frac{1}{3}x + \frac{1}{2}\cos\varphi, \frac{1}{3}y + \frac{1}{2}\sin\varphi\right),$$

корректно определено и инъективно.

Образ $f(M)$ пересекает любое сечение $C = \{\theta\} \times D^2$ множества M по двум непересекающимся дискам радиуса $1/5$, как это показано на рис. 13.2.1. Действительно, $C \cap f(M) = f(C_1) \cup f(C_2)$, где C_1 и C_2 — сечения. При итерациях эта картина повторяется в меньших масштабах.

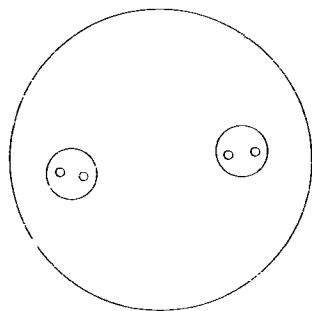


Рис. 13.2.1. Аттрактор Смейла и сечение

Ясно, что $f^2(M) \subset f(M)$, но, кроме того, $C \cap f^2(M) = f(C_1 \cap f(M)) \cup f(C_2 \cap f(M))$, где сечения C_1 и C_2 определены так же, как и прежде. Таким образом, множество $C \cap f^2(M)$ состоит из четырех небольших дисков, по два в каждом из множеств $f(C_1)$ и $f(C_2)$ (см. рис. 13.2.1), и множество $f^2(M)$ делает четыре оборота вокруг M .

Итак, мы заключаем по индукции, что множество $C \cap f^{l+1}(M)$ состоит из 2^{l+1} дисков, по два в каждом из множеств $C \cap f^l(M)$.

Геометрически эта картина напоминает работу автоматов, производящих сахарную вату, которые можно видеть на курортах. Эти машины постоянно растягивают и сворачивают вату (состоящую из патоки или сахара), и в результате получается хорошо организованная волокнистая структура, которая, конечно же, не является ни кривой, ни поверхностью. Имеются другие примеры использования подобного метода для практического производства. Так, при изготовлении японских мечей производится неоднократное сворачивание и сглаживание стального листа, чтобы получить хорошо выровненную молекулярную структуру. После 20 сворачиваний металлического листа толщиной меньше 1 см мы получаем слой толщиной меньше 2^{-20} см $< 10^{-6}$ см = 10 нм, а эта величина сопоставима с раз-

мером атома ($\approx 10^{-8}$ см). Это растяжение и сворачивание напоминает построение подковы в п. 7.4.4.

Эти два примера из реальной жизни отличаются от соответствующей модельной ситуации тем, что в наших моделях происходит некоторая потеря объема. В аттракторе Смейла пятикратное сжатие в двух направлениях, объединенное лишь с удвоением в оставшемся направлении, уменьшает объем больше чем в 10 раз при каждой итерации. В случае подковы мы теряем объем, удаляя части, которые не попадают в прямоугольник. Однако примеры с сахарной ватой и мечами служат иллюстрацией растяжения и сворачивания, приводящих к сложной динамике, которую мы наблюдали в гл. 7, и к подобной сложности в случае аттрактора Смейла. И как видно из определения 13.1.3, потеря объема неизбежна при наличии аттрактора.

Аттрактор Смейла является *гиперболическим*: растяжение равномерно (всегда в 2 раза), поскольку угловая координата умножается на постоянный множитель. Это делает динамику максимально хаотической и приводит к полной сложности, описанной в гл. 10. Поэтому аттрактор Смейла можно легко изучать с помощью методов, предложенных в гл. 10.

Имеется следующая непосредственная связь между аттрактором Смейла и конструкцией обратного предела (см. определение 7.1.12).

Предложение 13.2.2. *Аттрактор Смейла является обратным пределом удвоений окружности E_2 .*

Поскольку гиперболические аттракторы так удобны для изучения, термин «странный аттрактор» часто используется только применительно к аттракторам, обладающим сходной сложностью, но не имеющим равномерной нижней границы для коэффициентов растяжения. Рассмотрим известный пример такого рода.

§ 13.3. Аттрактор Лоренца

В 1961 г. метеоролог Эдвард Нортона Лоренц с помощью своего нового компьютера занимался моделированием погоды на основании законов, выражающих соотношения между температурой и давлением, давлением и скоростью ветра и т. д. Однажды он решил продолжить предыдущий эксперимент в течение более продолжительного времени и в качестве исходного значения ввел начальные данные предыдущего эксперимента. Однако вместо повторения эксперимента Лоренц получил быструю расходимость выходных данных. Тогда он обратил внимание на то, что ввел данные с точностью до 3 знаков после запятой вместо шести, и понял, что разница в нескольких тысячных долей привела к появлению макроскопических ошибок.

В 1963 г. Лоренц сильно упростил модель конвекции, которая использовалась в науке об атмосфере, сведя ее к дифференциальным уравнениям

$$\begin{aligned}\dot{x} &= \sigma(y - x), \\ \dot{y} &= rx - y - xz, \\ \dot{z} &= xy - bz,\end{aligned}\tag{13.3.1}$$

где $\sigma, r, b > 0$. (Он проделал тщательный анализ для частных значений $\sigma = 10$, $b = 8/3$, $r = 28$.) Те же самые уравнения с тех пор неоднократно появлялись в других моделях. Оказывается, они описывают динамо-машину (предтечу современных генераторов), и есть предположения, что наблюдаемые при их изучении изменения направления могут объяснить происходившие в разные геологические эпохи изменения магнитного поля Земли. В качестве механической реализации этих уравнений можно рассмотреть беспорядочно вращающееся водяное колесо¹.

Качественное поведение решений этих дифференциальных уравнений существенно зависит от значений параметров. Сейчас мы можем легко проверить это численно (с некоторой степенью точности, зависящей от того, насколько надежную картину может дать численный эксперимент), но стандартные методы теории обыкновенных дифференциальных уравнений позволили Лоренцу провести весьма глубокий анализ этих уравнений.

13.3.1. Случай $r < 1$. Самое простое поведение наблюдается для $r < 1$.

Предложение 13.3.1. Для $r < 1$ все решения уравнений Лоренца (13.3.1) стремятся к началу координат при $t \rightarrow \infty$.

Доказательство. Прежде всего заметим, что для $r < 1$ начало координат является единственной неподвижной точкой: из уравнений (13.3.1) видно, что если $\dot{x} = 0$, то $x = y$, и если $0 = \dot{z} = xy - bz = x^2 - bz$, то $z \geq 0$, а следовательно, $r - 1 - z < 0$. Если, кроме того, $0 = \dot{y} = rx - y - xz = x(r - 1 - z)$, мы получаем, что $x = 0$, а следовательно, $y = 0$ и (поскольку $\dot{z} = 0$) $z = 0$.

Хотя это и не требуется для доказательства, заметим, что линеаризация

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} \sigma(x - y) \\ rx - y \\ -bz \end{pmatrix} = \begin{pmatrix} -\sigma & \sigma & 0 \\ r & -1 & 0 \\ 0 & 0 & -b \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (13.3.2)$$

уравнений (13.3.1) имеет только отрицательные собственные значения при $r < 1$. Это означает, что начало координат является аттрактором.

Чтобы показать, что это глобальный аттрактор, используем функцию Ляпунова. Это положительная функция, убывающая вдоль орбит. Выбор этой функции для конкретной системы уравнений — это своего рода искусство. В данном случае в качестве функции Ляпунова мы выберем функцию $L(x, y, z) := x^2/\sigma + y^2 + z^2$, множества уровня которой — эллипсоиды с центрами в начале координат. Тогда L убывает вдоль каждой орбиты. В соответствии с правилом дифференцирования сложной функции

$$\begin{aligned} \dot{L} &= \frac{d}{dt}(x^2/\sigma + y^2 + z^2) = 2x\dot{x}/\sigma + 2y\dot{y} + 2z\dot{z} = \\ &= 2x(y - x) + 2y(rx - y - xz) + 2z(xy - bz) = 2(r + 1)xy - 2x^2 - 2y^2 - 2bz^2 = \\ &= -2\left(x - \frac{r+1}{2}y\right)^2 - 2\left(1 - \left(\frac{r+1}{2}\right)^2\right)y^2 - 2bz^2, \end{aligned}$$

¹ S. Strogatz. Nonlinear Dynamics and Chaos. Addison-Wesley, 1994. P. 302.

а это выражение всегда неположительно и обращается в нуль, только если каждое слагаемое равно нулю, т. е. $y = z = 0$ (из последних двух слагаемых) и $x = 0$ (из первого слагаемого). Таким образом, функция L убывает вдоль каждой орбиты, т. е. орбиты постоянно переходят во все меньшие и меньшие эллипсоиды, а это означает, что они приближаются к началу координат.

Убывание функции L вдоль каждой орбиты само по себе не означает, что она убывает к 0. Но это должно быть так, потому что для каждой орбиты мы имеем $\dot{L} \rightarrow 0$ (так как функция L ограничена снизу), а это означает, что $(x, y, z) \rightarrow 0$. Таким образом, 0 — глобальный аттрактор. $\square$

Внимательно посмотрев на эти рассуждения, можно увидеть, что для $r = 1$ качественное поведение в основном такое же. Анализ функции Ляпунова становится более сложным, поскольку исчезает слагаемое $(1 - (\frac{r+1}{2})^2)y^2$, но осложнения все же можно преодолеть с помощью стандартных методов теории обыкновенных дифференциальных уравнений.

13.3.2. Случай $r > 1$. При $r > 1$ возникают еще два положения равновесия. Из равенства $\dot{x} = 0$ все еще следует, что $x = y$. Из уравнений (13.3.1) видно, что если $0 = \dot{y} = rx - y - xz = x(r - 1 - z)$, то или $x = 0$ (а следовательно, $y = 0$ и $z = 0$, поскольку $0 = \dot{z} = xy - bz = -bz$), или $z = r - 1$. В первом случае мы получаем уже известное положение равновесия в начале координат. Если же $z = r - 1$, то с учетом равенства $0 = \dot{z} = xy - bz = x^2 - bz$ мы получаем $x^2 - b(r - 1) = 0$, т. е. $x = \pm\sqrt{b(r - 1)}$. Следовательно, появляются два новых положения равновесия $(x_0, y_0, z_0) = (\sqrt{b(r - 1)}, \sqrt{b(r - 1)}, r - 1)$ и $(-\sqrt{b(r - 1)}, -\sqrt{b(r - 1)}, r - 1)$, которые стремятся к 0 при $r \rightarrow 1^+$. Таким образом, $r = 1$ — бифуркационное значение. (Тот факт, что расстояния между положениями равновесия порядка квадратного корня, согласуется с тем, что мы наблюдали для бифуркаций удвоения периода квадратичного семейства; см. соотношение (11.2.2) и рис. 11.2.3.)

Предложение 13.3.2. Если $\sigma > b + 1$ и $1 < r < \sigma \frac{\sigma + b + 3}{\sigma - b - 1}$, тогда эти положения равновесия устойчивы.

Схема доказательства. Чтобы линеаризовать систему (13.3.1) в этих точках, положим $\xi := x - x_0$, $\eta := y - y_0$ и $\mu := z - z_0$. Тогда

$$\begin{aligned}\dot{\xi} &= \sigma(x_0 + \xi - y_0 - \eta) = [\sigma(x_0 - y_0)] + \sigma(\xi - \eta), \\ \dot{\eta} &= r(x_0 + \xi) - y_0 - \eta - (x_0 + \xi)(z_0 + \mu) = [rx_0 - y_0 - x_0z_0] + r\xi - \eta - (x_0\mu + z_0\xi + \xi\mu), \\ \dot{\mu} &= (x_0 + \xi)(y_0 + \eta) - b(z_0 + \mu) = [x_0y_0 - bz_0] + x_0\eta + y_0\xi + \xi\eta - b\mu.\end{aligned}$$

Слагаемые в квадратных скобках равны нулю. Опуская слагаемые, которые не линейны по ξ , η , μ , и используя равенства $x_0 = y_0$ и $z_0 = r - 1$, мы получаем

$$\begin{pmatrix} \dot{\xi} \\ \dot{\eta} \\ \dot{\mu} \end{pmatrix} = \begin{pmatrix} \sigma(\xi - \eta) \\ r\xi - \eta - x_0\mu - z_0\xi \\ x_0\eta + y_0\xi - b\mu \end{pmatrix} = \begin{pmatrix} \sigma(\xi - \eta) \\ \xi - \eta - x_0\mu \\ x_0\xi + x_0\eta - b\mu \end{pmatrix} = \begin{pmatrix} \sigma & -\sigma & 0 \\ 1 & -1 & -x_0 \\ x_0 & x_0 & -b \end{pmatrix} \begin{pmatrix} \xi \\ \eta \\ \mu \end{pmatrix}.$$

Характеристический многочлен последней матрицы равен

$$\lambda^3 + (b + 1 - \sigma)\lambda^2 + \lambda b(\sigma + r) - 2\sigma b(r - 1).$$

Проверьте, что решения имеют отрицательную действительную часть, если $1 < r < \sigma \frac{\sigma + b + 3}{\sigma - b - 1}$. $\square$

Для этих значений параметра имеются также два неустойчивых периодических решения, которые сливаются с положениями равновесия $(\sqrt{b(r-1)}, \sqrt{b(r-1)}, r-1)$ и $(-\sqrt{b(r-1)}, -\sqrt{b(r-1)}, r-1)$ соответственно, когда $r = \sigma \frac{\sigma + b + 3}{\sigma - b - 1}$.

13.3.3. Случай $r > \sigma \frac{\sigma + b + 3}{\sigma - b - 1}$. Это наиболее интересный случай. Уже для $r > 1$ начало координат перестает быть устойчивым, потому что характеристический многочлен матрицы линеаризации из уравнения (13.3.2) равен $(b + \lambda)[\lambda^2 + \lambda(\sigma + 1) + \sigma(1 - r)]$ и имеет корни $-b < 0$ и $(1/2)[- \sigma - 1 \pm \sqrt{(\sigma + 1)^2 - 4\sigma(1 - r)}]$, причем оба последних корня отрицательны, только если $r < 1$. Имеются две неустойчивые неподвижные точки, однако неустойчивые периодические решения исчезают. Лоренц дал возможное объяснение отсутствия устойчивых периодических решений. В то же время нетрудно заметить, что решения не могут уходить слишком далеко, а на самом деле даже ограничены.

Теперь становится понятно, что динамика должна усложняться. Орбиты заключены в компактную область, в которой нет точек устойчивого равновесия и, как оказывается, предельных циклов. В силу компактности они должны где-то накапливаться. Лоренц сосредоточил свое внимание на частных значениях параметров ($\sigma = 10, b = 8/3, r = 28 > 10 \frac{47}{19} = 10 \frac{10 + (8/3) + 3}{10 - (8/3) - 1} = \sigma \frac{\sigma + b + 3}{\sigma - b - 1}$) и использовал компьютерные рисунки. Объект, изображенный на рис. 13.3.1, известен как аттрактор Лоренца уже несколько десятилетий. В течение почти сорока лет исследователи пытались преодолеть технические сложности и доказать, что это действительно аттрактор. И вот совсем недавно, когда мы уже работали над этой книгой, доказательство существования аттрактора Лоренца было наконец получено.

13.3.4. Теорема Такера. Доказательство того, что для «классических» значений параметров в системе Лоренца на самом деле имеется аттрактор, основано на сложной комбинации математической теории и строгих вычислений. Уорик Такер разработал алгоритм, позволяющий численно находить точные решения большого класса обыкновенных дифференциальных уравнений. Этот алгоритм основан на процессе деления и использовании интервальной арифметики с направленным округлением.

Метод Такера дает больше чем просто существование аттрактора.

Теорема 13.3.3. *Уравнения Лоренца для $\sigma = 10, b = 8/3, r = 28$ имеют аттрактор. Кроме того, аттрактор устойчив, т. е. он сохраняется при малых изменениях параметров и имеет единственную «SRB-меру» (см. п. 10.4.3.3).*

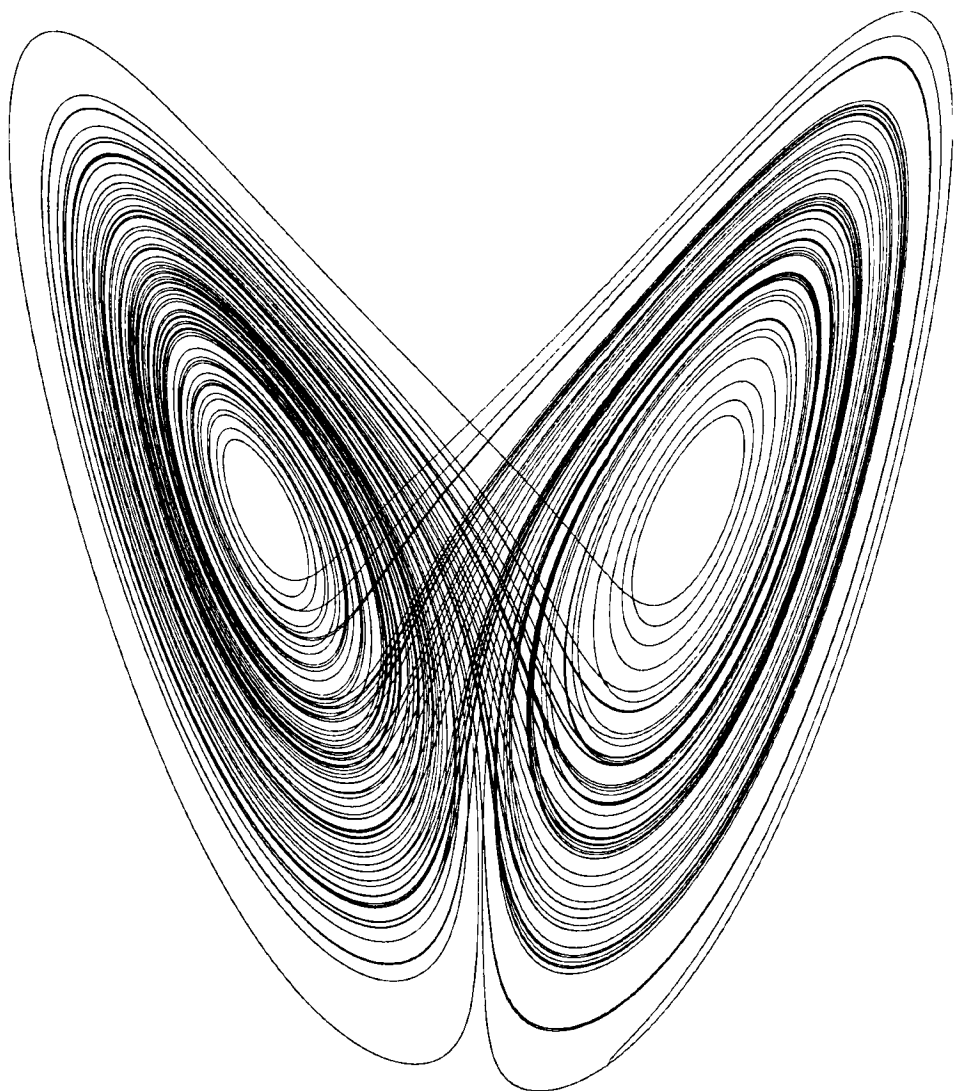


Рис. 13.3.1. Аттрактор Лоренца

По существу это означает, что объект, полученный с помощью численных методов, действительно является аттрактором. Точнее говоря, существует инвариантная мера (см. п. 10.4.3.3), определенная на аттракторе (т. е. мера, которая всем непрерывным функциям, обращающимся в нуль на аттракторе, ставит в соответствие нуль), относительно которой почти все орбиты, начинающиеся в окрестности аттрактора, равномерно распределены: среднее по времени любой непрерывной функции совпадает со *средним по пространству*, определенным инвариантной мерой. Поэтому данная мера называется «физически наблюдае-

мой» плотностью. Два дополнительных утверждения теоремы показывают, что имеется не просто какой-то аттрактор, но он еще физически значимый (устойчивый) и наблюдаемый.

Мы опишем схему доказательства.

*Набросок доказательства*². Первый шаг вполне стандартен и был предложен задолго до того, как было получено полное доказательство. Выберем прямоугольник S в плоскости $z = r - 1$, удовлетворяющий перечисленным условиям, и рассмотрим отображение последования R для потока Лоренца на этом прямоугольнике. Прямоугольник должен содержать точку $(0, 0, r - 1)$, и положения равновесия $(\sqrt{b(r - 1)}, \sqrt{b(r - 1)}, r - 1)$ и $(-\sqrt{b(r - 1)}, -\sqrt{b(r - 1)}, r - 1)$ должны являться серединами двух сторон. Поток Лоренца закручивается вокруг этих точек, пересекая прямоугольник сверху. Отображение R определено на всем прямоугольнике S , кроме прямой Γ , проходящей через точку $(0, 0, r - 1)$ параллельно сторонам через положения равновесия. Точки, лежащие на кривой Γ , притягиваются к началу координат. Вот почему отображение возвращения не определено на Γ , и именно поэтому численные методы не давали успешного решения вопроса о существовании аттрактора: поскольку время возвращения становится произвольно большим вблизи Γ , контролировать ошибки вычислений при решении уравнений Лоренца просто невозможно.

Оказывается, отображение последования на S переводит множество $S \setminus \Gamma$ в две треугольные области в S , каждая из которых пересекает Γ , как показано на рис. 13.3.2. Вершины этих треугольников соответствуют кривой Γ . Поэтому

Рис. 13.3.2. Сечение отображения Лоренца

кривая Γ может разделять орбиты, даже если производная отображения возвращения невелика. Это и есть та «особенность», о которой мы упоминали в начале главы.

²W. Tucker. The Lorenz Attractor Exists // Comptes Rendus des Séances de l'Académie des Sciences. Série I. Mathématique. 1999 V. 32, no. 12. P. 1197—1202; A Rigorous ODE Solver and Smale's 4th Problem // Foundations of Computational Mathematics. 2002. V. 2, no. 1. P. 53—117.

Достигнув этой стадии, исследователи, изучавшие систему Лоренца, обычно делали некоторые предположения о том, каким образом R отображает $S \setminus \Gamma$ в S . Они получили название «геометрического аттрактора Лоренца» и позволили установить существование аттрактора. Однако раньше не было ясно, выполняются ли на самом деле эти условия, т. е. является ли аттрактор, полученный таким образом, аттрактором Лоренца. Здесь уместно заметить, что отсутствие гиперболичности в данной ситуации не слишком существенно. В то время как равномерно гиперболический аттрактор типа соленоида структурно устойчив, для аттрактора Лоренца или геометрического аттрактора Лоренца это не так из-за особенности в начале координат. Однако структурная неустойчивость не слишком велика. В некоторой окрестности имеется более одного класса сопряженности, но классы сопряженности образуют двупараметрическое семейство. Это значительное упрощение по сравнению с общим случаем, когда может существовать бесконечно много модулей сопряженности или, что еще хуже, классы сопряженности могут быть нерегулярными.

Подход Такера служит хорошим примером того, как можно анализировать конкретные странные аттракторы. Он основан на комбинации аналитических и численных методов, и при этом численные методы используются при исследовании большей части аттрактора, а аналитические — при рассмотрении особенности, с которой численные методы не справляются. Это позволяет установить следующие три свойства.

Во-первых, имеется такое компактное множество $N \subset S$, что множество $N \setminus \Gamma$ строго R -инвариантно, т. е. $R(N \setminus \Gamma) \subset \text{Int}(N)$. Это доказывает существование аттрактора $\mathcal{L}$ для потока Лоренца, который пересекает множество S в $\Lambda := \bigcap_{n=0}^{\infty} R^n(N)$.

Чтобы убедиться в том, что аттрактор на самом деле нетривиален, доказываются следующие два результата. Имеется R -инвариантное поле конусов на N , т. е. если $C(x)$ — сектор с углом 10° и центром на кривой, которая приближает множество Λ , тогда $DR(x)C(x) \subset C(R(x))$ (см. определение 10.1.4). Наконец, векторы внутри конусов растягиваются после применения достаточного числа итераций отображения DR . Точнее, имеются такие константы $C > 0$ и $\lambda > \sqrt[29]{2}$, что если v — вектор в $C(x)$ для некоторого $x \in N$, то $\|DR^n(x)v\| \geq C\lambda^n\|v\|$ для всех $n \in \mathbb{N}$. Эти два утверждения относительно поля конусов соответствуют свойствам поля конусов в соленоиде вокруг «горизонтального» направления, которые следуют из гиперболичности отображения. Здесь же из этих двух результатов следует гиперболическое поведение (см. теорему 10.1.5). Такер, кстати, заметил, что положительная константа C в свойстве растяжения конуса не может быть равна 1, т. е. векторы в конусах могут сжиматься на нескольких первых шагах. Это был весьма неожиданный результат.

Доказательство этих свойств основано на двух дополняющих друг друга подходах. Один из них — теория нормальных форм. В малом кубе с центром в начале координат (положение равновесия) имеется почти линейная аналитическая замена координат, которая приводит уравнения Лоренца к виду $\dot{v} = Av + F(v)$, где A — диагональная матрица, а F имеет порядок 20 относительно $\|v\|$. Точнее, на кубе с ребром r выполняется неравенство $\|F\| \leq 7 \cdot 10^{-9} r^{20} / (1 - 3r)$. Это

означает, что на малом кубе новое координатное представление очень близко к линейному. На том же самом кубе нелинейная часть замены координат ограничена константой $r^2/2$.

Замена координат и замечательные оценки были получены методом Пуанкаре. Его суть состоит в том, что мы записываем замену переменных в уравнении в виде (неопределенного) ряда Тейлора для нового координатного представления и замены координат. Если мы хотим устранить нелинейные слагаемые до порядка 20, то коэффициенты Тейлора замены координат можно выписать явно в терминах коэффициентов потока Лоренца, появляющегося с коэффициентами типа $1/(n_1\lambda_1 + n_2\lambda_2 + n_3\lambda_3)$, где λ_i — собственные значения производной потока Лоренца в нуле и $n_i \in \mathbb{N}$. Если значения n_i достаточно велики, для этих дробей имеются явные верхние границы, а оставшиеся случаи можно проверить численными методами (их число конечно, но очень велико).

Вторая составляющая доказательства — строгое численное нахождение орбит. Это делается с использованием интервальной арифметики (по каждой координате). После первого шага интегрирования возникают ошибки, причем оценка сверху для погрешности используемого метода интегрирования известна. Прибавляя эту величину к найденному значению и вычитая ее из найденного значения, мы получим концы отрезка, содержащего истинное значение. На следующем шаге мы интегрируем эти точки, причем меньший результат уменьшается, а больший увеличивается на величину возможной погрешности. Повторяя эту процедуру, мы в конце концов приходим к отрезку (или параллелепипеду), который содержит истинный ответ. Явные значения частных производных потока Лоренца можно получить непосредственно из уравнений Лоренца и продолжить во времени тем же методом. Эта конструкция используется для оценок конуса.

Применяя этот метод, мы вновь сталкиваемся с проблемой, из-за которой завершились неудачей все предыдущие попытки установить существование аттрактора Лоренца: при вычислении орбит, которые приближаются к началу координат, время возвращения становится произвольно большим и погрешности выходят из-под контроля. Но теперь мы можем воспользоваться нормальной формой. Для орбиты, входящей в некоторую фиксированную область вблизи нуля, она позволяет получить точный контроль над тем, как орбита выходит из небольшой области.

Таким образом, мы можем следить за орбитами, используя комбинированный метод: когда они находятся вне некоторой области, лежащей вблизи начала координат, их можно интегрировать с помощью численных методов и интервальной арифметики, а когда они проникают в эту область, мы можем пройти сквозь нее с использованием отображения, заданного нормальной формой. $\square$

ВАРИАЦИОННЫЕ МЕТОДЫ, ЗАКРУЧИВАЮЩИЕ ОТОБРАЖЕНИЯ И ЗАМКНУТЫЕ ГЕОДЕЗИЧЕСКИЕ

В этой главе мы перейдем от гиперболической динамики, которая до сих пор занимала господствующее место в нашем обзоре, к описанию некоторых динамических приложений вариационного подхода. Это один из самых мощных методов, и в то же время его легко представить наглядно. Вариационный подход был впервые упомянут нами в гл. 6 в связи с лагранжевым формализмом и затем в контексте билиардов. Мы начнем с изучения закручивающих отображений, которые включают билиарды в выпуклых областях в виде частного случая и представляют идеальную ситуацию для применения вариационного метода благодаря сравнительно простой структуре фазового пространства. Теория Обри—Мазера устанавливает существование разнообразных видов движения с такими же свойствами орбит, как у одномерных отображений, для всех чисел вращения, согласованных с количеством закручивания в системе. Эти виды движения можно рассматривать как «следы» или «призраки» инвариантных кривых. Таким образом, мы можем наблюдать как остаточные явления, связанные с «порядком», так и наличие хаоса, поскольку в отсутствие настоящих инвариантных кривых взаимодействие между периодическими орбитами и множествами типа канторова производит не только гомоклинические и гетероклинические пересечения, но также сложное поведение более загадочного характера. Поэтому строгое описание типичного поведения выходит далеко за пределы возможностей для имеющихся методов.

Далее мы изложим результаты, требующие меньше предположений относительно динамики и предполагающие более глубокое использование топологии, и закончим описанием одного из самых впечатляющих приложений динамики к геометрии — существования бесконечного числа замкнутых геодезических для любой римановой метрики на двумерной сфере.

§ 14.1. Вариационный метод и биркгофовы периодические орбиты для билиардов

14.1.1. Периодические состояния и функционал действия. В п. 6.4.5 мы построили две орбиты периода 2 специального вида для выпуклого билиарда; для этого мы нашли критические точки функционала, построенного по производящей функции и определенного на пространстве «потенциальных орбит». В этом простом случае «потенциальными орбитами» были *состояния периода 2*, т. е. пары точек границы, а функционалом являлась длина хорды, соединяющей эти две точки. Подобного рода подход является частным случаем вариационного метода; в этом параграфе данный метод используется для нахождения большого числа орбит особого вида, в частности орбит любого заданного периода. Следуя

первоначальному подходу Дж.Д. Биркгофа, мы начнем с описания основных результатов в случае билиардов, где геометрические иллюстрации позволяют сделать изложение наглядным. В следующем параграфе мы представим более общую ситуацию, включающую билиарды в виде частного случая. Это позволит нам объяснить существование периодических орбит на более техническом уровне и обобщить указанный подход за пределы периодического случая.

Пусть числа $p, q \in \mathbb{Z}$ взаимно просты. Без потери общности мы считаем, что $q > 0$ и $1 \leq p \leq q - 1$. Для данного выпуклого билиярда мы ищем периодические орбиты периода q , которые обходят вокруг стола p раз таким образом, что каждая орбита делает в точности p шагов в положительном направлении (против часовой стрелки). Другими словами, билиардное отображение, ограниченное на такую орбиту, должно вести себя как поворот $R_{p/q}$ на угол $2\pi p/q$. Такие орбиты называются *биркгофовыми периодическими орбитами типа (p, q)* .

Мы теперь представим в общих чертах рассуждение, позволяющее построить по крайней мере две различные биркгофовы периодические орбиты типа (p, q) для любых $q > 0$ и $1 \leq p \leq q - 1$. Этим будет показано существование бесконечного числа периодических орбит с произвольно большими периодами. Как и следовало ожидать, пространством $C_{p,q}$ потенциальных орбит будет пространство (p, q) -периодических состояний. Такое состояние можно представлять себе как q -угольник (как правило, самопересекающийся), вписанный в билиардный стол с отмеченной вершиной, так что его ребра соединяют вершины, расположенные через p вершин друг от друга. Отмеченная вершина x_0 соответствует началу отсчета; далее в циклическом порядке идут вершины $x_1, \dots, x_{q-1}$. Продолжим эту последовательность периодически; другими словами, если $0 \leq k \leq q - 1$ и $l \in \mathbb{Z}$, определим $x_{k+lq} = x_k$. Функционал $A_{p,q}$ для нашей вариационной задачи будет общей длиной (периметром) описанного выше многоугольника, а именно многоугольника, ребра которого соединяют x_0 с x_q , x_q с x_{2q} и т. д. до $x_{pq} = x_0$. Мы теперь воспользуемся представлением билиардного отображения через производящую функцию (которая есть расстояние, взятое со знаком «минус»), заданной формулой (6.4.1). Пусть $s_0, s_1, \dots, s_{q-1}$ — значения параметра длины, соответствующие $x_0, x_q, \dots, x_{p(q-1)}$. Тогда

$$A_{p,q}(x_0, x_1, \dots, x_{q-1}) = -(H(s_0, s_1) + H(s_1, s_2) + \dots + H(s_{q-1}, s_0)).$$

Функционал $(-A_{p,q})$ часто называют функционалом действия. Уравнение (6.4.2) показывает, что три последовательные вершины образуют участок орбиты тогда и только тогда, когда частная производная функционала $A_{p,q}$ по переменной, представляющей среднюю вершину, равна нулю; стало быть, критические точки функционала $A_{p,q}$ на пространстве $C_{p,q}$ есть в точности конфигурации, соответствующие биркгофовым периодическим орбитам типа (p, q) .

14.1.2. Существование двух биркгофовых периодических орбит. Нам остается показать, что существуют по крайней мере две критические точки функционала периметра $A_{p,q}$ (см. рис. 14.1.1). В случае орбит периода 2 (т.е. при $p = 1, q = 2$) эти две критические точки соответствуют диаметру и ширине (см. п. 6.4.5). Первая орбита получается из конфигурации, соответствующей мак-

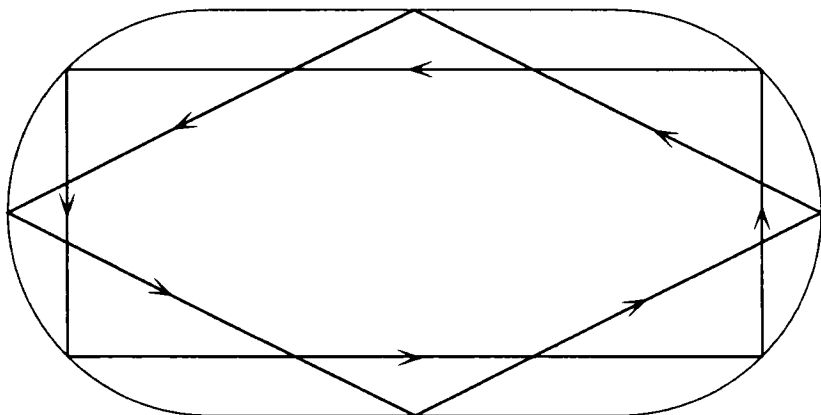


Рис. 14.1.1. Две биркгофовы орбиты

симальному значению функционала $A_{p,q}$. Поскольку пространство $S_{p,q}$ некомпактно, необходимо доказать еще, что это максимальное значение достигается. Это делается стандартным образом: пространство $A_{p,q}$ расширяется до своего естественного замыкания в пространстве последовательностей из q точек, а именно добавляются конфигурации, для которых порядок не является строгим, т. е. несколько последовательных точек могут совпадать. Такое пополненное пространство компактно, функционал $A_{p,q}$ естественным образом продолжается на это пространство и достигает на нем максимального значения. Остается показать, что максимальная длина не может достигаться ни на одной из добавленных вырожденных конфигураций. Это почти очевидно в случае орбит периода 3 и может быть проверено с помощью элементарных рассуждений в случае орбит периода 4, но для более длинных орбит доказательство сложнее, и мы отложим его до следующего параграфа, где данная задача рассматривается в более общем контексте. Идея состоит в том, что любую вырожденную конфигурацию можно «пошевелить» в пополненном пространстве таким образом, что длина соответствующего многоугольника станет больше, а сама конфигурация станет «менее вырожденной», т. е. будет иметь меньше совпадающих вершин. Детали представлены в доказательстве приведенной ниже теоремы 14.2.5.

Найдя максимум для функционала периметра $A_{p,q}$, мы немедленно замечаем, что одна и та же конфигурация дает q различных максимумов, когда мы передвигаем отмеченную точку вдоль орбиты. Это замечание является ключевым для построения второй биркгофовой периодической орбиты типа (p, q) , которое основано на *принципе минимакса* или *перевала*. Название предлагает наглядное представление доказательства: чтобы перейти горный хребет между двумя вершинами с минимальной потерей высоты, необходимо пересечь седло, или перевал. Изменяя знак функционала, который соответствует в данном представлении высоте (и возвращаясь, таким образом, к первоначальной производящей функции в построении функционала), мы получаем версию

для менее смелых альпинистов: чтобы пройти из одной горной долины в другую, поднимаясь при этом на минимальную высоту, необходимо пересечь перевал.

Для билиардов метод перевала работает следующим образом. Пусть $x = (x_0, x_1, \dots, x_{q-1}) \in C_{p,q}$ — конфигурация, на которой функционал $A_{p,q}$ достигает своего максимума. Рассмотрим гладкий путь $x(t) = (x_0(t), x_1(t), \dots, x_{q-1}(t))$, $0 \leq t \leq 1$, в $C_{p,q}$, соединяющий конфигурации x и $x' = (x_1, \dots, x_{q-1}, x_0)$ и обладающий тем свойством, что $x_i(t)$ остается между x_i и x_{i+1} для $i = 0, \dots, q-1$. Функционал $A_{p,q}$ либо имеет постоянное значение на этом пути (и тогда каждая конфигурация $x(t)$, очевидно, порождает свою биркгофovu периодическую орбиту типа (p, q)), либо, что гораздо более вероятно, достигает минимального значения, строго меньшего, чем общее значение в точках x и x' . Простое дифференцирование в этом случае показывает, что если такое значение является максимальным среди значений, принимаемых на всех возможных путях описанного выше типа, то оно должно соответствовать критической точке функционала $A_{p,q}$ (перевалу, см. рис. 14.1.2). И здесь, подобно случаю биркгофовой

Рис. 14.1.2. Перевал

периодической орбиты типа максимум, остается вопрос о существовании такого пути, на котором минимум достигает своего наибольшего значения. Этот вопрос является, несомненно, наиболее тонкой частью всего рассуждения, хотя на интуитивном уровне то, что более выгодно сдвинуть все вершины максимальной конфигурации, чем заморозить некоторые из них, выглядит вполне убедительным. (См. рис. 14.1.3.)

14.1.3. Поднятие на полосу. Альтернативное описание билиардной системы состоит в параметризации границы не простой замкнутой кривой, т.е. не параметром в S^1 , а периодической кривой с периодом 1. В этом случае каждой «физической» точке границы соответствует бесконечное число значений параметра, причем каждые два из них отличаются на целое число. Они называются поднятиями данной точки, и все точки $(x, y) \in S := \mathbb{R} \times (-1, 1)$, которые про-

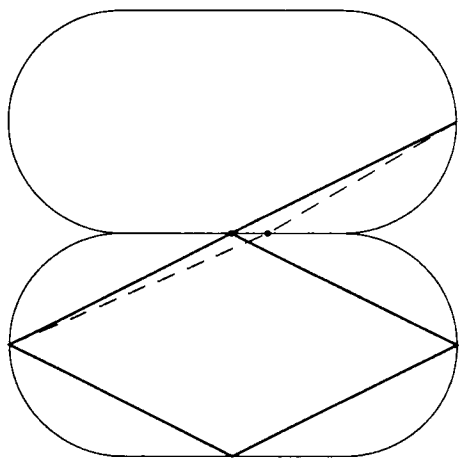


Рис. 14.1.3. Минимаксная орбита

ектируются в данную точку $(s, r) \in C$, называются поднятиями точки (s, r) . Это соответствует в точности «развертыванию» границы, которое является процессом, обратным к описанному в п. 2.6.2 (а конструкция поднятия соответствует предложению 4.3.1). Биллиардное отображение может быть описано однозначно и в этой модели: для данного параметра $x \in \mathbb{R}$ и угла y найдем соответствующий луч на биллиардном столе. Этот луч определяет новую точку и угол (это означает, что мы находим точку (s, y) в C , в которую (x, y) проектируется по модулю 1, и берем ее образ под действием биллиардного отображения ϕ). Для найденной точки выберем наименьшее возможное значение параметра $x' > x$ (т.е. для точки (s', y') , выберем такую точку (x', y') , что $x' = s' \pmod{1}$). Действуя таким образом, мы получим непрерывное отображение (в чем можно убедиться, перейдя к пределу при $y \rightarrow 0$ при фиксированном s). Из этого (периодического по s) нового отображения $\Phi: S \rightarrow S$, которое мы будем называть *поднятием отображения ϕ* , мы можем легко восстановить само отображение ϕ , взяв все граничные значения по модулю 1.

§ 14.2. Биркгофовы периодические орбиты и теория Обри—Мазера для закручивающих отображений

14.2.1. Закручивающие отображения. Любое отображение цилиндра может быть поднято до отображения полосы $\mathbb{R} \times (-1; 1)$ способом, совершенно аналогичным описанному выше для биллиардного отображения. Чтобы различать отображение и его поднятие, мы будем всегда обозначать циклическую координату на цилиндре через s , а первую координату в полосе через x .

Определение 14.2.1. Диффеоморфизм $\phi: C \rightarrow C$ открытого цилиндра $C = S^1 \times (-1; 1)$ называется *закручивающим отображением*, если

1) он сохраняет ориентацию и сохраняет компоненты границы в следующем смысле: существует такое $\varepsilon > 0$, что если $(x, y) \in S^1 \times (-1; \varepsilon - 1)$, то $\phi(x, y) \in S^1 \times (-1; 0)$;

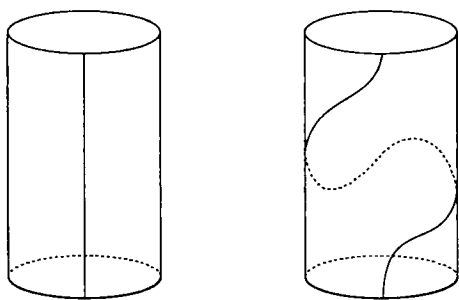


Рис. 14.2.1. Закручивающее отображение

2) $\frac{\partial}{\partial y} \Phi_1(x, y) > 0$, где $\Phi = (\Phi_1, \Phi_2)$ — поднятие отображения ϕ на $S = \mathbb{R} \times (0; 1)$ (см. рис. 14.2.1);

3) отображение ϕ продолжается до гомеоморфизма $\bar{\phi}$ (необязательно гладкого) замкнутого цилиндра $S^1 \times [-1; 1]$.

Отображение ϕ называется *дифференцируемым закручивающим отображением*, если для любого $\epsilon > 0$ существует такое $\delta > 0$, что $\frac{\partial}{\partial y} \Phi_1(x, y) > \delta$ на множестве $C_\epsilon := S^1 \times [\epsilon - 1; 1 - \epsilon]$.

Последнее условие в определении не является существенным. Однако оно упрощает некоторые рассуждения, например, с его помощью мы ниже определим производящую функцию. Кроме того, оно помогает численно выразить «количество вращения», соответствующее данному закручивающему отображению. Ограничение гомеоморфизма $\bar{\phi}$ на «нижнюю» окружность $S^1 \times \{-1\}$ имеет число вращения, определенное с точностью до прибавления целого числа (см. определение 4.3.6). Если мы зафиксируем поднятие этого ограничения с числом вращения ρ_- , это определит (единственное) поднятие отображения $\bar{\phi}$; ограничение этого поднятия на «верхнюю» окружность $S^1 \times \{1\}$ имеет однозначно определенное число вращения ρ_+ . При другом выборе поднятия отрезок $[\rho_-, \rho_+]$ изменится, но это изменение есть всего лишь сдвиг на целое число. Мы называем этот отрезок *отрезком закручивания* закручивающего отображения ϕ . Подобным образом можно определить понятие закручивающего отображения для любого гомеоморфизма замкнутого цилиндра, сохраняющего компоненты границы.

Из элементарных геометрических рассуждений, соотношения (6.4.4) и предложения 6.4.2 следует такой результат.

Предложение 14.2.2. *Бильярдное отображение $\phi: C \rightarrow C$ открытого цилиндра $C = S^1 \times (-1, 1)$ является сохраняющим площадь дифференцируемым закручивающим отображением, удовлетворяющим следующему дополнительному условию: для любого его поднятия Φ выполняются предельные соотношения $\Phi_1(x, y) \xrightarrow{y \rightarrow -1} x$ и $\Phi_1(x, y) \xrightarrow{y \rightarrow 1} x + 1$. Отсюда следует, что отрезком закручивания для любого бильярдного отображения является отрезок $[0; 1]$.*

14.2.2. Производящая функция закручивающего отображения. Сохраняющие площадь закручивающие отображения разделяют наиболее важные свой-

ства билиардных отображений. Сначала мы покажем, что любое сохраняющее площадь дифференцируемое скручивающее отображение можно представить с помощью гладкой производящей функции вида (6.4.3). Чтобы избежать возможных наложений областей цилиндра и подсчета площади с кратностью, мы опишем производящую функцию для поднятия.

Пусть $\Phi(x, y) = (x', y')$. Зафиксируем x и x' и рассмотрим «треугольник», образованный вертикальным сегментом с абсциссой x' , образом относительно Φ вертикального сегмента с абсциссой x и сегментом нижней границы, соединяющим нижние точки этих двух кривых. Пусть $H(x, x')$ — площадь такой фигуры (см. рис. 14.2.2). Тогда

$$\frac{\partial}{\partial s'} H(x, x') = y'.$$

Применяя отображение Φ^{-1} и используя сохранение площади, получаем

$$\frac{\partial}{\partial s} H(x, x') = -y.$$

Хотя определение производящей функции и не требует выполнения условия дифференцируемости закручивающего отображения, последнее необходимо для существования вторых производных $\frac{\partial^2 H}{\partial s^2}$ и $\frac{\partial^2 H}{\partial s'^2}$. С другой стороны, определение закручивающего отображения гарантирует, что смешанная производная $\frac{\partial^2 H}{\partial s \partial s'} = -\frac{\partial y}{\partial s'}$ существует и неположительна; она отрицательна для дифференцируемого закручивающего отображения.

Рис. 14.2.2. Производящая функция

Очевидно, что производящая функция корректно определена с точностью до прибавления константы. Другой способ ее определения заключается в следующем. Из условия закручивания следует, что y и y' однозначно определены значениями x и x' . Из условия дифференцируемого закручивания следует, что если функции $y(x, x')$ и $y'(x, x')$ определены для пары значений (x, x') , то они определены и дифференцируемы в окрестности этой пары. Чтобы найти функцию H локально, необходимо проверить известное условие точности $\partial y / \partial x' = \partial y' / \partial x$. Несложный подсчет показывает, что точность эквивалентна

сохранению площади. Отсюда следует, что производящая функция H определена локально с точностью до аддитивной константы и может быть продолжена на все допустимые пары (x, x') путем склейки локальных данных с помощью согласования констант. Величина $H(x + 1, x' + 1) - H(x, x')$ предположительно является константой, и, поскольку Φ сохраняет полосу, эта константа равна нулю.

Основные качественные характеристики, присущие билиардным отображениям, свойственны и закручивающим отображениям. Очевидное преимущество использования последнего понятия состоит в том, что оно покрывает многие другие важные случаи, такие как периодически вынужденные колебания, окрестности большинства эллиптических точек сохраняющих площадь диффеоморфизмов, малые возмущения гамильтоновых систем с двумя степенями свободы и внешние билиарды. Внешний билиард — это отображение, определенное на точках вне данной выпуклой кривой и действующее следующим образом. Для данной точки проведем касательную к нашей кривой и переместимся в точку, находящуюся на этой прямой на том же расстоянии от точки касания, что и исходная точка (см. рис. 14.2.3). С точки зрения внутренней геометрии разница между прямым

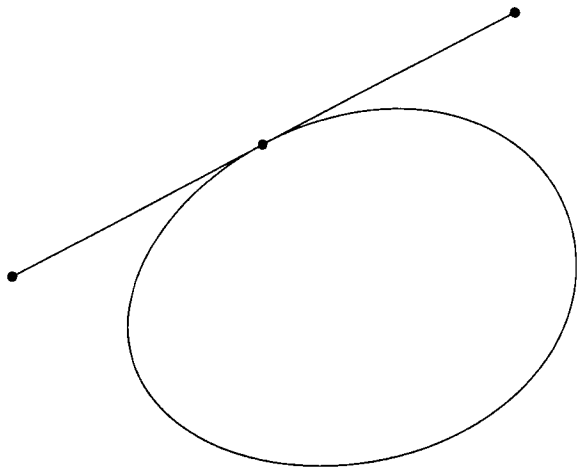


Рис. 14.2.3. Внешний билиард

рассмотрением билиардных отображений и рассмотрением их в качестве закручивающих отображений подобна разнице между лагранжевым и гамильтоновым формализмом в классической механике и на самом деле представляет собой частный случай этой дуальности, точнее ее версии, соответствующей дискретному времени. Вообще говоря, гамильтонов подход делает динамическую (в смысле, используемом в данной книге) природу задачи более явной путем рассмотрения динамической системы в фазовом пространстве, без непосредственного различения координаты и импульса. Лагранжев подход, который разделяет конфигурационное пространство и делит фазовое пространство на координаты и импульсы (скорости), иногда оказывается полезным, поскольку обеспечивает геометрическую наглядность методов и результатов.

14.2.3. Биркгофовы периодические орбиты. Мы рассмотрели биркгофовы периодические орбиты для билиардов в предыдущем параграфе. Здесь мы рассмотрим их в контексте закручивающих отображений.

Определение 14.2.3. Для данного закручивающего отображения ϕ и его поднятия Φ точка $w \in S$ называется *биркгофовой периодической точкой типа (p, q)* , а ее орбита — *биркгофовой периодической орбитой типа (p, q)* , если для любого поднятия $z \in S$ точки w существует такая последовательность $((x_n, y_n))_{n \in \mathbb{Z}}$ точек окружности S , что

- 1) $(x_0, y_0) = z$,
- 2) $x_{n+1} > x_n$ ($n \in \mathbb{N}$),
- 3) $(x_{n+q}, y_{n+q}) = (x_n + 1, y_n)$,
- 4) $(x_{n+p}, y_{n+p}) = \Phi(x_n, y_n)$.

Замечание 14.2.4. Последовательность (x_n, y_n) параметризует орбиту не в соответствии с «динамическим упорядочением», индуцированным переходом от (x, y) к $\Phi(x, y)$, а в соответствии с «геометрическим упорядочением» ее проекции на S^1 (см. п. 14.1.1). Этот порядок совпадает с порядком расположения образов при рациональном повороте $R_{p/q}$ на окружности. Более того, проекция биркгофовой периодической орбиты типа (p, q) на окружность есть конечное множество, и отображение, индуцированное проектированием Φ , может быть продолжено кусочно линейно до гомеоморфизма окружности.

Теперь, как было обещано в п. 14.1.1, мы представим более детальное доказательство существования первой (минимизирующей действие) биркгофовой периодической орбиты, включая объяснение того, почему минимум достигается внутри пространства (p, q) -состояний.

Теорема 14.2.5. Пусть $\phi: S \rightarrow S$ — дифференцируемое закручивающее отображение. Если числа $p, q \in \mathbb{N}$ взаимно просты и число p/q лежит в отрезке закручивания отображения ϕ , то у отображения ϕ существует биркгофова периодическая орбита типа (p, q) .

Доказательство. Возьмем такое поднятие Φ отображения ϕ , что число p/q лежит в отрезке закручивания этого поднятия. Обозначим ограничения поднятия Φ на «нижнюю» границу $\mathbb{R}^1 \times \{-1\}$ и «верхнюю» границу $\mathbb{R}^1 \times \{1\}$ через Φ_- и Φ_+ соответственно. Мы получим биркгофову периодическую орбиту, найдя последовательность ее абсцисс, на которой достигается глобальный минимум соответствующего действия, определенного на пространстве последовательностей действительных чисел. В качестве пространства, элементы которого могут служить абсциссами орбиты, рассмотрим следующее пространство Σ . Во-первых, пусть $\tilde{\Sigma}$ — множество таких невозрастающих последовательностей $(x_n)_{n \in \mathbb{Z}}$ действительных чисел, что

$$x_{n+q} = x_n + 1 \quad (14.2.1)$$

и

$$\Phi(x_n \times [\varepsilon - 1, 1 - \varepsilon]) \cap (x_{n+p} \times [\varepsilon - 1, 1 - \varepsilon]) \neq \emptyset, \quad (14.2.2)$$

где $\varepsilon > 0$ выбрано следующим образом. Поскольку число p/q принадлежит отрезку закручивания отображения Φ , существует такое $\delta \in (0, 1)$, что если $x_{k+1} \leq \Phi_-(x_k) + \delta$ ($k = 0, 1, \dots, q-1$), то $x_q < x_0 + p$, и аналогично, если $x_{k+1} \geq \Phi_+(x_k) - \delta$ ($k = 0, 1, \dots, q-1$), то $x_q > x_0 + p$. Возьмем такое $\varepsilon > 0$, что

$$\bigcup_{i=0}^{q-1} \Phi^i(\mathbb{R} \times ((-1, \varepsilon - 1] \cup [1 - \varepsilon, 1))) \subset \mathbb{R} \times ((-1, \delta - 1] \cup [1 - \delta, 1)).$$

Мы называем такие последовательности *упорядоченными состояниями типа* (p, q) .

Отсюда следует, что для любой орбиты в S , абсциссы которой удовлетворяют соотношениям (14.2.1) и (14.2.2), ее ординаты лежат в интервале $(\varepsilon - 1, 1 - \varepsilon)$. Определим отношение эквивалентности $\sim$ на $\tilde{\Sigma}$, положив $x \sim x'$, если $x_i - x'_i = k$ для всех i и некоторого фиксированного $k \in \mathbb{Z}$. Пусть $\Sigma := \tilde{\Sigma}/\sim$ — множество классов эквивалентности.

Условие (14.2.1) равносильно периодичности, и, поскольку число p/q принадлежит отрезку закручивания отображения Φ , условие (14.2.2) гарантирует существование такой точки (x_n, y_n) , что $\Phi(x_n, y_n) = (x_{n+p}, y_{n+p})$ для некоторого y_{n+p} . Последовательность, удовлетворяющая соотношениям (14.2.1) и (14.2.2), обычно не является проекцией орбиты на первую координату, но мы найдем последовательность, которая окажется таковой, и соответствующая орбита будет искомой биркгофовой периодической орбитой типа (p, q) .

Каждая последовательность имеет всего лишь q «независимых» переменных $x_0, \dots, x_{q-1}$, как следует, скажем, из соотношения (14.2.1), т. е. множество $\tilde{\Sigma}$ естественно вложено в $\mathbb{R}^q$. Условие (14.2.2), примененное индуктивно, показывает, что множество $\{x_n - x_0\}_{n=0}^{q-1}$ ограничено для любого $x \in \tilde{\Sigma}$, поэтому Σ есть замкнутое и ограниченное, а стало быть, компактное подмножество в $\mathbb{R}^q/\mathbb{Z} \sim \mathbb{R}^{q-1} \times S^1$.

Определим на Σ функционал действия

$$L(s) := \sum_{n=0}^{q-1} H(x_n, x_{n+p}),$$

где H — производящая функция. Поскольку числа p и q взаимно просты, из (14.2.1) следует, что $L(s) = \sum_{n=0}^{q-1} H(x_j, x_{j+np})$ для любого $j \in \mathbb{Z}$. Поскольку функция L инвариантна относительно сдвигов на целое число, она корректно определена на компактном множестве Σ и потому достигает своего максимума и минимума, но они могут достигаться на границе. Мы покажем, что минимум соответствует биркгофовой периодической орбите типа (p, q) , и выведем отсюда, что он достигается не на границе.

Рассмотрим произвольную последовательность $x \in \Sigma$. Она не постоянна в соответствии с (14.2.1). Поэтому для любого $m \in \mathbb{Z}$ существуют такие $n \in \mathbb{Z}$ и $k \geq 0$, что $n \leq m \leq n + k$ и $x_{n-1} < x_n = \dots = x_{n+k} < x_{n+k+1}$. (Если $k > 0$, то x — гранич-

ная точка множества $\tilde{\Sigma}$.) Определим $h_1(x, x')$ и $h_2(x, x')$ из условия

$$\Phi(x, h_2(x, x')) = (x', h_1(x, x')). \quad (14.2.3)$$

Поскольку последовательность x неубывающая, из условия закручивания (определение 14.2.1, 14.2.1)) следует, что

$$\varepsilon - 1 \leq h_1(x_{n+k-p}, x_{n+k}) \leq \dots \leq h_1(x_{n-p}, x_n) \leq 1 - \varepsilon,$$

$$\varepsilon - 1 \leq h_2(x_n, x_{n+p}) \leq \dots \leq h_2(x_{n+k}, x_{n+k+p}) \leq 1 - \varepsilon,$$

так что либо

$$h_2(x_n, x_{n+p}) < h_1(x_{n-p}, x_n), \quad (14.2.4)$$

либо

$$h_1(x_{n+k-p}, x_{n+k}) < h_2(x_{n+k}, x_{n+k+p}), \quad (14.2.5)$$

либо

$$h_1(x_{n+l-p}, x_{n+l}) = h_2(x_{n+l}, x_{n+l+p}) \text{ для } l \in \{0, \dots, k\}. \quad (14.2.6)$$

В случае (14.2.4) заметим, что, рассматривая $s = x_n$ как независимую переменную и сохраняя все остальные значения x_i фиксированными, мы получаем

$$\begin{aligned} \frac{d}{ds} \Big|_{s=x_n} L(x) &= \frac{d}{ds} \Big|_{s=x_n} \sum_{i=0}^{q-1} H(x_i, x_{i+p}) = \frac{d}{ds} \Big|_{s=x_n} (H(x_{n-p}, s) + H(s, x_{n+p})) = \\ &= h_1(x_{n-p}, x_n) - h_2(x_n, x_{n+p}) > 0 \end{aligned}$$

и что в соответствии с (14.2.4) мы можем, не покидая множества Σ , немного уменьшить x_n , а значит, также и $L(x)$, так что x не может быть минимумом. В случае (14.2.5) аналогично, полагая $s = x_{n+k}$, мы получим, что $\frac{d}{ds} \Big|_{s=x_{n+k}} L(x) < 0$, и в соответствии с (14.2.5) мы можем, не покидая множества Σ , немного увеличить x_{n+k} , а значит, уменьшить $L(x)$, так что x опять не является минимумом. Таким образом, если $x = (x_m)_{m \in \mathbb{Z}}$ — минимум, то для любого $m \in \mathbb{Z}$ наш анализ приводит к равенству (14.2.6). Поэтому

$$h_1(x_{m-p}, x_m) = h_2(x_m, x_{m+p}) \text{ для всех } m \in \mathbb{Z}. \quad (14.2.7)$$

Положив $(s_n, y_n) = (x_n, h_1(x_{n-p}, x_n))$, мы получим периодическую орбиту.

Далее, $y_n \in (\varepsilon - 1, 1 - \varepsilon)$ для любого $n \in \mathbb{Z}$, поскольку из неравенства $y_n \leq \varepsilon - 1$, выполненного для любого $n \in \mathbb{Z}$, следует, что $y_n < \delta$ для всех $n \in \mathbb{Z}$, что противоречит соотношениям (14.2.1), (14.2.2) и выбору δ . Поэтому для того чтобы показать, что (x_n, y_n) — биркгофова периодическая орбита типа (p, q) и s не находится на границе множества Σ , достаточно доказать, что последовательность $s_n = x_n$ строго возрастает.

Предположим, что $s_n = s_{n+1}$. Поскольку s не есть постоянная последовательность, выбрав при необходимости другое n , мы можем считать, что либо $s_{n-1} < s_n$, либо $s_n < s_{n+1}$. Далее, поскольку последовательность s не убывает, из условия закручивания и из соотношения (14.2.7) следует, что $y_{n+1} = h_1(s_{n-p+1}, s_{n+1}) \leq h_1(s_{n-p}, s_{n+1}) \leq h_1(s_{n-p}, s_n) = y_n = h_2(s_n, s_{n+p}) \leq h_2(s_n, s_{n+p+1}) = y_{n+1}$, причем по крайней мере одно неравенство строгое, что невозможно.

Итак, мы нашли такую биркгофову периодическую орбиту типа (p, q) , что последовательность ее абсцисс является глобальным минимумом функционала L на внутренности множества Σ . $\square$

При построении второй биркгофовой периодической орбиты (типа минимакс) используется версия принципа перевала, примененная к тому же самому функционалу, ограниченному на пространство состояний, которые остаются на своем протяжении между состоянием, определенным максимальной биркгофовой периодической орбитой, и его сдвигом. Однако для того чтобы приступить к изучению неперiodических орбит в оставшейся части главы, нам важно знать лишь о существовании одной биркгофовой периодической орбиты.

14.2.4. Упорядоченные орбиты. В этом пункте мы доказываем, что любая сохраняющая порядок орбита закручивающего отображения представляет собой часть графика липшицевой функции с константой липшицевости, ограниченной на любом замкнутом кольце, содержащемся в S . Как и в п. 14.2.1, мы часто работаем с поднятиями.

Определение 14.2.6. (Ср. определение 14.2.3.) Рассмотрим закручивающее отображение $\phi: C \rightarrow C$. Сегмент орбиты (или целая орбита) $\{(x_m, y_m), \dots, (x_n, y_n)\}$ отображения ϕ , $-\infty \leq m < n \leq \infty$, который может быть бесконечным в одну или в обе стороны, называется *упорядоченным* или *сохраняющим порядок*, если $x_i \neq x_j$ для $i \neq j$ и $(i, j) \neq (n, m)$ и ϕ сохраняет циклический порядок абсцисс, т. е. если абсциссы x_i, x_j, x_k , где $i, j, k < n$, положительно ориентированы (относительно выбранной ориентации на S^1), то $x_{i+1}, x_{j+1}, x_{k+1}$ также положительно ориентированы.

Лемма 14.2.7. Пусть $\Phi: \mathbb{R} \times (-1, 1) \rightarrow \mathbb{R} \times (-1, 1)$ — поднятие закручивающего диффеоморфизма $\phi: C \rightarrow C$ (не обязательно сохраняющего площадь). Если $(x_i, y_i) = F^i(x_0, y_0)$, $(x'_i, y'_i) = F^i(x'_0, y'_0)$ и $x'_i > x_i$ для $i = -1, 0, 1$, то существует такое число $M \in \mathbb{R}$, что $|y'_0 - y_0| < M|x'_0 - x_0|$. Константа M может быть выбрана равномерно на любом замкнутом кольце в C .

Доказательство. Предположим, что $y'_0 < y_0$. Если обозначить $(\tilde{x}, \tilde{y}) = \Phi(x'_0, y_0)$, то из условия закручивания следует, что

$$\tilde{x} > x'_1 + c(y_0 - y'_0),$$

где константа c отделена от 0 на любом замкнутом кольце в C . С другой стороны, из дифференцируемости отображения ϕ следует существование такой константы L (ограниченной на компактных кольцах в C), что

$$x'_1 > x_1 > \tilde{x} - L(x'_0 - x_0).$$

Положив $M = Lc^{-1}$, мы докажем утверждение леммы. В случае если $y'_0 > y_0$, необходимо повторить проведенное выше рассуждение, заменив ϕ на ϕ^{-1} . $\square$

Следствие 14.2.8. Рассмотрим сохраняющее площадь закручивающее отображение $\phi: C \rightarrow C$ и сохраняющий ориентацию сегмент орбиты этого отображения $\{(x_m, y_m), \dots, (x_n, y_n)\}$, $-\infty \leq m < n \leq \infty$, который содер-

жить в замкнутом кольце, принадлежащем множеству C . Тогда $|y_i - y_j| < M|x_i - x_j|$ для всех таких i, j , что $m < i, j < n$.

Доказательство. Достаточно применить лемму 14.2.7 к тройкам $(i-1, i, i+1)$ и $(j-1, j, j+1)$. $\square$

Данное следствие показывает, что замыкание E упорядоченной орбиты является подмножеством графика некоторой липшицевой функции $\varphi: S^1 \rightarrow (-1, 1)$. Заметим, что $\varphi|_E$ проектируется в гомеоморфизм образа множества E при проекции на S^1 ; последний гомеоморфизм можно доопределить по линейности на промежутках, не входящих в это множество, получив таким образом гомеоморфизм окружности. Мы можем, стало быть, определить *число вращения* упорядоченной орбиты как число вращения этого индуцированного гомеоморфизма окружности. Поэтому внутренняя динамика упорядоченной орбиты закручивающего отображения кольца является по существу одномерной. Мы увидим, что для любого числа вращения из отрезка закручивания одномерная динамика с этим числом вращения представлена в закручивающем отображении.

14.2.5. Множества Обри—Мазера. Наша следующая цель состоит в том, чтобы показать, что любое иррациональное число из отрезка закручивания является числом вращения некоторой упорядоченной орбиты. Более того, мы увидим, что такие орбиты не являются исключительными, подобно биркгофовым периодическим орбитам; для любого числа вращения имеется много таких орбит. Их можно построить с помощью весьма усложненной версии вариационного подхода, примененной к соответствующим образом определенному бесконечномерному пространству. Это сложный, но мощный метод; его дальнейшее развитие позволяет получить много новых результатов об упорядоченных орбитах и о более сложных видах орбит. Замечательно, однако, то, что из довольно простых рассуждений, использующих непрерывность, вытекает, что упорядоченная периодическая орбита с иррациональным числом вращения является пределом биркгофовых периодических орбит. Поэтому при доказательстве всех результатов из этого пункта (за исключением теоремы 14.2.15) не используется непосредственно сохранение площади, а достаточно знать только о существовании биркгофовых периодических орбит (что можно установить при более слабых предположениях).

Определение 14.2.9. Пусть $\phi: C \rightarrow C$ — закручивающее отображение. Замкнутое инвариантное множество $E \subset C$ называется *упорядоченным множеством*, если оно взаимно однозначно проектируется на подмножество окружности и ϕ сохраняет циклический порядок на E . *Множеством Обри—Мазера* называется минимальное упорядоченное инвариантное множество, взаимно однозначно проектирующееся на канторово множество в S^1 .

Любая орбита упорядоченного множества есть упорядоченное множество. Дополнение к образу множества Обри—Мазера при проекции является объединением счетного числа интервалов на окружности. Мы называем эти интервалы *разрывами* множества Обри—Мазера. Концевые точки каждого из таких интервалов есть проекции точек множества Обри—Мазера; последние также

называются *концевыми точками*. Из следствия 14.2.8 немедленно выводится такой результат.

Следствие 14.2.10. Пусть $\phi: C \rightarrow C$ — закручивающий диффеоморфизм и A — множество Обри—Мазера для ϕ . Тогда существует липшицева функция $\varphi: S^1 \rightarrow (-1, 1)$, график которой содержит множество A .

Доказательство. Как мы знаем из следствия 14.2.8, такая функция определена на образе множества A при проекции на S^1 . Продолжая ее линейно на разрывах соответствующего канторова множества, мы получим функцию на S^1 с той же самой константой Липшица. $\square$

Определим *число вращения* множества Обри—Мазера или инвариантной окружности как число вращения любой из орбит, содержащейся в данном множестве; последнее определено в конце п. 14.2.4. Теперь мы можем доказать один из основных результатов теории закручивающих отображений.

Теорема 14.2.11. Пусть $\phi: C \rightarrow C$ — сохраняющее площадь дифференцируемое закручивающее отображение. Тогда для любого иррационального числа α из отрезка закручивания для ϕ существует множество Обри—Мазера A с числом вращения α или инвариантная окружность с числом вращения α , представленная графиком липшицевой функции φ .

Доказательство. Пусть p_n/q_n — последовательность несократимых рациональных чисел, аппроксимирующая число α . Воспользуемся теоремой 14.2.5 и возьмем любую последовательность ω_n биркгофовых периодических орбит типа (p_n, q_n) . В соответствии со следствием 14.2.8 мы можем построить липшицеву функцию $\varphi_n: S^1 \rightarrow (-1, 1)$, график которой содержит ω_n . С помощью рассуждения, подобному примененному при доказательстве соотношения (14.2.2), мы замечаем, что все эти орбиты содержатся в замкнутом кольце внутри C , так что константа Липшица может быть выбрана независимой от n . Используя предкомпактность этого равномерно непрерывного семейства функций (теорема Арцела—Асколи), мы можем без потери общности предположить, что эти функции сходятся к липшицевой функции φ . График функции φ может не быть ϕ -инвариантным, но он всегда содержит замкнутое ϕ -инвариантное множество A , которое можно получить следующим образом. Область определения функции φ_n содержит образы биркгофовых периодических орбит типа (p_n, q_n) при проекции на S^1 . Эти биркгофовы периодические орбиты типа (p_n, q_n) являются замкнутыми ϕ -инвариантными подмножествами множества C , и поэтому в топологии, заданной метрикой Хаусдорфа (см. определение А.1.28), они имеют точку накопления $A \subset C$. Это множество A , очевидно, принадлежит графику функции φ и является ϕ -инвариантным по лемме А.1.27. Более того, отображение ϕ сохраняет циклический порядок на A (поскольку это верно для биркгофовых периодических орбит ω_n , а сохранение порядка — замкнутое свойство). Если мы обозначим через φ_n продолжение на S^1 проекции отображения ϕ , ограниченного на биркгофову периодическую орбиту типа (p, q) , а через φ_α продолжение проекции отображения $\phi|_A$ на S^1 , то $\varphi_n \rightarrow \varphi_\alpha$ равномерно. В силу непрерывности числа вращения в C^0 -топологии (предложение 4.4.5) числом вращения множества A является α .

Рассмотрим теперь минимальное множество отображения ϕ_α . По дихотомии из предложения 4.3.19 оно является или целой окружностью, или инвариантным канторовым множеством. В последнем случае образ этого канторова множества под действием отображения $\text{Id} \times \phi$ есть искомое множество Обри—Мазера с числом вращения α . $\square$

Замечание 14.2.12. Множество Обри—Мазера, полученное в теореме 14.2.11, может оказаться подмножеством окружности, инвариантной относительно ϕ . Однако если и отображение, и инвариантная окружность принадлежат классу C^2 , то ограничение этого отображения на окружность есть C^2 -диффеоморфизм окружности, который по теореме Данжуа (см. п. 4.4.3) является транзитивным. Поэтому для того чтобы множество Обри—Мазера лежало на инвариантной окружности, либо отображение, либо окружность (а возможно, и то и другое) должны иметь порядок гладкости меньше чем 2. Мишель Эрман придумал замечательное построение, с помощью которого ему удалось вложить примеры типа Данжуа нетранзитивных $C^{2-\epsilon}$ -диффеоморфизмов окружности ($\epsilon > 0$) в пространство сохраняющих ориентацию дифференцируемых закручивающих отображений класса $C^{3-\epsilon}$, таким образом получив дополнительную производную по сравнению с очевидной конструкцией. До сих пор неизвестно, может ли C^3 -диффеоморфизм иметь инвариантную окружность, содержащую множество Обри—Мазера.

Хаусдорфов предел биркгофовых периодических орбит типа (p_n, q_n) может оказаться обширнее множества Обри—Мазера, хотя он всегда является сохраняющим порядок множеством. Если это множество не минимально, то оно содержит множество орбит, гомоклинических к множеству Обри—Мазера. Перейдя к хаусдорфову пределу биркгофовых периодических орбит типа минимакс и используя некоторые аккуратные вариационные оценки, можно доказать, что такие орбиты всегда существуют.

Заменяя биркгофовы периодические орбиты w_n в предыдущем рассуждении произвольными упорядоченными инвариантными множествами, сходящимися в метрике Хаусдорфа, мы получаем следующее предложение.

Предложение 14.2.13. *Число вращения упорядоченного инвариантного множества непрерывно в топологии метрики Хаусдорфа.*

Из этого, в свою очередь, следует такой результат.

Следствие 14.2.14. *Число вращения упорядоченной орбиты непрерывно зависит от начального значения.*

Доказательство. Пусть $x_n \rightarrow x$ — сходящаяся последовательность точек с упорядоченными орбитами. Без потери общности мы можем предполагать, что числа вращения α_n орбит точек x_n сходятся. Рассмотрим набор орбит, соответствующих x_n . В силу компактности топологии, задаваемой метрикой Хаусдорфа (см. лемму А.1.26), он содержит подпоследовательность, сходящуюся к упорядоченному множеству, содержащему замыкание орбиты точки x . Отсюда по

предложению 14.2.13 предел чисел вращения орбит точек x_n равен числу вращения орбиты точки x . $\square$

Теперь мы можем показать, что для любого иррационального числа существует не более одной инвариантной окружности с заданным числом вращения.

Теорема 14.2.15. Пусть $\phi: C \rightarrow C$ — сохраняющее площадь закручивающее отображение и α — иррациональное число, принадлежащее отрезку закручивания. Тогда отображение ϕ имеет не более одной инвариантной окружности с числом вращения α , являющейся графиком некоторой функции ϕ . Если такая инвариантная окружность существует, то не существует множества Обри—Мазера с числом вращения α вне этой окружности; следовательно, существует не более одного множества Обри—Мазера.

Замечание 14.2.16. Известно, что закручивающее отображение может иметь несколько инвариантных окружностей с одним и тем же рациональным числом вращения. Это так, например, в случае эллиптического бильярда (см. рис. 6.3.9), где две ветви гетероклинических циклов образуют пару инвариантных окружностей с числом вращения $1/2$. Подобные явления с нулевым числом вращения возникают для отображения за время t (при малых t) математического маятника (см. п. 6.2.2).

Лемма 14.2.17. Предположим, что отображение ϕ имеет инвариантную окружность R , являющуюся графиком функции ϕ , с числом вращения α . Тогда любая сохраняющая порядок орбита, замыкание которой не пересекает R , имеет число вращения, отличное от α .

Доказательство. Окружность R делит кольцо C на верхнюю и нижнюю компоненты. Предположим, что x — точка верхней компоненты множества $C \setminus R$, орбита которой сохраняет порядок и отделена от R . Тогда отображение ϕ , ограниченное на орбиту точки x , проектируется в отображение подмножества E окружности S^1 . Мы хотим продолжить его до отображения ϕ_2 окружности S^1 , которое строго больше (в смысле определения 4.4.6) отображения ϕ_1 , индуцированного отображением $\phi|_R$, т. е. $\phi_1 < \phi_2$. Это соотношение уже выполнено на E , так что мы должны позаботиться только об аккуратном продолжении на дополнение $S^1 \setminus E$. Продолжение на замыкание множества E не меняет строгого неравенства благодаря условию закручивания и предположению, что орбита отделена от R . Чтобы определить ϕ_2 на интервалах, дополнительных к $\bar{E}$, обозначим концевые точки такого интервала через x_1 и x_2 ; пусть $\delta := \min\{\phi_2(x_1) - \phi_1(x_1), \phi_2(x_2) - \phi_1(x_2)\}$. Положим $\phi_2(tx_1 + (1-t)x_2) = \max(t\phi_2(x_1) + (1-t)\phi_2(x_2), \delta + \phi_1(tx_1 + (1-t)x_2))$. Тогда отображение ϕ_2 монотонно и $\phi_1 < \phi_2$. Следовательно, в соответствии с предположением 4.4.9 число вращения отображения ϕ_2 больше чем α . Аналогично не может существовать сохраняющей порядок орбиты с числом вращения α в нижней компоненте множества $C \setminus R$. $\square$

Доказательство теоремы 14.2.15. Предположим, что существуют две инвариантные окружности с числом вращения α . Их пересечение также инвариант-

но, так что если по крайней мере одна из них транзитивна, то они не пересекаются, что невозможно по предыдущей лемме. В противном случае пересечение содержит общее для двух окружностей множество Обри—Мазера A и окружности представляют собой графики двух различных функций φ_1 и φ_2 , совпадающих на проекции множества A . Графики функций $\max(\varphi_1, \varphi_2)$ и $\min(\varphi_1, \varphi_2)$ являются инвариантными, и потому таким же является пространство между этими графиками. Но это пространство должно состоять из бесконечного числа компонент, поскольку оно проектируется на непериодические интервалы, дополнительные к множеству Обри—Мазера. Таким образом, мы получаем открытый диск с попарно непересекающимися образами, что невозможно из-за сохранения площади (см. теорему Пуанкаре о возвращении 6.1.6). Здесь мы используем то обстоятельство, что число вращения иррационально; в противном случае возможно существование конечного числа компонент, переставляемых отображением ϕ .

Из леммы также следует, что не существует множества Обри—Мазера с числом вращения α вне инвариантной окружности с числом вращения α . $\square$

Замечание 14.2.18. В отсутствие инвариантной окружности с числом вращения α несколько множеств Обри—Мазера могут сосуществовать. Более того, часто встречаются многопараметрические семейства таких множеств¹.

14.2.6. Гомоклинические и гетероклинические орбиты. Теперь мы обратим процесс и будем приближать рациональные числа иррациональными, рассматривая пределы соответствующих множеств Обри—Мазера с целью построить непериодические орбиты с рациональным числом вращения.

Предложение 14.2.19. Пусть $\phi: S \rightarrow S$ — сохраняющее площадь закручивающее отображение и p/q — рациональное число, принадлежащее отрезку закручивания. Тогда существует сохраняющее порядок замкнутое ϕ -инвариантное множество с числом вращения p/q , которое либо является инвариантной окружностью, состоящей из периодических орбит, либо содержит непериодические точки. Более того, в последнем случае обе концевые точки каждого интервала, дополнительного к этому множеству, являются непериодическими.

Доказательство. Пусть $(\alpha_n)_{n \in \mathbb{N}}$ — последовательность иррациональных чисел, принадлежащих отрезку закручивания, аппроксимирующая число p/q . Рассмотрим соответствующие инвариантные минимальные сохраняющие порядок множества A_n с числами вращения α_n . Без потери общности мы предполагаем, что множества A_n сходятся к множеству A в топологии метрики Хаусдорфа при $n \rightarrow \infty$. Очевидно, что множество A является ϕ -инвариантным и упорядоченным. Если среди множеств A_n бесконечно много окружностей, то A тоже окружность и в силу непрерывности числа вращения ограничение отображения ϕ на эту окружность имеет число вращения p/q . В соответствии с классификацией отображений окружности с рациональным числом вращения (см. предложение 4.3.12) в этом случае предложение доказано. Поэтому мы можем предполагать, что все

¹J. Mather. More Denjoy Minimal Sets for Area Preserving Diffeomorphisms // Commentarii Mathematici Helvetici. 1985. V. 60, no. 4. P. 508—557.

множества A_n — множества Обри—Мазера. Чтобы понять, какова динамика на множестве A , рассмотрим разрывы, т. е. интервалы на окружности S^1 , дополнительные к образу проекции множества A на S^1 . Каждый из этих разрывов $G \subset S^1$ имеет корректно определенную длину $l(G)$, и мы хотим показать, что обе концевые точки такого разрыва непериодичны.

Разрыв G множества A есть предел в метрике Хаусдорфа соответствующих разрывов G_n множеств A_n . Обозначим через ϕ_n продолжение до гомеоморфизма окружности отображения, получаемого при проекции $\phi|_{A_n}$ на S^1 , а через ϕ_0 — подобное продолжение, соответствующее $\phi|_A$. Поскольку отображение ϕ_n имеет иррациональное число вращения, образы разрыва G_n под действием итераций отображения ϕ_n попарно не пересекаются, так что $\sum_{m \in \mathbb{N}} l(\phi_n^m(G_n)) \leq 1$. Если обе конечные точки разрыва G периодичны, то разрыв G периодичен, т. е. ряд $\sum_{n \in \mathbb{N}} l(\phi_n^m G_n)$ расходится. Но $l(\phi_n^m(G_n)) \rightarrow l(\phi_0^m(G))$ для всех $m \in \mathbb{N}$, что приводит к противоречию. Таким образом, одна из концевых точек разрыва G непериодична.

Другая концевая точка в таком случае также непериодична, так как иначе $\phi_0^q(G)$ будет разрывом, нетривиально пересекающимся с разрывом G и не совпадающим с ним. $\square$

Таким образом, мы можем описать структуру такого инвариантного множества в типичном случае, когда оно содержит конечное число периодических орбит.

Следствие 14.2.20. *Если замкнутое сохраняющее порядок ϕ -инвариантное множество A с рациональным числом вращения p/q содержит лишь конечное число периодических орбит, то существует полный набор гетероклинических связок в следующем смысле: если $\gamma_1, \dots, \gamma_s$ обозначают периодические орбиты в A , упорядоченные в соответствии с индуцированным циклическим порядком на окружности, то существуют такие гетероклинические орбиты, что или*

$$\begin{aligned}\gamma_1 &= \omega(\sigma_s) = \alpha(\sigma_1), \\ \gamma_2 &= \omega(\sigma_1) = \alpha(\sigma_2), \\ &\dots\dots\dots \\ \gamma_s &= \omega(\sigma_{s-1}) = \alpha(\sigma_s),\end{aligned}$$

или то же самое верно, если α и ω поменять местами. Здесь α и ω означают α - и ω -предельные множества орбиты (см. определение 4.3.18). Если $s = 1$, то орбита σ_1 , разумеется, является гомоклинической.

§ 14.3. Инвариантные окружности и области неустойчивости

14.3.1. Глобальное устройство инвариантных окружностей. В предыдущем параграфе мы встретились с инвариантными кривыми закручивающего отображения, которые появляются как пределы биркгофовых периодических орбит и являются поэтому графиками липшицевых отображений $S^1 \rightarrow [-1; 1]$. Суще-

ствование подобной окружности является альтернативой существованию канторова множества Обри—Мазера. Хотя в общем случае эти возможности не являются взаимно исключающими, они часто оказываются таковыми; например, это так в случае, когда орбиты плотны на окружности. Конечно, основное различие между инвариантной окружностью и множеством Обри—Мазера состоит в том, что первая разделяет фазовое пространство. Поскольку компоненты границы сохраняются, любая орбита, начинающаяся по одну сторону от инвариантной окружности, остается там навсегда. Поэтому существование даже одной инвариантной окружности предоставляет существенную информацию о поведении *всех* орбит. Естественно задаться вопросом, существуют ли другие инвариантные множества, кроме графиков липшицевых функций, которые разделяют фазовое пространство. Подобные множества могут возникать в окрестности определенных периодических орбит, как мы уже видели в случае эллиптического бильярда и пары инвариантных кривых, окружающих устойчивые орбиты периода 2, соответствующие гиперболе в качестве каустики (см. п. 6.3.5.3). Однако если мы рассматриваем только множества, которые делят цилиндр таким образом, что компоненты границы лежат в различных кусках, то следующий классический результат Биркгофа показывает, что это может случиться лишь благодаря тому, что существуют инвариантные кривые, являющиеся графиками липшицевых функций.

Теорема 14.3.1². *Если U — открытое инвариантное множество для дифференцируемого закручивающего отображения ϕ , содержащее окрестность «нижней» окружности $S^1 \times \{-1\}$ и имеющее связную границу, то граница множества U является графиком липшицевой функции.*

Динамику закручивающего отображения на объединении инвариантных окружностей можно описать в терминах динамики отображений окружности (см. § 4.3). Поэтому необходимо лишь понять, что происходит вне объединения инвариантных окружностей.

Рассмотрим сначала простой пример. Для бильярда в эллипсе существует в точности одна инвариантная окружность с числом вращения, отличным от $1/2$; соответствующие каустики являются софокусными эллипсами. Для числа вращения $1/2$ существует две инвариантные окружности, соответствующие орбитам, проходящим через фокусы: в картине фазового пространства (рис. 6.3.9) они образованы верхней и нижней ветвью сепаратрис гиперболической орбиты периода 2 (большая ось или диаметр). Другие орбиты оборачиваются вокруг эллиптической орбиты периода 2 (малая ось); соответствующие каустики являются софокусными гиперболами. Подобная картина оказывается возможной благодаря тому, что инвариантная окружность с рациональным числом вращения не обязательно единственна. По теореме 14.2.15 существует не более одной инвариантной окружности с данным иррациональным числом вращения, так что инвариантные окружности упорядочены в соответствии со своими числами вращения. Каждая окружность полностью лежит в одной из компонент дополнения до любой другой

²А.Каток, Б.Хасселблат. Введение в современную теорию динамических систем, теорема 13.2.13.

окружности; более того, предел последовательности инвариантных окружностей есть инвариантная окружность, и число вращения — непрерывная функция на множестве инвариантных окружностей.

14.3.2. Области неустойчивости. Таким образом, любой отрезок $[a, b]$, лежащий в дополнении к множеству чисел вращения инвариантных окружностей, единственным образом определяет область, граница которой состоит из непересекающихся инвариантных окружностей с числами вращения a и b соответственно. Подобная область называется *областью неустойчивости*.

Мы видели, что существуют билиардные отображения без каустик (теорема 6.4.7) и поэтому без инвариантных кривых. Для таких отображений весь цилиндр является одной областью неустойчивости, и поэтому для *любого* иррационального числа от 0 до 1 существует нигде не плотное множество Обри—Мазера.

Простая замена координат позволяет нам рассматривать ограничение закручивающего отображения на область неустойчивости как закручивающее отображение, отрезок закручивания которого есть отрезок между числами вращения двух граничных инвариантных окружностей. Поэтому для любого рационального (соответственно иррационального) числа из этого отрезка существует биркгофова периодическая орбита (соответственно множество Обри—Мазера) внутри области неустойчивости. При этом не существует «барьеров» в виде инвариантных окружностей, препятствующих орбитам блуждать по всей области, в частности осциллировать между граничными компонентами.

1. *Невообразимые трудности.* Динамика внутри области неустойчивости сложна. Как в общем случае, так и в «типичных» ситуациях можно найти много орбит со специальными свойствами, и существуют правдоподобные гипотезы о поведении большинства орбит («большинство» может означать или орбиты, покрывающие открытое плотное множество, или дополнение к множеству нулевой меры). По всей видимости, строгий анализ типичных орбит выходит за рамки существующих и вообразимых методов даже при самых оптимистических предположениях. Сложность этой задачи, вероятно, превосходит сложность некоторых знаменитых задач, представляющих большую математическую ценность (таких как гипотеза Пуанкаре в трехмерной топологии), и едва ли можно ожидать существенного прогресса в течение XXI в.

2. *Энтропия и подковы.* Запрещение экспоненциального роста серьезно ограничивает динамику закручивающего отображения.

Теорема 14.3.2³. *Закручивающее отображение с топологической энтропией, равной нулю, имеет инвариантные окружности для любого числа вращения внутри отрезка закручивания; в частности, для такого отображения не существует областей неустойчивости.*

Используя теорему 12.4.1, мы получаем такой результат.

³S. B. Angenent. A Remark on the Topological Entropy and Invariant Circles of an Area Preserving Twistmap // Twist Mappings and Their Applications. New York: Springer, 1992. P. 1—5.

Следствие 14.3.3. *Для любого закручивающего диффеоморфизма класса C^2 и любой области неустойчивости существует подкова, а стало быть, и гиперболическая периодическая точка с трансверсальной гомоклинической точкой в этой области.*

Поэтому внутри любой области неустойчивости возникают все сложные явления, совместные с сохранением площади, о которых мы говорили в гл. 12.

3. *Специальные орбиты, полученные вариационными методами.* Дальнейшее развитие вариационного подхода к закручивающим отображениям состоит в использовании критических точек должным образом построенных функционалов на аккуратно и искусно выбранных пространствах состояний, удовлетворяющих условиям, которые позволяют исключить простые решения, такие как упорядоченные орбиты. Этот подход был далеко продвинут Джоном Мазером.

При таком подходе мы получаем орбиты, которые перемещаются от одной компоненты границы области неустойчивости до другой в любом направлении, т. е. орбиты, гомоклинические и гетероклинические по отношению к компонентам границы, а также орбиты, которые осциллируют между ними в предписанном порядке. Более того, обнаруживаются гетероклинические орбиты по отношению к парам множеств Обри—Мазера с различными числами вращения и более сложные орбиты, которые блуждают между различными наборами таких множеств предписанным образом. И все богатство этих орбит покрывает множество, которое предположительно всегда является «тощим» как метрически (множество меры нуль), так и топологически (нигде не плотное множество).

4. *Сложность в типичной ситуации.* Существуют динамические ограничения на поведение биркгофовой периодической орбиты. Например, максимальная орбита не может быть эллиптической, т. е. иметь пару комплексно-сопряженных собственных значений. Такие орбиты, если они невырождены, всегда гиперболически, обычно с гетероклиническими связками между различными точками орбиты. Поэтому подкова, которая всегда возникает благодаря положительности энтропии, в этом случае появляется в весьма специфическом виде (теорема 12.3.1).

Биркгофовы орбиты типа минимакс обычно эллиптические. Широко распространено заблуждение, особенно среди специалистов в естественных науках и инженеров, работающих с моделями, которые позволяют получать закручивающие отображения или отображения подобной природы, что за исключением вырожденной орбиты (с двойным собственным значением 1) орбиты типа минимакс всегда эллиптические. Это не так. Орбиты типа минимакс могут быть гиперболическими, но с отрицательными собственными значениями, как в знаменитом примере биллиарда на «стадионе» (см. рис. 14.1.3)⁴. И все же эллиптичность орбит такого типа — это обычное явление, например, оно имеет место для малых возмущений интегрируемых закручивающих отображений. Эллиптические периодические орбиты часто порождают острова относительной устойчивости, благодаря тому что в окрестности подобной орбиты отображение за период

⁴А. Каток, Б. Хасселблат. Введение в современную теорию динамических систем, § 9.2.

в соответствующим образом выбранных координатах становится закручиванием и имеет инвариантные кривые, окружающие орбиты (представьте, что происходит в областях неустойчивости для *этого* закручивающего отображения, и попытайтесь итерировать картинку в вашем воображении!). Эти острова, таким образом, исключаются по крайней мере из общей игры сложности, поскольку, например, все богатство построенных с помощью вариационных методов орбит, а также гетероклинических пересечений, лежит за пределами этих островов.

5. Неразрешимая проблема мирного сосуществования. Отсюда возникает картина поведения орбит в области неустойчивости для «типичного» сохраняющего площадь закручивающего отображения. Существуют эллиптические точки, окруженные островами относительной стабильности, из которых орбиты не могут выйти. Существуют также гиперболические точки с гомоклиническими и гетероклиническими пересечениями, равно как и другие орбиты, демонстрирующие гиперболическое поведение разными способами. Многие вопросы, которые выше были охарактеризованы как невообразимо трудные, относятся к преобладанию одного из типов поведения и механизмам вынужденного сосуществования этих типов. Вот примеры подобных вопросов.

1) Является ли объединение эллиптических островов когда-либо (или в типичном случае) плотным?

2) Является ли дополнение к объединению эллиптических островов когда-либо (или в типичном случае) множеством меры нуль?

3) Верно ли, что замыкание устойчивого многообразия гиперболической биркгофовой периодической орбиты когда-либо (или в типичном случае, или всегда при естественных условиях невырожденности) содержит открытое множество?

3) Является ли объединение замыканий из предыдущего пункта когда-либо (или в типичном случае, или всегда) множеством меры нуль?

4) Верно ли, что энтропия Колмогорова сохраняющего площадь закручивающего отображения класса C^2 когда-либо (или в типичном случае) положительна?

§ 14.4. Периодические точки для отображений цилиндра

14.4.1. Отображения цилиндра со слабым условием закручивания. Некоторые результаты, приведенные в предыдущем параграфе, можно обобщить на отображения со значительно ослабленными условиями типа закручивания. Это относится к существованию бесконечного числа периодических орбит с возрастающими периодами. Один из классических результатов в этом направлении гораздо старше, чем теория закручивающих отображений. Его сформулировал и доказал в некоторых частных случаях Пуанкаре незадолго до своей смерти. Биркгоф, который осознал важность этого результата для решения многих задач классической механики и геометрии, привел строгое доказательство.

Напомним, что понятие закручивающего отображения было определено в п. 14.2.1 для любого гомеоморфизма замкнутого цилиндра, сохраняющего компоненты границы. В этом пункте мы предполагаем выполнение последнего условия, не оговаривая это далее дополнительно.

Теорема 14.4.1 (Последняя геометрическая теорема Пуанкаре)⁵. Пусть f — сохраняющий площадь гомеоморфизм замкнутого цилиндра, отрезок закручивания которого содержит 0 как внутреннюю точку. Тогда f имеет неподвижную точку внутри цилиндра.

Применяя эту теорему к соответствующим образом выбранной итерации отображения, можно получить утверждение о существовании периодических орбит, которые, хотя и не обязательно являются упорядоченными, все же сохраняют некоторые свойства биркгофовых орбит.

Следствие 14.4.2. Пусть p и q — взаимно простые целые числа, $q > 0$ и f — сохраняющий площадь гомеоморфизм замкнутого цилиндра, отрезок закручивания которого содержит p/q как внутреннюю точку. Тогда f имеет такую периодическую точку (s, y) периода q , что для некоторого поднятия F гомеоморфизма f и любого поднятия (x, y) точки (s, y) выполняется равенство $F^q(x, y) = (x + p, q)$.

Разумеется, в отсутствие всякого контроля над тем, что происходит внутри цилиндра, предельный процесс, описанный в § 14.2, не работает, и поэтому не существует естественного продолжения этого результата на иррациональные числа вращения.

В 1980-х гг. Джон Франкс обобщил предыдущий результат, используя гораздо более слабое свойство, чем то, что числа вращения компонент границы находятся с разных сторон от p/q . Он использовал самое слабое условие типа закручивания, какое только можно представить. Франкс также предложил более слабое условие рекуррентности, чем сохранение площади, однако мы оставляем последнее в нашей формулировке его теоремы.

Теорема 14.4.3. Пусть p и q — взаимно простые целые числа, $q > 0$, и f — сохраняющий площадь гомеоморфизм замкнутого цилиндра и F — поднятие гомеоморфизма f . Предположим, что существуют такие точки u и v на универсальной накрывающей, что

$$\lim_{n \rightarrow \infty} \frac{F^n(u) - u}{n} \leq \frac{p}{q} \leq \overline{\lim}_{n \rightarrow \infty} \frac{F^n(v) - v}{n}. \quad (14.4.1)$$

Тогда гомеоморфизм f имеет такую периодическую точку периода q , что любое поднятие w этой точки удовлетворяет соотношению $F^q(w) = w + (0, p)$.

При решении упражнения 4.3.8 используется условие, подобное условию (14.4.1), но более сильное.

Если u, v — точки на универсальной накрывающей и

$$\lim_{n \rightarrow \infty} \frac{F^n(u) - u}{n} < \overline{\lim}_{n \rightarrow \infty} \frac{F^n(v) - v}{n}, \quad (14.4.2)$$

⁵G. D. Birkhoff. Dynamical Systems. Providence, RI: AMS, 1927. (American Mathematical Society Colloquium Publications, v. 9). Имеется перевод: Дж. Биркгоф. Динамические системы. М.—Л., Гостехиздат, 1941.

то существует бесконечно много различных рациональных чисел, удовлетворяющих неравенству (14.4.1), откуда следует, что f имеет бесконечно много периодических орбит.

14.4.2. Периодические точки без закручивания. Таким образом, существование конечного числа периодических точек (или их отсутствие) означает, что неравенство (14.4.2) не выполняется. Это эквивалентно тому, что все точки кольца имеют одно и то же число вращения, т. е.

$$\lim_{n \rightarrow \infty} \frac{F^n(v) - v}{n}$$

существует для любой точки v на универсальной накрывающей и не зависит от v . (На самом деле независимость от v следует из существования предела для всех v .) Тогда мы можем назвать этот предел числом вращения гомеоморфизма f . Если число вращения иррационально, то периодических точек нет совсем. Если оно рационально, то по теореме 14.4.3, примененной к граничной окружности, существует по крайней мере одна периодическая орбита. Франкс сумел значительно усилить последнее заключение.

Теорема 14.4.4. *Сохраняющий площадь гомеоморфизм замкнутого кольца с рациональным числом вращения имеет бесконечное число внутренних периодических точек.*

Поэтому из существования конечного числа периодических точек следует иррациональность числа вращения диффеоморфизма, а стало быть, полное отсутствие периодических точек, что ведет к заключительному выводу из теории Франкса.

Теорема 14.4.5⁶. *Если гомеоморфизм замкнутого кольца, сохраняющий площадь и граничные компоненты, имеет хотя бы одну периодическую (или неподвижную) точку, то он имеет бесконечное число внутренних периодических точек.*

Это основной динамический результат, используемый при доказательстве существования бесконечного числа замкнутых геодезических на любой двумерной сфере, о котором говорится в следующем параграфе.

Теорема 14.4.4 не гарантирует, что существуют точки с произвольно большим периодом, поскольку тождественное отображение имеет только неподвижные точки. Оказывается, это единственное исключение: любое отображение, никакая степень которого не является тождественным отображением, имеет периодические точки, минимальный период которых произвольно велик⁷.

Схема доказательства теоремы 14.4.4. Во-первых, предположим, что существует только конечное число периодических точек. Перейдя к соответствующей итерации, можно сделать все эти точки неподвижными, а число вращения равным 0. Удалив эти точки, получим некомпактную поверхность S конечного

⁶J. Franks. Geodesics on S^2 and Periodic Points of Annulus Homeomorphisms // Inventiones Mathematicae. 1992. V. 108, no. 2. P. 403—418.

⁷Об этом сообщил одному из авторов в частной беседе Патрис ле Калвез; не опубликовано.

рода и сохраняющий площадь гомеоморфизм этой поверхности без неподвижных точек. Используя классификацию Уильяма Тёрстона гомеоморфизмов двумерных поверхностей с точностью до гомотопии, которая является важным инструментом в маломерной топологии, можно убедиться, что, переходя к другой итерации, мы можем сделать гомеоморфизм гомотопным тождественному, т. е. что он может быть непрерывно продеформирован (по поверхности) в тождественное отображение. Теперь поднимем гомеоморфизм (с выколотыми неподвижными точками) на универсальную накрывающую поверхности S (которая на вид совсем непохожа на полосу). Можно придать смысл понятию числа вращения в этой ситуации, и это число вращения оказывается равным 0. Это позволяет нам построить возмущение нашего отображения, которое имеет внутреннюю периодическую точку с нулевым числом вращения, что противоречит глубокому результату Майкла Ханделя, гласящему, что в отсутствие неподвижной точки любая периодическая точка должна иметь ненулевое число вращения. $\square$

§ 14.5. Геодезические на сфере

В п. 1.3.3 обсуждалась задача нахождения замкнутых геодезических на (деформированной) сфере. Можно также рассматривать движение частиц по (неравномерно изогнутой) сфере в отсутствие внешних сил (см. п. 6.2.8). Вопрос заключается в том, существует ли бесконечное число различных путей, движение по которым периодически. Специалисты верили в то, что это утверждение верно, на протяжении длительного времени, однако оно было доказано в общем случае только недавно. При доказательстве используется замечательное сочетание методов из области дифференциальной геометрии, вариационного анализа, маломерной топологии и динамических систем. Хотя мы не можем представить его здесь, мы объясним, как используются некоторые динамические результаты.

Мы начнем с краткого исторического экскурса в проблему геодезических. Существование по крайней мере одной замкнутой геодезической не является специфическим для сферы (это верно для любого компактного риманова многообразия). Доказательство основано на еще одной версии принципа перевала. В случае сферы оно проводится следующим образом. Рассмотрим гладкое однопараметрическое семейство замкнутых кривых, «прикрепленных» в некоторой точке и покрывающих всю сферу (т. е. гладкое сюръективное отображение из квадрата $[0, 1] \times [0, 1]$ в S^2 , отображающее границу квадрата в одну точку). В каждом таком семействе существует кривая наибольшей длины. Очевидно, что длина такой кривой отделена от нуля числом, не зависящим от выбора семейства. Если существует семейство, на котором длина такой длиннейшей кривой достигает минимума по всем семействам, то из вариационного описания геодезической следует, что такая кривая — замкнутая геодезическая. Доказательство существования такого семейства основано на том общем рассуждении, что нижняя граница относительно всех семейств будет такой же, как относительно определенного набора семейств, обладающего тем свойством, что его можно сделать компактным в топологии, в которой функционал «максимальной длины» остается непрерывным.

Немного более тонким фактом является существование *простой*, т. е. несамопересекающейся, замкнутой геодезической. Его также можно получить с помощью вариационного рассуждения, путем нахождения критической точки функционала длины в подходящем пространстве простых замкнутых кривых.

Для случая двумерной сферы гораздо более глубокий результат, также доказанный с помощью вариационных методов, принадлежит Л. А. Люстернику и Л. Г. Шнирельману.

Теорема 14.5.1⁸. *Существует по крайней мере три различных простых замкнутых геодезических на двумерной сфере.*

Первая статья этих авторов, содержащая доказательство данной теоремы и относящаяся к 1930 г., была доступна на Западе только в очень сокращенном переводе, и другие доказательства были опубликованы позднее различными математиками.

При доказательстве теоремы Люстерника—Шнирельмана используется довольно грубый, но все же весьма замечательный инвариант топологических пространств, называемый *категорией Люстерника—Шнирельмана*. Это всего-навсего минимальное число подмножеств, на которые можно разложить пространство, причем таким образом, что каждое подмножество *стягиваемо*, т. е. может быть непрерывно продеформировано (по себе) в точку. Связь с вариационными задачами здесь следующая. Для пространств (даже бесконечномерных), на которых можно определить дифференцирование, а стало быть, понятие критической точки функции имеет смысл, категория Люстерника—Шнирельмана дает нижнюю оценку на число критических точек *любой* дифференцируемой функции.

Пространство, к которому этот критерий применяется в задаче о геодезических, строится из пространства (параметризованных) простых гладких кривых. Функционал связан с длиной, и условие на критическую точку гарантирует, что соответствующая кривая является замкнутой геодезической, параметризованной своей длиной. Топологическая часть доказательства состоит в том, чтобы показать, что категория Люстерника—Шнирельмана равна 3; разумеется, она не зависит от метрики.

Результат Люстерника—Шнирельмана оптимален в том смысле, что существуют метрики с ровно тремя простыми замкнутыми геодезическими. Примером может служить трехосный эллипсоид.

Еще до работы Люстерника и Шнирельмана Биркгоф предложил метод нахождения бесконечного числа геодезических. Он основан на существовании простой замкнутой геодезической. Биркгоф не знал, как доказать это в общем случае, хотя ему принадлежит приведенное выше рассуждение о минимаксе, которое обеспечивает существование одной замкнутой геодезической. По теореме Жор-

⁸L. A. Lusternik, L. G. Shnirelman. Sur le problème de trois géodesiques fermées sur les surfaces de genre 0 // Comptes Rendus des Séances de l'Académie des Sciences. Série I. Mathématique. 1929. V. 189. P. 269—271; Л. А. Люстерник, Л. Г. Шнирельман. Топологические методы в вариационных задачах // Труды Института математики и механики. 1930; Топологические методы в вариационных задачах и их приложение к дифференциальной геометрии поверхностей // УМН. 1947. Т. 2, № 1(17). С. 166—217; Л. А. Люстерник. Топология вариационного анализа бесконечно больших // Труды математического института им. Стеклова. 1947. Т. 19.

дана о замкнутой кривой замкнутая простая геодезическая делит сферу на две связные компоненты, которые мы будем называть северной и южной полусферой. Саму геодезическую мы будем называть экватором. Рассмотрим все единичные касательные векторы к сфере с началом на экваторе, направленные в сторону северной полусферы (при этом неважно, какую из полусфер выбрали в качестве «северной»). Множество S таких векторов параметризовано декартовым произведением окружности (экватора) и интервала (углы в $(0, \pi)$), и потому представляет собой открытый цилиндр. Каждый из этих векторов определяет единственную геодезическую. Если эта геодезическая, входящая в северную полусферу, когда-то покидает ее, а потом опять возвращается из южной полусферы, то она определяет при этом вектор того же типа. Тем самым задается отображение из подмножества R пространства S в само пространство S . Периодические орбиты этого отображения соответствуют замкнутым геодезическим, и различные периодические орбиты соответствуют разным геодезическим. Если $R = S$, возникает вопрос о продолжении этого отображения на границу открытого цилиндра S . Если подобное продолжение возможно, то из очевидной симметрии следует, что отображения двух граничных компонент являются обратными друг другу, а значит, отрезок закручивания содержит 0. Теперь возникает дихотомия. Одна возможность состоит в том, что отрезок закручивания имеет положительную длину, и тогда теорема 14.4.1 обеспечивает бесконечное число периодических орбит, а значит, бесконечное число замкнутых геодезических (в частности, по этой причине Биркгоф проявлял интерес к последней геометрической теореме Пуанкаре). Альтернатива состоит в том, что числа вращения граничных компонент равны 0. Именно на этом этапе решение задержалось примерно на 60 лет.

Результаты Франкса, теорема 14.4.4 и теорема 14.4.3 завершают доказательство существования бесконечного числа различных замкнутых геодезических в случае, когда описанное выше отображение последования определено на открытом цилиндре S и продолжается по непрерывности на замкнутый цилиндр.

Доказательство для оставшихся случаев было получено Виктором Бангертом; его результаты частично основаны на его же более ранней совместной работе с Вильгельмом Клингенбергом. Нужно рассмотреть два случая. Если отображение определено на открытом цилиндре, но не продолжается на границу, то, оказывается, вдоль первоначальной замкнутой простой геодезической нет «сопряженных точек». Грубо говоря, близлежащие геодезические расходятся. Из этого следует, что функционал длины имеет специальный вид, который позволяет использовать его при применении вариационного подхода и доказать, что число замкнутых геодезических бесконечно. Если же отображение последования не определено, то, как показывает Бангерт, существует другая простая замкнутая геодезическая без сопряженных точек и задача сводится к предыдущей.

Окончательный результат этих рассуждений — следующая замечательная теорема.

Теорема 14.5.2. *На любой гладкой сфере существует бесконечное число замкнутых геодезических.*

ДИНАМИКА, ТЕОРИЯ ЧИСЕЛ И ДИОФАНТОВЫ ПРИБЛИЖЕНИЯ

Некоторые вопросы теории чисел, восходящие еще к работам Дирихле, Якоби, Кронекера и Вейля, можно рассматривать как важные истоки современной динамики. Плодотворное взаимодействие развивается в двух направлениях: динамические методы часто предлагают новые и порой неожиданные походы к задачам аналитической теории чисел, а с другой стороны, алгебраическая теория чисел предоставляет аппарат для более глубокого исследования модельных динамических систем, чем позволяют традиционные аналитические, топологические и геометрические методы.

Содержание первых четырех параграфов этой главы является классическим и призвано продемонстрировать пользу динамического подхода в задачах о равномерном распределении и диофантовых приближениях чисел. Постепенно становится ясной значительная роль гиперболической геометрии. В последнем параграфе представлено краткое описание одного из основных достижений динамического подхода в аналитической теории чисел: доказательство гипотезы Оппенгейма о малых значениях квадратичных форм от трех переменных.

§ 15.1. Равномерная распределенность дробных частей для многочленов

В этом параграфе мы покажем, как некоторые общие динамические методы позволяют установить равномерную распределенность (строгую эргодичность, см. п. 4.1.4) для специальной динамической системы, что в свою очередь используется для решения теоретико-числовой задачи, рассматриваемой во введении (см. п. 1.3.5), и ее обобщений. Строгая эргодичность является подходящим инструментом для доказательства равномерной распределенности заданной последовательности, потому что каждой последовательности соответствует орбита динамической системы. Сходимости вне множества меры нуль недостаточно, так как данная орбита может принадлежать исключительному множеству меры нуль.

15.1.1. Квадратные многочлены и аффинные отображения двумерного тора. Вернемся к задаче распределения последних цифр перед запятой в последовательности $x_n = n^2\sqrt{2}$, которую мы рассматривали в п. 1.3.5. Число $\sqrt{2}$ было выбрано только потому, что оно иррационально, поэтому далее будем рассматривать последовательность αn^2 для любого иррационального числа α . Вместо того чтобы работать с последними цифрами, будем рассматривать последовательность первых цифр дробных частей $\alpha n^2/10$. В отличие от некоторых других последовательностей, о которых говорилось ранее в этой книге, эта последовательность не является последовательностью значений подходящих итераций одномерного

отображения и не может быть естественным образом превращена в такую последовательность. Однако есть метод, преодолевающий эту трудность. Мы построим динамическую систему более высокой размерности, а требуемую последовательность будем интерпретировать как последовательность координат или, в более общем виде, значений функции от подходящих итераций отображения, взятых в фиксированной начальной точке. В нашем случае подходящая динамическая система — это следующее *аффинное* отображение двумерного тора:

$$A_{\alpha/5}(x, y) = (x + \alpha/5, y + x) \pmod{1}.$$

Это отображение «интегрируемо» в следующем смысле: его итерации задаются явной формулой, а именно

$$A_{\alpha/5}^n(x, y) = (x + n\alpha/5, y + nx + n(n-1)\alpha/10) \pmod{1}.$$

Последовательность абсцисс знакома нам из п. 4.1.1, а последовательность ординат содержит нужные квадратичные члены. Чтобы избавиться от ненужных линейных частей, возьмем $x = \alpha/10$ и $y = 0$.

Используя соображения, приведенные в п. 4.2.2, мы видим, что для того чтобы установить равномерность распределения последних цифр последовательности αn^2 , достаточно показать, что для любого десятичного отрезка $\Delta_k = [k/10, (k+1)/10]$, $k = 0, \dots, 9$, средние Биркгофа характеристической функции χ_{Δ_k} , имеющие вид

$$\frac{1}{n} \sum_{i=0}^n \chi_{\Delta_k} \circ A_{\alpha/5}^i,$$

равномерно сходятся к $\frac{1}{10}$. И вновь единственное свойство характеристических функций, которое мы используем, — это то, что они интегрируемы по Риману (см. п. 4.1.5).

Есть два основных метода доказательства равномерной сходимости средних Биркгофа этих отображений для всех интегрируемых по Риману функций. Первый метод был предложен Германом Вейлем в 1916 г. Это усовершенствованный метод Кронекера—Вейля, который мы использовали для поворотов окружности в п. 4.1.6 и для сдвигов тора в п. 5.1.6. Он включает в себя оценку средних Биркгофа для характеров $\exp 2\pi i(kx + ly)$ ($k, l \in \mathbb{Z}$) с помощью более сложных вычислений, чем простое суммирование геометрической прогрессии в п. 4.1.6. Выражения в вычислениях Вейля являются частным случаем *тригонометрических сумм*, играющих значительную роль в аналитической теории чисел. Далее по теореме Вейерштрасса устанавливается равномерная сходимость для всех непрерывных функций (строгая эргодичность, см. определение 4.1.18), и стандартное рассуждение из п. 4.1.6 дает возможность распространить это свойство на все интегрируемые по Риману функции. Основное достоинство этого метода в том, что он позволяет для определенных классов функций установить *скорость* сходимости средних Биркгофа.

15.1.2. Эргодичность и строгая эргодичность. Рассмотрим теперь более качественный подход, предложенный Хиллелом Фюрстенбергом примерно

в 1960 г. Этот метод не позволяет получить оценки скорости сходимости средних Биркгофа, но зато он не опирается на сложные выкладки с тригонометрическими суммами и поэтому более широко применим. Этот подход основан на том, что и строгая эргодичность (равномерная сходимость средних для непрерывных функций), и более слабое свойство — сходимость средних вне множества меры нуль (см. п. 7.5.3) эквивалентны некоторым чисто качественным свойствам.

Для строгой эргодичности соответствующее качественное свойство — это *единственность инвариантной меры* или интеграла (см. определение 10.4.1, предложение 10.4.2). Для сходимости средних вне множества меры нуль соответствующее качественное свойство называется *эргодичностью*. Для сохраняющего объем преобразования f эргодичность означает, что каждое разумным образом выбранное f -инвариантное множество является либо множеством меры нуль, либо дополнением до множества меры нуль, или, что эквивалентно, каждая разумным образом выбранная f -инвариантная функция постоянна вне множества меры нуль.

Строгое определение «разумного» выбора в каждом случае требует использования понятия измеримости, которое не обсуждается в этой книге, за исключением понятия множества меры нуль.

Второй метод доказательства равномерности распределения для многочленов опирается на основной критерий строгой эргодичности для класса отображений, называемых групповыми расширениями, который мы сформулируем в частном случае.

Предложение 15.1.1. Пусть f — непрерывное, сохраняющее объем, строго эргодическое отображение тора $\mathbb{T}^k$ и отображение $\phi: \mathbb{T}^k \rightarrow S^1$ непрерывно. Тогда если отображение $f_\phi: \mathbb{T}^{k+1} \rightarrow \mathbb{T}^{k+1}$, определенное формулой

$$f_\phi(x, y) = (f(x), y + \phi(x)),$$

эргодично, то оно строго эргодично.

Набросок доказательства. Мы будем использовать то обстоятельство, что строгая эргодичность эквивалентна единственности инвариантного интеграла, а также свойства инвариантного интеграла, следующие из *разложения на эргодические компоненты*.

Так как отображение f_ϕ эргодично (по отношению к обычной мере Лебега на торе $\mathbb{T}^{k+1}$), каждая орбита вне множества меры нуль равномерно распределена. Но так как объем инвариантен относительно сдвигов, в частности «вертикальных» сдвигов $(x, y) \rightarrow (x, y + \beta)$, и отображение f_ϕ коммутирует с вертикальными сдвигами, множество A равномерно распределенных орбит также инвариантно относительно вертикальных сдвигов, поэтому оно содержит целые окружности $x = \text{const}$. Мы называем такое множество *насыщенным*. Множество A не является множеством меры нуль, так как его дополнение — множество меры нуль (объединение двух множеств меры нуль — множество меры нуль, а весь тор не является множеством меры нуль).

Теперь предположим, что отображение f_ϕ не строго эргодично. Это означает, что существует по крайней мере еще один инвариантный интеграл, отличный

от порожденного стандартным объемом, а значит (это следует из разложения на эргодические компоненты), еще одно асимптотически равномерное распределение. Множество точек с этим новым асимптотическим распределением не пересекается с множеством точек, равномерно распределенных относительно обычной меры Лебега, а следовательно, принадлежит множеству окружностей $x = \text{const}$, имеющему меру нуль. Проектируя этот инвариантный интеграл на координату x , получаем инвариантный интеграл для f , который по предположению единствен и, следовательно, совпадает со стандартным. Это приводит к противоречию: множество A всех равномерно распределенных точек по обычной мере Лебега на торе $\mathbb{T}^{k+1}$ является множеством меры нуль для второго инвариантного интеграла, но последний проектируется в стандартный интеграл на $\mathbb{T}^k$, поэтому множество A , являющееся насыщенным, проектируется в множество меры нуль. $\square$

15.1.3. Равномерное распределение для квадратичных многочленов.

С помощью этих методов несложно доказать равномерность распределения для квадратов.

Предложение 15.1.2. *Отображение $A_\beta(x, y) = (x + \beta, y + x) \pmod{1}$ тора $\mathbb{T}^2$ является строго эргодическим.*

Доказательство. Отображение принадлежит классу отображений, рассматриваемых в предложении 15.1.1, причем $f = R_\beta$. Поэтому достаточно доказать эргодичность, т.е. убедиться, что любая разумным образом выбранная инвариантная функция постоянна вне множества меры нуль. Достаточно рассматривать функции класса L^2 , для которых ряд Фурье однозначно определен, хотя может и не сходиться в некоторых точках. Проверка отсутствия инвариантных непостоянных функций сводится к написанию уравнений на коэффициенты Фурье этих функций, из которых следует, что все коэффициенты, кроме одного, обращаются в нуль. А именно, пусть g — инвариантная функция из L^2 . Тогда в терминах рядов Фурье

$$\begin{aligned} \sum_{i,j} a_{i+j} e^{2\pi i x} e^{2\pi i j x} e^{2\pi i j y} &= \sum_{i,j} a_{ij} e^{2\pi i x} e^{2\pi i j y} = g(x, y) = \\ &= g(A_\beta(x, y)) = g(x + \beta, y + x) = \sum_{i,j} a_{ij} e^{2\pi i (x+\beta)} e^{2\pi i j (y+x)}, \end{aligned}$$

т.е. $a_{i+j} = a_{ij} e^{2\pi i \beta}$ для любых $i, j \in \mathbb{Z}$. Отсюда следует, что $|a_{i+kj}| = |a_{ij}|$ для любых $i, j, k \in \mathbb{Z}$. Квадраты коэффициентов Фурье функции из L^2 суммируемы, поэтому $\lim_{i \rightarrow \infty} a_{ij} = 0$ и $a_{ij} = 0$ для всех $i, j \in \mathbb{Z} \setminus \{0\}$ (мы используем неравенство $e^{2\pi i \beta} \neq 1$ для $i \neq 0$, так как $\beta \notin \mathbb{Q}$). Если $i = 0$, то $a_{i+j} = a_{ij} e^{2\pi i \beta}$, и мы получаем $a_{0j} = a_{ij} = 0$ для $j \neq 0$. Если $j = 0$, то $a_{i+j} = a_{ij} e^{2\pi i \beta}$, и мы получаем $a_{i0} = a_{i0} e^{2\pi i \beta}$, откуда следует, что $a_{i0} = 0$ для $i \neq 0$. Таким образом, во всех случаях, кроме случая $i = j = 0$, выполняется равенство $a_{ij} = 0$. $\square$

Следствие 15.1.3. *Если $\alpha \notin \mathbb{Q}$, то последние цифры последовательности αn^2 равномерно распределены.*

Доказательство. Так как для $A_{\alpha/5}^n$ справедлива формула $A_{\alpha/5}^n(x, y) = (x + n\alpha/5, y + nx + n(n-1)\alpha/10) \pmod{1}$ и отображение $A_{\alpha/5}$ строго эргодично, каждая орбита равномерно распределена. Положив $x = \alpha/10$ и $y = 0$, получаем, что последовательность $y + nx + n(n-1)\alpha/10 = \alpha n^2/10$ равномерно распределена на единичном отрезке, в частности, равномерно распределены первые цифры после запятой в десятичной записи ее элементов. $\square$

15.1.4. Равномерное распределение для многочленов. Метод, описанный выше, можно применить индуктивно к доказательству следующего утверждения.

Предложение 15.1.4. *Если α — иррациональное число, то отображение $A_{k,\alpha}: \mathbb{T}^{k+1} \rightarrow \mathbb{T}^{k+1}$, определенное формулой*

$$A_{k,\alpha}(x_0, x_1, \dots, x_k) = (x_0 + \alpha, x_1 + x_0, \dots, x_k + x_{k-1}) \pmod{1},$$

строго эргодично.

Для любого многочлена P степени $k+1$ с иррациональным старшим коэффициентом существуют такое число α и такие начальные условия $x_0, x_1, \dots, x_k$, что последняя координата итерации $A_{k,\alpha}^n(x_0, x_1, \dots, x_k)$ равна $P(n)$. Чтобы это доказать, нужно найти явную формулу для итераций $A_{k,\alpha}^n$, а затем решить систему линейных уравнений. Из периодичности дробных частей многочленов с рациональными коэффициентами следует теорема о равномерном распределении для многочленов.

Теорема 15.1.5. *Дробные части значений $P(n)$ для многочлена с хотя бы одним иррациональным коэффициентом равномерно распределены.*

Используя то обстоятельство, что последние k цифр перед запятой у действительного числа становятся первыми k цифрами после деления его на 10^k , получаем следствие из теоремы.

Следствие 15.1.6. *Если P — многочлен с хотя бы одним иррациональным коэффициентом, то последние k цифр перед запятой чисел $(P(n))_{n \in \mathbb{N}}$ равномерно распределены.*

§ 15.2. Цепные дроби и рациональные приближения

Мы уже знаем (см. предложение 6.1.12), что все орбиты поворота окружности R_α (равномерно) рекуррентны. Как видно из рассмотрения орбиты точки 0, это означает, что $n\alpha$ подходит сколь угодно близко к целым числам для достаточно больших $n \in \mathbb{N}$, или что величина $|n\alpha - m|$ может быть сделана сколь угодно малой при надлежащем выборе чисел m, n .

15.2.1. Наилучшие приближения. Для выяснения количественной стороны вопроса попробуем выяснить, насколько большими должны быть числа m, n , чтобы выполнялось условие $|n\alpha - m| < \epsilon$. Поставим вопрос немного иначе: пусть заданы границы для n , тогда нам требуется найти минимум величины $\min_{m \in \mathbb{Z}} |n\alpha - m|$.

Для рационального α эта задача сводится к нахождению минимума на конечном множестве, и для нас она не интересна. Впредь будем считать, что $\alpha \notin \mathbb{Q}$, и пока полагать, что $0 < \alpha < 1$.

Определение 15.2.1. Рациональное число p/q называется *наилучшим рациональным приближением*, или просто *наилучшим приближением* иррационального числа α , если $q > 0$ и $|q\alpha - p| \leq |n\alpha - m|$ для любых $m, n \in \mathbb{Z}$, $|n| \leq q$.

Из определения, очевидно, следует, что числа p и q должны быть взаимно просты. Из топологической транзитивности поворота на иррациональный угол следует, что существует бесконечное число различных наилучших приближений данного иррационального числа.

Теперь для поиска наилучшего приближения будем внимательно следить за тем, как близко число $n\alpha$ подходит к целым числам (и каковы эти целые числа). В частности, будем отмечать те числа $n \in \mathbb{N}$, для которых $n\alpha$ ближе к $\mathbb{Z}$, чем $i\alpha$ для любого $i < n$. Мы хотим таким образом получить некоторую рекурсивную процедуру. С динамической точки зрения мы смотрим, как близко орбита точки 0 подходит к 0 при итерациях поворота R_α , отмечая ближайшие подходы.

Обозначим первое такое целое число через a_1 . Оно определяется условием $a_1\alpha < 1 < (a_1 + 1)\alpha$, т. е. находится к $\mathbb{Z}$ ближе, чем α . Это означает, что $a_1 = \lfloor 1/\alpha \rfloor$, где $\lfloor \cdot \rfloor$ — обозначение для целой части. В терминах приближения числа α рациональными числами это означает, что $a_1\alpha \approx 1$, или $\alpha \approx 1/a_1$. Чтобы перейти к следующему шагу, удобно считать, что приближение является хорошим, т. е. ошибка $\delta := 1 - a_1\alpha$ мала.

Следующим близким приближением целого числа считаем такое число $n \in \mathbb{N}$, что $n\alpha$ попадает в δ -окрестность целого числа. Предположим, что δ мало. Заметим, что после a_1 шагов после предыдущего перехода через целое число $i\alpha$ для $a_1 < i \leq 2a_1$ лежат между 1 и 2, и в итоге мы получаем $2a_1\alpha = 2(1 - \delta) = 2 - 2\delta$; поэтому мы отошли от целого числа на еще одно значение δ .

Пусть a_2 — последнее такое число n , что $n\delta < \alpha$, т. е. $a_2 = \lfloor \alpha/\delta \rfloor$. Тогда $0 < \delta_1 := \alpha - a_2\delta < \delta$ и число $a_1a_2\alpha$ «отстает от целого» приблизительно на α (точнее, на $\alpha - \delta_1$). Таким образом, число $(a_1a_2 + 1)\alpha$ ближе к целому числу (а именно, к a_2), чем все предыдущие числа $i\alpha$, потому что выполняется равенство $(a_1a_2 + 1)\alpha - a_2 = \alpha - a_2(1 - a_1\alpha) = \alpha - a_2\delta = \delta_1 < \delta$. Заметим, что соответствующим рациональным приближением является $(a_1a_2 + 1)\alpha \approx a_2$, или

$$\alpha \approx \frac{a_2}{a_1a_2 + 1} = \frac{1}{a_1 + \frac{1}{a_2}}.$$

Чтобы понять, как выбрать число a_2 , заметим, что $\alpha_2 := \delta/\alpha = (1 - a_1\alpha)/\alpha = (1/\alpha) - a_1 = \{1/\alpha\}$, где $\{\cdot\}$ означает дробную часть числа, и поэтому $a_2 = \lfloor 1/\alpha_2 \rfloor$.

15.2.2. Представление в виде цепной дроби. Приведенная схема дает алгоритм наилучших рациональных приближений иррациональных чисел и представления любого действительного числа в виде цепной дроби.

Теорема 15.2.2. Пусть $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Определим последовательности $(a_i)_{i=0}^\infty$ и $(\alpha_i)_{i=0}^\infty$ рекуррентно формулами

$$a_0 := [\alpha], \quad \alpha_1 := \{\alpha\}, \quad a_i := \left[\frac{1}{\alpha_i} \right] \quad \alpha_{i+1} = \left\{ \frac{1}{\alpha_i} \right\}$$

и положим

$$a_0 + \frac{1}{a_1 + \frac{1}{\ddots + \frac{1}{a_n}}} =: \frac{p_n}{q_n};$$

считаем, что $q_n > 0$ и дроби $\frac{p_n}{q_n}$ несократимы. (Эти дроби называются подходящими.)

Тогда

$$\alpha = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots}} := \lim_{n \rightarrow \infty} a_0 + \frac{1}{a_1 + \frac{1}{\ddots + \frac{1}{a_n}}}.$$

Если $\alpha \in \mathbb{Q}$, то данная рекуррентная формула заканчивается на $\alpha_{i+1} = 0$ для некоторого i и

$$\alpha = a_0 + \frac{1}{a_1 + \frac{1}{\ddots + \frac{1}{a_i}}}.$$

Обе последовательности p_n и q_n удовлетворяют рекуррентному соотношению второго порядка

$$x_{n+1} = x_{n-1} + a_{n+1}x_n$$

с начальными условиями $p_0 = a_0$, $p_1 = a_0 a_1 + 1$, $q_0 = 1$, $q_1 = a_0$;

$$p_{n+1}q_n - p_nq_{n+1} = (-1)^n;$$

и, наконец, каждое наилучшее рациональное приближение иррационального числа является подходящей дробью, и каждая подходящая дробь является наилучшим приближением.

Замечание 15.2.3. Несложные рассуждения показывают единственность разложения числа в цепную дробь. Есть несколько различий между этим и десятичным разложениями. Десятичное разложение не всегда единственно. Для натуральных чисел, используемых в задании цепной дроби, нет априори заданной границы, а в десятичном разложении используются только цифры, т. е. числа от 0 до 9. Десятичное разложение различает рациональные и иррациональные числа (для первых оно периодически, для вторых — нет), но мы можем увидеть это различие, только если дано все разложение. В случае же цепной дроби число рационально тогда и только тогда, когда разложение заканчивается на некотором шаге, а это видно из конечной части цепной дроби.

Доказательство правильности выбора чисел a_i . Пусть дано разложение в цепную дробь. Легко проверить, что значения a_i определяются приведенной выше процедурой: если

$$\alpha = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots}},$$

то $[\alpha] = a_0$ и, положив $\alpha_1 := \alpha - a_0 = \{\alpha\}$, мы получим

$$\frac{1}{\alpha_1} = a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}},$$

поэтому $a_1 = \left\lfloor \frac{1}{\alpha_1} \right\rfloor$ и

$$\alpha_2 := \left\{ \frac{1}{\alpha_1} \right\} = \frac{1}{\alpha_1} - a_1 = \frac{1}{a_2 + \frac{1}{a_3 + \dots}},$$

поэтому $a_2 = \left\lfloor \frac{1}{\alpha_2} \right\rfloor$, и т. д. □

Замечание 15.2.4. Заметим, что дроби p_n/q_n поочередно принимают значения то больше, то меньше α . В этом можно убедиться, рассмотрев ближайшее приближение или заметив, что добавление к конечной цепной дроби элемента то увеличивает, то уменьшает дробь.

Пример 15.2.5. Из рекуррентной формулы для p_n и q_n следует, что числа Фибоначчи $(b_n)_{n \in \mathbb{N}_0} = (1, 1, 2, 3, 5, 8, \dots)$ образуют подходящие дроби $1, 2, 3/2, 5/3, 8/5, \dots$ золотого сечения

$$\frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \dots}}} = \frac{\sqrt{5} - 1}{2}.$$

Пример 15.2.6. Рассмотрим простую цепную дробь

$$x = \frac{1}{2 + \frac{1}{2 + \frac{1}{2 + \dots}}},$$

которая является положительным корнем уравнения $1/x = x + 2$. Таким образом, $x^2 + 2x - 1 = 0$, и $x = \sqrt{2} - 1$. Предыдущий пример можно было рассмотреть аналогичным образом.

Эти примеры наводят на мысль о плодотворном обобщении. Предположим, что разложение числа в цепную дробь периодически, начиная с некоторого момента. Тогда это число можно представить как корень квадратного уравнения с рациональными коэффициентами. Оказывается, обратное тоже верно. Таким образом, разложение действительного числа в цепную дробь является периодическим,

начиная с некоторого момента тогда и только тогда, когда оно является *квадратичной иррациональностью*, т. е. имеет вид $a + \sqrt{b}$, где a и b — рациональные числа.

15.2.3. Скорость аппроксимации и динамика. Есть другой естественный способ оценить качество рационального приближения p/q иррационального числа α , а именно, мы можем сравнивать абсолютную величину разности $\alpha - p/q$ с некой заданной функцией от знаменателя q . Как и следовало ожидать, при этом появляются подходящие дроби.

Предложение 15.2.7. *Если $\alpha \notin \mathbb{Q}$ и $n \in \mathbb{N}$, то выполняется неравенство*

$$\left| \alpha - \frac{p_n}{q_n} \right| < \frac{1}{a_{n+1} q_n^2} \leq \frac{1}{q_n^2}.$$

Доказательство. Соседние подходящие дроби находятся по разные стороны от α , поэтому по теореме 15.2.2 получаем

$$\left| \alpha - \frac{p_n}{q_n} \right| < \left| \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} \right| = \frac{|p_{n+1}q_n - p_nq_{n+1}|}{q_nq_{n+1}} = \frac{1}{q_nq_{n+1}} \leq \frac{1}{a_{n+1}q_n^2}. \quad (15.2.1) \quad \square$$

Можно также показать, что

$$\left| \alpha - \frac{p_n}{q_n} \right| > \frac{1}{(a_{n+1} + 2)q_n^2}. \quad (15.2.2)$$

Отсюда видно, что существует тесная связь между величиной a_n и фактической ошибкой. Чуть более тонкие рассуждения позволяют усилить результат предложения 15.2.7, установив для бесконечного числа пар неравенство $|\alpha - p_n/q_n| < 1/(2q_n^2)$; эта оценка всегда выполняется по крайней мере для одного из двух последовательных значений n . Она оптимальна в следующем смысле: для двух взаимно простых чисел p и q неравенство

$$|\alpha - p/q| < 1/(2q^2);$$

выполняется, только если p/q является подходящей дробью числа α .

Наиболее сильный возможный результат в этом направлении таков: для бесконечного числа пар (p, q) выполняется неравенство $|\alpha - p/q| \leq 1/(\sqrt{5}q^2)$. Для дроби золотого сечения

$$\frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \dots}}}$$

достигается равенство.

Таким образом, скорость, обратно пропорциональная квадрату, — это наилучшая возможная для всех чисел скорость приближения, и, чтобы найти наиболее быстрое приближение, надо рассматривать только подходящие дроби. Из формул (15.2.1) и (15.2.2) понятно, что рост последовательности a_n определяет скорость приближения числа α рациональными числами.

Теперь посмотрим, в каких случаях аппроксимация может быть очень быстрой. Возможно, например, что первая ошибка δ будет мала по сравнению с α . Это

означает, конечно, что $\alpha \approx 1/a_1$ — очень хорошее приближение (с ошибкой δ/a_1). Это было бы весьма удачно. При этом значение $a_2 = \lfloor \alpha/\delta \rfloor$ довольно велико. Таким образом, величина a_i в разложении в цепную дробь оценивает качество приближения на предыдущем шаге. (Действительно, из этого следует, что q_n на следующем шаге становится значительно больше, т.е. если приближение является хорошим, то требуется много времени, чтобы его улучшить.) Хорошим примером служит замечательное приближение числа π дробью $p_3/q_3 = 355/113$, которое было известно в Древнем Китае. Оно совпадает с π в шести десятичных знаках, потому что a_4 является неожиданно большим для такого малого числа шагов: оно равно 292. Соответственно значение q_4 подсказывает до 33 102.

Такое быстрое приближение влияет на динамику поворота окружности на угол α . Вернемся к нашему первому примеру, спроектируем его на окружность $\mathbb{R}/\mathbb{Z}$ и возьмем $a_1 = 5$, $\delta \approx 10^{-6}\alpha$. Рассматривая орбиту точки 0, мы увидим, что первые пять итераций равномерно распределяются по окружности (с точностью, определяемой выбором величины δ). Первые 500 000 итераций лежат на расстоянии $1/10$ от первых пяти, поэтому есть 5 интервалов, длины $1/10$ каждый, не содержащих ни одной из первых полумиллиона итераций. Хотя это и не противоречит равномерному распределению (см. предложение 4.1.7), нужно огромное число итераций, чтобы сбалансировать распределение, хотя бы до некоторой степени точности. Имеется возможность, что такие точные приближения будут получаться на любом шаге аппроксимационной схемы, откуда следует, что относительно быстрая аппроксимация может влиять на динамику нежелательным образом. При этом возникает явление *Лиувилля*. Одним из его эффектов является то, что динамика поворота окружности на угол, хорошо приближаемый рациональными числами, гораздо сложнее динамика поворота на угол с небольшими знаменателями в разложении в цепную дробь.

15.2.4. Метрическая теория диофантовых приближений. Задачи, которые мы до сих пор рассматривали, относились к возможной степени приближения подходящими дробями или к степени приближения определенного иррационального числа. В метрической теории рассматривается обратный вопрос: для заданной степени аппроксимации требуется определить, насколько много чисел имеют такую степень? Этот вопрос оказывается важным для динамики. Пусть $r: \mathbb{N} \rightarrow (0, \infty)$ — невозрастающая функция «степени». Тогда число $\alpha \in \mathbb{R}$ называется r -аппроксимируемым, если для бесконечного количества пар $(p, q) \in \mathbb{Z} \times \mathbb{N}$ выполняется условие $|q\alpha - p| \leq r(|q|)$.

Случай, когда степень является обратной пропорциональностью, имеет большое значение как для теории чисел, так и для многих приложений рациональных приближений.

Определение 15.2.8. Диофантово условие с показателем $\beta \geq 0$ и константой $C > 0$ для числа α состоит в том, что

$$|\alpha - (p/q)| \geq C/q^{2+\beta}$$

для любых $p, q \in \mathbb{Z}$, $q \neq 0$. Это значит, что число α не является $C/x^{1+\beta}$ -аппроксимируемым для некоторого $C > 0$. Обозначим множество таких чисел $D_{\beta, C}$.

Число α называется *диофантовым*, если оно удовлетворяет диофантову условию с некоторыми $\beta, C > 0$; в противном случае, т.е. если существуют такие последовательности $p_n, q_n \in \mathbb{Z}$, что

$$\left| \alpha - \frac{p_n}{q_n} \right| = o(q_n^{-\gamma}) \quad \text{для всех } \gamma > 2, \quad (15.2.3)$$

число α называется *лиувиллевым*.

Каждое множество $D_{\beta,C}$ замкнуто и нигде не плотно, но в то же время большинство чисел удовлетворяют диофантову условию с положительным показателем β и константой, зависящей от β : дополнение до множества $\bigcap_{\beta>0} \bigcup_C D_{\beta,C}$ есть множество меры нуль.

Основные результаты о размерах множеств, допускающих аппроксимацию с определенной скоростью, принадлежат А. Я. Хинчину¹. Если ряд $\sum_{n=1}^{\infty} r(n)$ сходится, то множество r -аппроксимируемых чисел имеет меру нуль. Справедливо и обратное утверждение, хотя его и труднее доказать: если ряд расходится, то множество чисел, не являющихся r -аппроксимируемыми, имеет меру нуль.

§ 15.3. Отображение Гаусса

Теоремы Хинчина можно доказать с помощью стандартного, но в то же время весьма искусного метода из действительного анализа: непосредственно изучается структура множеств, допускающих (не допускающих) аппроксимацию с фиксированной скоростью, и строятся подходящие покрытия. Этот метод успешно применяется для определения наилучшей степени аппроксимации различных классов чисел. Более детальное исследование рациональной аппроксимации состоит в выяснении того, насколько часто имеет место быстрая или медленная «наилучшая» аппроксимация. Иными словами, мы хотим знать распределения значений a_n для разных классов чисел, в частности асимптотические статистические свойства этих распределений.

15.3.1. Кодирование разложения в цепную дробь. Вспомним, что распределение цифр в десятичном разложении (или, в более общем виде, в разложении по основанию m) связано с асимптотическим поведением линейного растягивающего отображения E_{10} (E_m). Аналогично разложение для цепной дроби естественным образом связано с итерациями следующего отображения.

Определение 15.3.1. Отображение $G: [0, 1] \rightarrow [0, 1]$, определенное формулой $x \rightarrow \{1/x\}$ (дробная часть), называется *отображением Гаусса* (см. рис 15.3.1).

Это отображение имеет разрывы первого рода в точках $1/i$, $i \in \mathbb{N}$, и разрыв второго рода в точке 0. При проектировании на окружность $S^1 = \mathbb{R}/\mathbb{Z}$ все

¹А. Ya. Khinchine. Einige Sätze über Kettenbrüche, mit Anwendungen auf die Theorie der diophantischen Approximationen // Mathematische Annalen. 1924. Bd. 92. S. 115—125. Имеется перевод: А. Я. Хинчин. Одна теорема о непрерывных дробях и ее арифметические приложения // Избранные труды по теории чисел. М.: МЦНМО, 2005.

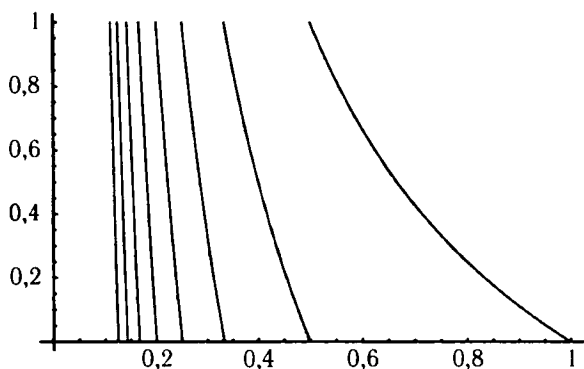


Рис. 15.3.1. Отображение Гаусса

разрывы кроме нуля исчезают, но отображение более естественным образом определено на отрезке, если рассматривать его как динамическую систему.

Если $x \in (0, 1]$, то первый член a_1 в разложении числа x в цепную дробь равен i тогда и только тогда, когда $1/(i+1) < x \leq 1/i$. Аналогично второй член a_2 равен i тогда и только тогда, когда $1/(i+1) < G(x) \leq 1/i$, и т. д. Поэтому распределение чисел a_n в разложении x в цепную дробь совпадает с распределением итераций $G^n(x)$ на полуинтервалах $(1/(i+1), 1/i]$, $i \in \mathbb{N}$.

Отображение Гаусса разрывно, и разрыв в нуле имеет довольно серьезный характер, но оно почти растягивающее там, где оно непрерывно. Уже вторая итерация G^2 — растягивающее отображение, потому что минимум производной больше 1. Поэтому вовсе не удивительно, что отображение Гаусса имеет некоторые свойства, связанные с гиперболическим (в частности, растягивающим) поведением, которые мы изучали в предыдущих главах. Его поведение очень далеко от строгой эргодичности: оно имеет огромное разнообразие инвариантных интегралов и асимптотических распределений, периодические орбиты плотны (они являются квадратичными иррациональностями определенного вида), и возможно большое разнообразие в поведении орбит.

15.3.2. Равномерное распределение. Основной результат об отображении Гаусса, связанный с нашей темой, — это существование равномерного распределения, заданного плотностью, совсем как для многих квадратичных отображений (см. п. 11.4.3.2) или растягивающих и тентообразных отображений (теорема 7.5.6 и п. 11.4.3 соответственно).

Предложение 15.3.2. *Отображение Гаусса сохраняет меру, заданную плотностью $1/(1+x)$.*

Доказательство. Мера отрезка $[a, b]$ равна

$$\int_a^b \frac{1}{1+x} dx = \ln(b+1) - \ln(a+1).$$

Прообраз этого множества получается, если заметить, что $a = \{1/x\}$ тогда и только тогда, когда $1/x = a + n$ для некоторого $n \in \mathbb{N}$ и то же самое верно для b .

Поэтому мера множества $G^{-1}([a, b])$ равна следующей сумме взвешенных длин:

$$\begin{aligned} \sum_{n \in \mathbb{N}} \int_{\frac{1}{b+n}}^{\frac{1}{a+n}} \frac{1}{1+x} dx &= \sum_{n \in \mathbb{N}} \ln\left(1 + \frac{1}{a+n}\right) - \ln\left(1 + \frac{1}{b+n}\right) = \\ &= \sum_{n \in \mathbb{N}} \ln(a+n+1) - \ln(a+n) - \ln(b+n+1) + \ln(b+n) = \ln(b+1) - \ln(a+1). \end{aligned}$$

Меры образа и прообраза совпадают, что и требовалось доказать. $\square$

С небольшими усилиями можно доказать такой результат.

Предложение 15.3.3. *Отображение Гаусса эргодично.*

Следствие 15.3.4. *Множество чисел $\alpha \in (0, 1]$, которые разлагаются в цепную дробь с ограниченными коэффициентами, есть множество меры нуль.*

Доказательство. Поскольку $a_i = [1/G^{(i-1)}(\alpha)]$, это множество есть множество точек, орбиты которых под действием отображения Гаусса отделены от нуля и потому не являются равномерно распределенными. Отсюда ясно, что это множество меры нуль. $\square$

Предложение 15.3.5. *Для почти любого числа (см. определение 7.5.3) коэффициенты соответствующей цепной дроби распределены следующим образом: число $n \in \mathbb{N}$ встречается как коэффициент цепной дроби с асимптотической вероятностью*

$$\frac{\ln\left(1 + \frac{1}{n}\right) - \ln\left(1 + \frac{1}{n+1}\right)}{\ln 2}. \quad (15.3.1)$$

Набросок доказательства. Отображение Гаусса эргодично относительно меры, определенной плотностью $1/(1+x)$. Поэтому орбиты отображения Гаусса равномерно распределены относительно плотности $1/(1+x)$, так что вероятность того, что $a_i = [1/G^{i-1}(\alpha)]$, совпадает с вероятностью того, что выполнено неравенство

$$\frac{1}{n+1} < G^{i-1}(\alpha) \leq \frac{1}{n},$$

которая равна $\int_{1/(n+1)}^{1/n} \frac{1}{1+x} dx / \ln 2$, где $\ln 2 = \int_0^1 (1/(1+x)) dx$ — нормализующая константа. Из этого следует формула (15.3.1). $\square$

15.3.3. Отображение Гаусса и индуцирование для отображений окружности. Приведем теперь другую интерпретацию отображения Гаусса, которое служит примером плодотворного подхода к решению задач динамики. Динамические системы из определенного класса часто можно представлять как элементы некоторого пространства с естественной («глобальной») динамикой на нем. Другими словами, в этом классе существует канонически определенная операция, и асимптотическое поведение орбиты относительно этой операции часто раскры-

вает существенные свойства системы, соответствующей начальному значению. При этом мы часто используем глобальные динамические свойства, связанные с так называемой «ренормализацией». Мы не будем подробно объяснять, что означает это понятие, а проиллюстрируем его на примере.

Отождествим действительное число α , $0 < \alpha < 1$, с поворотом R_α . Разрезав окружность в нуле, мы можем представить вращение как следующее отображение полуинтервала $[0, 1)$:

$$R_\alpha(x) = \begin{cases} x + \alpha, & \text{если } 0 \leq x < 1 - \alpha, \\ x + \alpha - 1, & \text{если } 1 - \alpha \leq x < 1. \end{cases}$$

Геометрически это можно представить себе как перекалывание двух полуинтервалов $[0, 1 - \alpha)$ и $[1 - \alpha, 1)$. Теперь рассмотрим отображение *первого возвращения* на полуинтервале $[0, \alpha)$. Проверка немедленно показывает, что это отображение является перекалыванием двух полуинтервалов $[0, \beta)$ и $[\beta, \alpha)$, где $\beta = \{n\alpha\}$ и целое число n определяется из неравенства $(n - 1)\alpha < 1 \leq n\alpha$, т. е. $n = \left[\frac{1}{\alpha}\right] + 1$, а значит, $\beta = \left[\frac{1}{\alpha}\right]\alpha + \alpha - 1$.

Теперь рассмотрим это перекалывание как представление поворота единичной окружности. Конечно, для этого нам нужно нормализовать полуинтервал $[0, \alpha)$ до длины 1. При этом индуцированное отображение становится представлением поворота на угол $1 - \frac{\beta}{\alpha} = 1 - \left[\frac{1}{\alpha}\right] + 1 + \frac{1}{\alpha} = \left\{\frac{1}{\alpha}\right\} = G(\alpha)$.

Таким образом, итерации отображения Гаусса соответствуют процессу последовательного взятия отображения первого возвращения. Поскольку операция взятия отображения первого возвращения на уменьшающихся полуинтервалах, очевидно, транзитивна, мы можем рассматривать итерации как взятие первого возвращения на специально выбранных полуинтервалах, которые становятся меньше и меньше (длины этих полуинтервалов могут быть легко подсчитаны и близко связаны с разложением в цепную дробь), и «ренормализовать» их обратно к полной длине. Эта процедура раскрывает микроскопическое поведение первоначального поворота, и поэтому динамика отображения Гаусса говорит нам о свойствах индуцирования.

Например, если α — рациональное число, то некоторая итерация $G^n(\alpha)$ становится равной 0, после чего процесс останавливается: индуцированное отображение тождественно, а это в точности означает, что первоначальный поворот периодичен.

Если α — квадратичная иррациональность, то соответствующая орбита отображения Гаусса, начиная с некоторого момента, становится периодичной, и поэтому процесс индуцирования демонстрирует *самоподобие*: начиная с некоторого конечного шага процесс индуцирования становится периодическим, так что в некотором смысле микроскопическая структура на любом уровне та же, что и макроскопическая.

Наконец, для типичного α (лежащего вне некоторого множества меры нуль) орбита отображения Гаусса равномерно распределена относительно плотности $1/(1+x)$. Поэтому индуцированные повороты распределены в соответствии с той

же плотностью, и макроскопическая структура на разных уровнях меняется довольно случайным образом.

Мы уже встречались с процессами ренормализации различного рода в п. 11.3.1. Основное их отличие от рассматриваемого здесь процесса состоит в том, что там мы имели дело со вспомогательной динамикой в окрестности гиперболической неподвижной точки и, таким образом, значительной рекуррентности не возникало. Действительно, применение процесса ренормализации к отдельно отображению делает его все более похожим на отображение с аттрактором типа Фейгенбаума—Мизюревича.

§ 15.4. Однородная динамика, геометрия и теория чисел

Естественный класс динамических систем появляется из следующей общей алгебраической конструкции. Пусть G — локально-компактная группа, $H \subset G$ — ее замкнутая подгруппа и $M = H \backslash G$ — левое однородное пространство, которое в этом случае имеет естественную локально-компактную топологию. Иногда для некомпактных групп G и H однородное пространство оказывается компактным, как, например, в абелевом случае $G = \mathbb{R}^n$, $H = \mathbb{Z}^n$. Поскольку левое и правое умножения в группе коммутируют, действие группы G на себе правыми сдвигами переносится на M . Ограничение этого действия на подгруппу $\Lambda \subset G$ называется (*правым*) *однородным действием*; если Λ — однопараметрическая подгруппа, оно называется (*правым*) *однородным потоком*. Разумеется, рассматривая правое однородное пространство и левые действия, мы определим левое однородное действие. Неудивительно, что некоторые однородные действия играют центральную роль во взаимодействии между теорией чисел, геометрией и динамикой.

15.4.1. Модулярная поверхность, геодезический и орициклические потоки. Мы начнем с описания очень известного однородного действия, близко связанного с цепными дробями и рациональными приближениями. С точки зрения гиперболической геометрии это действие есть геодезический поток на некоторой специальной поверхности постоянной отрицательной кривизны. Соответственно мы начинаем с предварительных геометрических рассуждений.

1. Верхняя полуплоскость. Рассмотрим верхнюю половину комплексной плоскости, определенную условием положительности мнимой части: $\mathbb{H} = \{(x + iy) \in \mathbb{C} : y > 0\}$. Мы определим на ней неевклидово расстояние, положив длину вектора v в точке $(x, y) \in \mathbb{H}$ равной $\|v\|/y$. Можно дать эквивалентное определение, положив длину кривой в $\mathbb{H}$ равной интегралу $\int (1/y)(dx + i dy)$ вдоль кривой. Это определяет *гиперболическую метрику*.

Геодезические в этой метрике легко описать: любая геодезическая является либо вертикальным лучом $\{x + iy : y > 0\}$, либо полуокружностью, конечные точки которой вещественны, т. е. лежат на оси абсцисс.

Геодезический поток определяется, как обычно, на пространстве $S\mathbb{H}$ единичных касательных векторов к $\mathbb{H}$. В отличие от случая сферы или тора, с геодезическими потоками на которых мы встречались ранее (в п. 1.3.3 и 5.2.2)

(и аналогично случаю евклидовой плоскости), это пространство довольно «большое». Оно некомпактно и имеет бесконечный объем. Имеется тем не менее, большая разница между поведением геодезических в евклидовой и неевклидовой геометрии. Чтобы получить менее обширное пространство, мы позднее проведем построение, напоминающее построение из п. 2.6.4, где тор был получен с помощью отождествления точек пространства $\mathbb{R}^2$, являющихся целыми сдвигами друг друга.

2. *Преобразования Мёбиуса.* Группа $SL(2, \mathbb{R})$ действует транзитивно и изометрично на $\mathbb{H}$ с помощью дробно-линейных преобразований, часто называемых *преобразованиями Мёбиуса*:

$$f_g(z) = \frac{az + b}{cz + d}, \quad z \in \mathbb{H}, \quad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{R}),$$

с ядром действия $Z_2 = \{\pm \text{Id}\} \subset SL(2, \mathbb{R})$. Эти преобразования сохраняют определенную выше структуру, а именно они сохраняют гиперболическую метрику (т. е. длины и углы) и (поэтому) переводят геодезические в геодезические. Факторгруппа $G = PSL(2, \mathbb{R}) = SL(2, \mathbb{R})/Z_2 \simeq SO(1, 2)^0$ эффективно действует на $\mathbb{H}$, и стационарная подгруппа точки $z_0 = i \in \mathbb{H}$ есть $C = PSO(2) = SO(2)/Z_2$; поэтому полуплоскость $\mathbb{H}$ можно отождествить с G/C .

Дифференциал действия факторгруппы $PSL_2(\mathbb{R})$ на $\mathbb{H}$ определяет транзитивное и свободное действие на $S\mathbb{H}$, так что последнее пространство можно отождествить с группой G , если зафиксировать любой единичный касательный вектор $\mathbb{H}$ в качестве единичного элемента группы. Удобно выбрать для этой цели направленный вверх вертикальный вектор в точке i .

3. *Фуксовы группы и поверхности постоянной отрицательной кривизны.* Любая гладкая поверхность M постоянной гауссовой кривизны -1 имеет вид $M = \Gamma \backslash \mathbb{H} = \Gamma \backslash G/C$ для некоторой дискретной подгруппы C группы G (такие группы называют *фуксовыми*), не имеющей кручения. Расслоение единичных касательных векторов SM является, таким образом, *левым* однородным пространством $\Gamma \backslash S\mathbb{H} = \Gamma \backslash G$. Форма объема для римановой метрики на M индуцирует G -инвариантную меру на SM , называемую *мерой Лиувилля*, которая, очевидно, совпадает с соответствующим образом нормализованной мерой Хаара на $\Gamma \backslash G$. В частности, поверхность $M = \Gamma \backslash \mathbb{H}$ имеет конечный объем тогда и только тогда, когда Γ — решетка в G (т. е. фактор $SM = \Gamma \backslash G$ имеет конечный объем).

4. *Модулярная поверхность.* Рассмотрим теперь специальную фуксову группу, которая является близким аналогом решетки целочисленных сдвигов в евклидовом пространстве. Это *модулярная группа* $PSL(2, \mathbb{Z})$, состоящая из преобразований

$$z \mapsto \frac{az + b}{cz + d}, \quad \text{где } a, b, c, d \in \mathbb{Z} \text{ и } ad - bc = 1. \quad (15.4.1)$$

В п. 2.6.4 мы получили тор в результате склеивания противоположных сторон фундаментальной области $[0, 1] \times [0, 1]$. В нашей ситуации удобной фундамен-

тальной областью является множество

$$\{x + iy: |x| \leq 1/2, x^2 + y^2 \geq 1\},$$

ограниченное лучами $\{1/2 + iy: y \geq \sqrt{3}/2\}$ и $\{-1/2 + iy: y \geq \sqrt{3}/2\}$ и соединяющей их дугой единичной окружности $\{x + iy: -1 \leq 2x \leq 1, y = \sqrt{1 - x^2}\}$. В случае тора нужно было отождествить две вертикальные стороны фундаментального квадрата, потому что целый сдвиг на $(1, 0)$ переводит левую сторону в правую. Аналогично вертикальные лучи, ограничивающие гиперболическую фундаментальную область, отождествляются с помощью того же сдвига, возникающего, если в (15.4.1) положить $a = b = d = 1, c = 0$. Более того, половинки дуги окружности отождествляются при отображении $z \rightarrow -1/z$ (которое получается, если в (15.4.1) положить $a = -b = c = 1, d = 0$), действующем на этой дуге как зеркальная симметрия. Можно представлять себе описанное отождествление как свертывание области вокруг мнимой оси подобно газете с последующим «застегиванием» нижних полудуг, при котором получившееся множество «закрывается». Топологически результат представляет собой полубесконечный цилиндр, замкнутый снизу. Геометрически, однако, везде за исключением «нижних» точек результат больше похож на псевдосферу: длина отрезка вдоль геодезических, представленных вертикальными линиями, уменьшается экспоненциально. Эти геодезические *параллельны* в неевклидовом смысле: расстояние между любыми двумя из них уменьшается экспоненциально.

Геодезический поток на этой *модулярной поверхности* можно описать как следование вдоль геодезической в описанной области до тех пор, пока геодезическая не встретится с границей области. Тогда геодезическая возвращается обратно в область в другой граничной точке. Для вертикальных отрезков это будет «противоположная» точка, и геодезическая будет продолжаться в том же направлении. При встрече с дугой окружности произойдет прыжок от $x + iy$ к $-x + iy$, сохраняющий угол с границей. См. рис. 15.4.1.

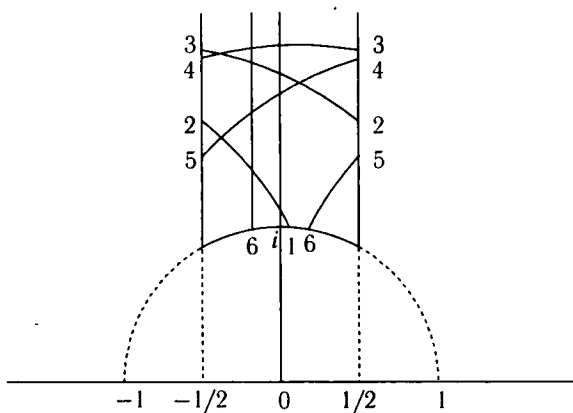


Рис. 15.4.1. Геодезический луч на модулярной поверхности

Геодезический поток имеет сечение, отображение возвращения на которое близко связано с отображением Гаусса (определение 15.3.1). Неудивительно поэтому, что свойства рациональных приближений близко связаны со свойствами геодезического потока на модулярной поверхности.

При отождествлении между SH и $PSL(2, \mathbb{R})$, описанном в п. 15.4.1.2, геодезический поток отождествляется с правым однородным потоком на $PSL(2, \mathbb{R})$, порожденным подгруппой диагональных матриц. Мы кратко упоминаем динамические свойства геодезического потока в п. 15.4.4.

5. Орициклический поток. Существует другой однородный поток на $PSL(2, \mathbb{Z}) \setminus PSL(2, \mathbb{R})$ (или, что эквивалентно, на расслоении единичных касательных векторов к модулярной поверхности), даже более важный для теоретико-числовых рассуждений, чем геодезический поток, и близко связанный с последним. Это правый однородный поток, определенный однопараметрической подгруппой

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \quad t \in \mathbb{R},$$

и называемый *орициклическим потоком*² на модулярной поверхности. Орбиты орициклического потока на самом деле являются неустойчивыми многообразиями геодезического потока (см. п. 15.4.4). Существует однопараметрическое семейство замкнутых орбит, представленных горизонтальными отрезками в фундаментальной области, с вертикальными касательными векторами, которые направлены вверх, так же как и их образы при действии геодезического потока. Однако любая другая орбита плотна и, более того, равномерно распределена относительно меры Лиувилля. Поэтому общая картина ближе к строгой эргодичности, чем к сложному поведению, представленному гиперболичностью³.

15.4.2. Пространство решеток. Естественное обобщение модулярной поверхности, играющее важную роль в приложениях однородной динамики к теории чисел, — однородные пространства $\Omega_k := SL(k/\mathbb{Z}) \setminus SL(k, \mathbb{R})$. Стандартное правое действие $SL(k, \mathbb{R})$ на $\mathbb{R}^k$ сохраняет решетку $\mathbb{Z}^k \subset \mathbb{R}^k$ инвариантной, и для любого данного $g \in G$ решетка $g\mathbb{Z}^k$ *унимодулярна*, т. е. ее фундаментальное множество имеет конечный объем. С другой стороны, любая унимодулярная решетка Λ в $\mathbb{R}^k$ имеет вид $\mathbb{Z}^k g$ для некоторого $g \in G$. Поэтому пространство Ω_k отождествляется с пространством всех унимодулярных решеток в $\mathbb{R}^k$, и это отождествление, очевидно, переводит правое действие группы $SL(k, \mathbb{R})$ на $SL(k, \mathbb{Z}) \setminus SL(k, \mathbb{R})$ в линейное действие на решетках: для решетки $\Lambda \subset \mathbb{R}^k$ мы имеем $g\Lambda = \{gx : x \in \Lambda\}$.

Отсюда, в частности, следует, что касательное расслоение единичных векторов к модулярной поверхности естественно отождествляется с пространством

²Название происходит от слова *орицикл*, обозначающего предел окружностей в гиперболической плоскости, когда центр окружности уходит на бесконечность вдоль геодезической. Орициклы представлены горизонтальными линиями и окружностями, касательными к вещественной оси; геометрически орициклический поток представляет собой движение единичного вектора вдоль одного из двух орициклов, перпендикулярных ему.

³Для такой фуксовой группы Γ , что факторпространство $\Gamma \setminus PSL(2, \mathbb{R})$ компактно, соответствующий орициклический поток является строго эргодичным.

всех решеток евклидовой плоскости, что свидетельствует о фундаментальной связи между евклидовой и неевклидовой геометриями. Аналогично случаю модулярной поверхности однородные пространства Ω_k некомпактны, но имеют конечный объем, инвариантный относительно действия группы G левыми сдвигами. Описанная выше особенность типа «касп» на бесконечности, появляющаяся для модулярной поверхности и легко представимая наглядно, заменяется более сложной асимптотической геометрией.

Основной результат, демонстрирующий связь между теорией чисел и теорией однородных действий, — это следующий критерий Малера, дающий определенное представление об этой асимптотической геометрии.

Теорема 15.4.1. *Последовательность решеток $g_i SL(k, \mathbb{Z})$ стремится к бесконечности в Ω_k тогда и только тогда, когда существует такая последовательность $\{x_i \in \mathbb{Z}^k \setminus \{0\}\}$, что $g_i(x_i) \rightarrow 0$ при $i \rightarrow \infty$. Эквивалентная формулировка такова: зафиксируем норму в $\mathbb{R}^k$ и определим функцию δ на Ω_k , положив*

$$\delta(\Lambda) := \max_{x \in \Lambda \setminus \{0\}} \|x\|. \quad (15.4.2)$$

Подмножество K пространства Ω_k ограничено тогда и только тогда, когда ограничение функции δ на K отделено от нуля.

Правые однородные действия группы $SL(k, \mathbb{R})$ и ее подгрупп на пространстве Ω_k решеток в $\mathbb{R}^k$ очень важны для приложений однородной динамики к теории чисел.

15.4.3. Неопределенные квадратичные формы на решетках. Рассмотрим теперь модельную теоретико-числовую ситуацию, в которой возникают описанные действия; в следующем параграфе мы кратко опишем один из самых замечательных успехов приложения динамических методов к классическим задачам теории чисел. Пусть $Q(x)$ — однородный многочлен от k переменных (например, квадратичная форма). Предположим, что мы хотим исследовать ненулевые векторы x , имеющие целые координаты и такие, что значение $Q(x)$ мало. Обозначим через H стабилизатор формы Q в $SL(k, \mathbb{R})$ (т. е. $H = \{g \in SL(k, \mathbb{R}) : Q(gx) = Q(x) \text{ для всех } x \in \mathbb{R}^k\}$). Тогда можно сформулировать следующую элементарную лемму.

Лемма 15.4.2. *Последовательность ненулевых векторов x_n , имеющих целочисленные координаты и таких, что $Q(x_n) \rightarrow 0$, существует тогда и только тогда, когда существует такая последовательность $h_n \in H$, что $\delta(h_n \mathbb{Z}^k) \rightarrow 0$.*

Последнее условие ввиду критерия Малера равносильно тому, что орбита $H\mathbb{Z}^k$ неограничена в Ω_k . Это придает изучению долгосрочного поведения различных траекторий в Ω_k особое теоретико-числовое значение.

Мы рассматриваем вещественные квадратичные формы от k переменных и их значения в целых точках. Разумеется, если такая форма Q положительно или отрицательно определена, множество $Q(\mathbb{Z}^k \setminus \{0\})$ имеет пустое пересечение с некоторой окрестностью нуля. Теперь возьмем неопределенную форму и назовем ее *рациональной*, если она кратна форме с рациональными коэффициентами,

и *иррациональной* в противном случае. Для $k = 2, 3, 4$ легко построить рациональные формы, которые не принимают малых значений на ненулевых точках с целыми координатами⁴; поэтому естественно предположить, что рассматриваемая форма иррациональна.

15.4.4. Квадратичные формы от двух переменных. Случай $k = 2$ для неопределенных форм является особым. Мы ограничимся тем, что рассмотрим частный, но довольно характерный случай форм $Q_\lambda(x_1, x_2) := x_1^2 - \lambda x_2^2$, где $\lambda > 0$. Заметим, что $x_1^2 - \lambda x_2^2 = (x_1 - \sqrt{\lambda}x_2)(x_1 + \sqrt{\lambda}x_2)$, и, предполагая, что x_1 и x_2 — целые числа одного знака (без потери общности мы можем считать, что они оба положительны), получим, что второй сомножитель должен быть большим. Поэтому для того чтобы произведение было малым, первый сомножитель должен быть мал сравнительно с $(\max(x_1, x_2))^{-1}$. Другими словами, должны существовать такие последовательности натуральных чисел p_n и q_n , что

$$\lim_{n \rightarrow \infty} q_n^2 |\sqrt{\lambda} - p_n/q_n| = 0.$$

В соответствии с (15.2.1) и (15.2.2) это условие эквивалентно неограниченности коэффициентов a_n в разложении числа $\sqrt{\lambda}$ в цепную дробь. Если эти коэффициенты оказываются периодическими, то, как мы знаем, $\sqrt{\lambda}$ является квадратичной иррациональностью, и λ тоже. Но существует несчетное (хотя все еще меры нуль) множество других чисел *ограниченного типа*, т.е. с ограниченными коэффициентами цепной дроби.

Заметим, что динамически такие числа характеризуются тем, что они имеют орбиты, отделенные от нуля, под действием отображения Гаусса, или, что то же самое, имеют компактное замыкание в полуинтервале $(0, 1]$. Стабилизатор формы Q_λ в $SL(2, \mathbb{R})$ является однопараметрической гиперболической подгруппой, которая, стало быть, сопряжена с подгруппой диагональных матриц, а ее действие сопряжено с геодезическим потоком на модулярной поверхности. Этот поток имеет все черты гиперболического поведения, рассмотренные в гл. 10 и очень похожие на свойства отображения Гаусса. В частности, типичные орбиты равномерно распределены и потому неограничены, хотя в то же время существует богатый набор ограниченных орбит. Критерий Малера говорит об ограниченности или неограниченности орбиты с данным начальным условием (стандартной решеткой $\mathbb{Z}^2$), но подходящая замена координат сводит этот вопрос к соответствующему вопросу о геодезическом потоке (или левом однородном действии подгруппы диагональных матриц) с переменным набором начальных значений. Оказывается, если рассматривать не специальную форму Q_λ , а произвольную неопределенную квадратичную форму, это множество начальных значений становится полным. Как и раньше, соответствие можно описать явно. Поэтому можно

⁴Однако по теореме Мейера (см.: *J. W. S. Cassels. An Introduction to Diophantine Approximation. New York: Cambridge University Press, 1957. (Cambridge Tracts in Mathematics and Mathematical Physics, v. 45);* имеется перевод: *Дж. В. С. Касселс. Введение в теорию диофантовых приближений. М.: ИЛ, 1961*) если Q — невырожденная неопределенная рациональная квадратичная форма от $k \geq 5$ переменных, то Q нетривиальным образом представляет нуль над $\mathbb{Z}$, т.е. существует такой ненулевой вектор x с целочисленными координатами, что $Q(x) = 0$.

охарактеризовать квадратичные формы, которые принимают произвольно малые положительные значения на векторах с целыми координатами, как те формы, которые при описанном соответствии отображаются в единичные касательные векторы с неограниченными геодезическими на модулярной поверхности.

§ 15.5. Квадратичные формы от трех переменных

В более высоких размерностях ситуация выглядит совсем по-другому.

15.5.1. Гипотеза Оппенгейма. В 1986 г. Г. А. Маргулис⁵ получил следующий результат (и тем самым доказал гипотезу, выдвинутую за 60 лет до этого А. Оппенгеймом).

Теорема 15.5.1. Пусть Q — действительная неопределенная невырожденная иррациональная квадратичная форма от $k \geq 3$ переменных⁶. Тогда для любого заданного $\varepsilon > 0$ существует такой вектор с целыми координатами $x \in \mathbb{Z}^k \setminus \{0\}$, что $|Q(x)| < \varepsilon$.

Для $k \geq 21$ этот результат был доказан методами аналитической теории чисел в 1950-х гг.⁷. В частности, было известно, что из справедливости гипотезы для некоторых k_0 следует ее справедливость для всех $k > k_0$; другими словами, теорема 15.5.1 сводится к случаю $k = 3$.

Ключевое наблюдение было неявно сделано в 1950-х гг. в работе Касселса и Свиннerton-Дайера и позже в более явном виде в работе Рагунатана (благодаря этому наблюдению Рагунатан и выдвинул свою знаменитую гипотезу). Оно состоит в том, что теорема 15.5.1 эквивалентна определенному утверждению о динамике некоторого однородного действия.

15.5.2. Динамический подход и теорема Маргулиса. Приведем утверждение, о котором идет речь и которое было окончательно доказано Маргулисом.

Теорема 15.5.2. Пусть Q — действительная неопределенная невырожденная квадратичная форма от 3 переменных, H_Q — стабилизатор формы Q в $SL(3, \mathbb{R})$ и Λ — решетка в $\mathbb{R}^3$. Тогда любая орбита $H_Q \Lambda$ правого однородного действия H_Q на Ω_3 , имеющая компактное замыкание, сама компактна.

Идея редукции теоремы 15.5.1 к теореме 15.5.2. Мы воспользуемся леммой 15.4.2 и определенными сведениями из теории решеток в группах Ли. Действительно, предположим, что для некоторого $\varepsilon > 0$ мы имеем $\inf_{x \in \mathbb{Z}^3 \setminus \{0\}} |Q(x)| \geq \varepsilon$.

Тогда по лемме 15.4.2 орбита $H_Q \mathbb{Z}^3$ неограничена в пространстве Ω_3 унимоду-

⁵G. A. Margulis. Formes quadratiques indéfinies et flots unipotents sur les espaces homogènes // C. R. Acad. Sci. Paris Sér. I Math. 1987. V. 304, no. 10. P. 249—253; Indefinite Quadratic Forms and Unipotent Flows on Homogeneous Spaces // Dynamical systems and ergodic theory (Warsaw, 1986). Warsaw: PWN, 1989. P. 399—409.

⁶Сам Оппенгейм формулировал свою гипотезу для $k \geq 5$; позже Давенпорт обобщил формулировку на случай $k \geq 3$.

⁷Об истории этой проблемы можно прочитать в статье: G. A. Margulis. Oppenheim Conjecture // Fields Medallists' lectures. River Edge, NJ: World Sci. Publishing, 1997. P. 272—327.

лярных решеток в $\mathbb{R}^3$; поэтому ввиду теоремы 15.5.2 она компактна. Но поскольку эту орбиту можно отождествить с $H_Q/H_Q \cap SL(3, \mathbb{Z})$, это показывает, что $H_Q \cap SL(3, \mathbb{Z})$ — кокомпактная решетка (дискретная подгруппа с компактным однородным пространством) в H_Q ; стало быть, она плотна по Зарисскому в соответствии с теоремой Бореля о плотности. Последнее, как нетрудно показать, эквивалентно тому, что группа H_Q определена над $\mathbb{Q}$, что, в свою очередь, эквивалентно тому, что форма Q пропорциональна форме с рациональными коэффициентами⁸. $\square$

15.5.3. Жесткость динамики для унитарных действий. Теорема Маргулиса была одним из ключевых результатов, которые позволили глубже понять динамические и статистические свойства некоторого класса однородных действий. Простейшим примером действия из этого класса является орициклический поток, описанный в п. 15.4.1.5. Как и в предыдущем параграфе, мы рассматриваем правые однородные действия на левом однородном пространстве $\Gamma \backslash G$, где G — связная группа Ли и Γ — *решетка*, т. е. такая дискретная подгруппа группы G , что однородное пространство имеет конечный объем, инвариантный при правом действии группы G . Последнее условие, очевидным образом, выполнено, если однородное пространство компактно, но во многих случаях, включая наиболее интересные (например Ω_k), это не так.

Определяющее локальное свойство орициклического потока — это его *унитарность*. Это означает, что в подходящей (подвижной) системе координат производные преобразований выглядят как линейные преобразования с собственными значениями, которые все равны единице (в нашем частном случае орициклического потока отображения, о которых идет речь, имеют единственную жорданову клетку размерности три). Фюрстенберг, Дани и другие в 1970-х гг. показали, что орициклический поток имеет определенные свойства жесткости, начиная со строгой эргодичности в кокомпактном случае. В конце 1970-х гг. это привело Рагунатана к общей гипотезе, согласно которой замыкание орбиты для однородного действия, порожденного унитарными элементами, является алгебраическим в естественном смысле, т. е. является проекцией единственного смежного класса замкнутой подгруппы.

Первоначально прогресс был достигнут для орициклических потоков и более общих унитарных действий, близко связанных с определенными гиперболическими и частично гиперболическими действиями; подобные унитарные подгруппы имеют общее название орисферических. Поскольку гиперболическое действие предоставляет ренормализацию для орисферического действия, эта связь помогает при изучении замыканий орбит и инвариантных мер. В начале 1980-х гг. Марина Ратнер пошла значительно дальше и установила не просто строгую эргодичность, а универсальный набор свойств жесткости для орициклических потоков, их произведений и т. п. В 1986 г. Маргулис установил справедливость гипотезы Рагунатана для определенного класса унитарных подгрупп, не явля-

⁸Детальное доказательство можно найти в книге: A. N. Starkov, Dynamical systems on homogeneous spaces. Providence, RI: AMS, 2000. Имеется перевод: А. Н. Старков. Динамические системы на однородных пространствах. М.: Фазис, 1999.

ющихся орисферическими, который включал в себя подгруппы, порождающие H_Q (см. ниже). Наконец, в 1990 г. Ратнер установила справедливость гипотезы Рагунатана в общем случае⁹, а также соответствующего утверждения об инвариантных мерах¹⁰. Теоремы Ратнер и их следствия имеют многочисленные применения в вопросах, относящихся к теории чисел, геометрии и даже теории групп, которые являются предметом активных исследований до настоящего времени.

Схема доказательства теоремы 15.5.2. Любую неопределенную квадратичную форму от трех переменных унимодулярной заменой координат можно привести к виду $\pm Q_0$, где

$$Q_0(x_1, x_2, x_3) = 2x_1x_2 + x_3^2.$$

Поэтому группа H_Q изоморфна группе H_{Q_0} , содержащей унипотентную однопараметрическую подгруппу

$$\begin{pmatrix} 1 & t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad t \in \mathbb{R},$$

и, как легко показать, порождена унипотентными подгруппами. Замыкание орбиты H_Q в Ω_3 по теореме Маргулиса или общей теореме Ратнер должно быть однородным пространством соответствующей замкнутой подгруппы H группы $SL(3, \mathbb{R})$, где, очевидно, $H_Q \subset H$. Поскольку не существует промежуточной подгруппы между H_{Q_0} и $SL(3, \mathbb{R})$ (а тогда и между H_Q и $SL(3, \mathbb{R})$), мы имеем $H = H_Q$; поэтому действие транзитивно на замыкании орбиты, и орбита компактна. $\square$

⁹*M. Ratner.* Raghunathan's Topological Conjecture and Distributions of Unipotent Flows // Duke Math. J. 1991. V. 63, no. 1. P. 235—280.

¹⁰*M. Ratner.* On Raghunathan's Measure Conjecture // Ann. of Math (2). 1991., V. 134, no. 3. P. 545—607.

ЛИТЕРАТУРА

Для параллельного чтения

Несколько популярных книг могут послужить хорошим дополнением к данной книге. Замечательное описание приложений приводится в книге

S. H. Strogatz. Nonlinear Dynamics and Chaos. Reading, MA: Addison-Wesley, 1994.

Следующие две книги предлагают другой подход к описанию возрастания сложности, основанный на переходе от размерностей 1 и 2 к более высоким размерностям:

R. Devaney. An Introduction to Chaotic Dynamical Systems. Redwood City, CA: Addison-Wesley, 1989.

R. Devaney. A First Course in Chaotic Dynamical Systems. Reading, MA: Addison-Wesley, 1992.

В них также на доступном уровне изложены основы комплексной динамики (множества Жюлиа и др.).

Для дальнейшего чтения

Есть много книг, которые можно использовать для дальнейшего углубленного изучения динамических систем. Для их чтения требуются более глубокие предварительные знания. Ниже мы приводим литературу по соответствующим вопросам.

В качестве наиболее естественного продолжения можно рекомендовать книгу

A. Katok, B. Hasselblatt. Introduction to the Modern Theory of Dynamical Systems. Cambridge: Cambridge University Press, 1995. Имеется перевод: *А. Каток, Б. Хасселблат. Введение в современную теорию динамических систем. М.: Факториал, 1999.*

Эта книга представляет собой весьма полное и систематическое самостоятельное изложение теории динамических систем.

Геометрическая теория динамических систем очень хорошо изложена в книге

C. Robinson. Dynamical Systems, Stability, Symbolic Dynamics, and Chaos. Second edition. Boca Raton, FL: CRC Press, 1999.

Мы также рекомендуем ряд книг, в которых представлены отдельные важные разделы динамики. Для изучения систем с непрерывным временем, механики и приложений полезно обратиться к теории дифференциальных уравнений. Клас-

сический подход к дифференциальным уравнениям с точки зрения динамических систем представлен в книгах

В. И. Арнольд. Дополнительные главы теории обыкновенных дифференциальных уравнений. М.: Наука, 1978.

В. И. Арнольд. Геометрические методы в теории обыкновенных дифференциальных уравнений. М.: МЦНМО, 2002.

Мы также рекомендуем следующую книгу В. И. Арнольда по механике, в которой большое внимание уделяется геометрическим, глобальным и структурным вопросам:

В. И. Арнольд. Математические методы классической механики. М.: УРСС, 2000.

Одномерная динамика систематически и весьма полно изложена в книге *W. de Melo, S. van Strien*. One-Dimensional Dynamics. Berlin: Springer-Verlag, 1993.

Хорошее изложение эргодической теории имеется в книге *K. Petersen*. Ergodic Theory. Cambridge: Cambridge University Press, 1983, 1989.

Символическая динамика и ее интересные приложения очень хорошо изложены в книге

D. Lind, B. Marcus. An Introduction to Symbolic Dynamics and Coding. Cambridge: Cambridge University Press, 1995.

Имеются две книги, представляющие особый интерес для тех читателей второй части нашей книги, которые хотят получить более полное представление о современном состоянии динамических систем. Первая из них написана на основе трудов исследовательского института в Сиэтле летом 1999 г., в котором читались курсы лекций и обзорные доклады, рассчитанные на аспирантов. В нее входит изложение курсов лекций, обзоры и статьи, которые будут весьма интересны для читателя:

Smooth Ergodic Theory and Its Applications: Proceedings of Symposia in Pure Mathematics, Summer Research Institute Seattle, WA, 1999 / A. Katok, R. de la Llave, Y. Pesin, H. Weiss, eds. Providence, RI: AMS, 2001.

Вторая книга содержит обзорение, охватывающее широкий круг вопросов, связанных с новейшими исследованиями в динамике. Она входит в серию справочников по теории динамических систем, в которую войдут не менее четырех книг:

Handbook of Dynamical Systems. Vol. 1A / B. Hasselblatt, A. Katok, eds. Elsevier, 2001.

Литература по основам

Для изучения динамических систем требуется хорошее знание многих других разделов математики. При чтении этой книги можно обращаться к соответствующим учебникам, а при дальнейшем углубленном изучении можно использовать

литературу, приведенную ниже. Краткий обзор разнообразных дисциплин, используемых при изучении динамических систем, приведен в приложении к нашей книге «Введение в современную теорию динамических систем».

Полезно обратиться к книгам

W. Rudin. Principles of Mathematical Analysis. New York—Auckland—Düsseldorf: McGraw-Hill, 1976. Имеется перевод: *У. Рудин*. Основы математического анализа. М.: Мир, 1969.

J. E. Marsden, M. J. Hoffman. Elementary Classical Analysis. New York: W. H. Freeman, 1993.

Есть много книг, посвященных введению в теорию обыкновенных дифференциальных уравнений, среди них, например, книга

В. И. Арнольд. Обыкновенные дифференциальные уравнения. М.: Наука, 1985.

Изучение статистических свойств (эргодической теории) требует знаний теории меры и функционального анализа. Три следующие классические книги содержат результаты, используемые в динамике:

H. L. Royden. Real Analysis. New York: Macmillan, 1988.

W. Rudin. Real and Complex Analysis. New York: McGraw-Hill, 1987.

А. Н. Колмогоров, С. В. Фомин. Элементы теории функций и функционального анализа. 6-е изд. М.: Наука, 1989.

Еще более исчерпывающее изложение теории меры содержится в книге

P. R. Halmos. Measure Theory. New York: Springer-Verlag, 1974. Имеется русский перевод более раннего издания: *П. Халмош*. Теория меры. М.: ИЛ, 1953; М.: Факториал-Пресс, 2005.

Изучение гладкой динамики также требует знаний из области топологии и геометрии. Мы рекомендуем книги

J. W. Milnor. Topology from the Differentiable Viewpoint. Princeton, NJ: Princeton University Press, 1997. Имеется перевод: *Дж. Милнор*. Топология с дифференциальной точки зрения // *Дж. Милнор, А. Уоллес*. Дифференциальная топология. Начальный курс. М.: Мир, 1972.

J. R. Munkres. Elementary Differential Topology. Princeton, NJ: Princeton University Press, 1966. (Annals of Mathematics Studies, v. 54). Имеется перевод: *Дж. Манкрс*. Элементарная дифференциальная топология // *Дж. Милнор, Дж. Сташеф*. Характеристические классы. М.: Мир, 1979.

§ A.1. Метрические пространства

Некоторые интересные динамические системы более естественно рассматривать не в евклидовом пространстве, а в некотором вспомогательном пространстве. Поэтому мы обсуждаем более общее понятие метрического пространства.

A.1.1. Определения.

Определение A.1.1. Пусть X — некоторое множество. Отображение $d: X \times X \rightarrow \mathbb{R}$ называется *метрикой* если выполнены следующие условия:

- 1) $d(x, y) = d(y, x)$ (*симметричность*),
- 2) $d(x, y) = 0 \Leftrightarrow x = y$ (*положительность*),
- 3) $d(x, y) + d(y, z) \geq d(x, z)$ (*неравенство треугольника*).

Если d — метрика, то пару (X, d) называют *метрическим пространством*.

Замечание A.1.2. Подставляя $z = x$ в 3) и используя 1) и 2), получаем $d(x, y) \geq 0$.

Замечание A.1.3. Подмножество метрического пространства само является метрическим пространством с метрикой исходного пространства (она в этом случае называется *индуцированной метрикой*).

Далее обобщим основные понятия, знакомые из евклидова пространства, на произвольное метрическое пространство.

Определение A.1.4. Множество $B(x, r) = \{y \in X: d(x, y) < r\}$ называется *открытым шаром радиуса r с центром в точке x* . Множество $A \subset X$ называется *ограниченным*, если оно содержится в некотором шаре.

Множество $O \subset X$ называется *открытым*, если для любой точки $x \in O$ существует такое $r > 0$, что $B(x, r) \subset O$. (Отсюда немедленно следует, что любое объединение открытых множеств открыто.) *Внутренность* $\text{Int } S$ множества S — это объединение всех открытых множеств, лежащих в S . Иначе говоря, это множество таких точек $x \in S$, что $B(x, r) \subset S$ для некоторого $r > 0$. Если $x \in X$ и O — открытое множество, содержащее точку x , то O называется *окрестностью* точки x . Точка $x \in X$ называется *граничной точкой* множества $S \subset X$, если для любой окрестности U точки x выполняются соотношения $U \cap S \neq \emptyset$ и $U \setminus S \neq \emptyset$. *Граница* ∂S множества S — это множество граничных точек.

Для множества $A \subset X$ множество $\bar{A} := \{x \in X: B(x, r) \cap A \neq \emptyset \forall r > 0\}$ называется *замыканием* множества A . Множество A называется *замкнутым*, если $\bar{A} = A$. Множество $A \subset X$ называется *плотным* в X , если $\bar{A} = X$, и *ε -плотным*,

если $X \subset \cup\{B(x, \epsilon) : x \in A\}$. Множество называется *нигде не плотным*, если его замыкание имеет пустую внутренность (т. е. не содержит непустого открытого множества). Например, на прямой конечное множество является *нигде не плотным*, а $\mathbb{Q}$ и интервал не являются. Последовательность $(x_n)_{n \in \mathbb{N}}$ в X называется *сходящейся* к точке $x \in X$, если для любого $\epsilon > 0$ существует такое $N \in \mathbb{N}$, что для всех $n \geq N$ выполняется неравенство $d(x_n, x) < \epsilon$.

Легко видеть, что множество замкнуто тогда и только тогда, когда его дополнение открыто. (Поэтому любое пересечение замкнутых множеств замкнуто.) Можно дать другое определение замкнутого множества — с помощью предельных точек.

Определение А.1.5. *Предельная точка* множества A — это такая точка x , что $B(x, \epsilon) \cap (A \setminus \{x\}) \neq \emptyset$. Множество предельных точек множества A называется *производным множеством* множества A и обозначается A' . Множество замкнуто, если $A' \subset A$ и для замыкания $\bar{A}$ множества A выполняется равенство $\bar{A} = A \cup A'$. Множество A называется *совершенным*, если $A' = A$, т. е. все предельные точки принадлежат множеству A и в нем нет изолированных точек.

Заметим, что $x \in A'$ тогда и только тогда, когда существует последовательность точек в A , не содержащая x и сходящаяся к x .

Пример А.1.6. Совершенные множества замкнуты. Множества $\mathbb{R}$, $[0, 1]$, замкнутые диски в $\mathbb{R}^n$, S^1 , стандартное канторово множество (см. п. А.1.7) совершенны. Примеры несовершенных множеств — $\mathbb{Z}$, конечное подмножество в $\mathbb{R}^n$ (так как они не содержат предельных точек), множество рациональных чисел $\mathbb{Q}$ (так как иррациональные числа также являются предельными точками).

Как было отмечено выше, на прямой любое конечное множество является *нигде не плотным*, а $\mathbb{Q}$ и интервалы — нет. Стандартное канторово множество *нигде не плотно*, потому что оно замкнуто и не содержит внутренних точек (не содержит интервалов).

Отметим следующий интересный факт (частный случай теоремы А.1.38).

Предложение А.1.7. *Любое ограниченное, совершенное и нигде не плотное подмножество множества $\mathbb{R}$ гомеоморфно стандартному канторову множеству.*

Определение А.1.8. Метрическое пространство X *связно*, если оно не является объединением двух открытых непересекающихся множеств. *Вполне несвязное* пространство — это такое пространство X , что для любых двух точек $x_1, x_2 \in X$ существуют два непересекающихся открытых множества $O_1, O_2 \subset X$, содержащие точки x_1, x_2 соответственно и такие, что $O_1 \cup O_2 = X$.

Множество $\mathbb{R}$, любой интервал в $\mathbb{R}$, $\mathbb{R}^n$, открытый диск в $\mathbb{R}^n$, а также окружность в $\mathbb{R}^2$ связны. Примеры вполне несвязных множеств — конечное подмножество множества $\mathbb{R}$, содержащее по крайней мере 2 элемента, множество рациональных чисел, а также любое счетное подмножество множества $\mathbb{R}$. Стандартное канторово множество — пример несчетного вполне несвязного множества.

А.1.2. Полнота. Имеется важное свойство, которое, например, отличает действительные числа от рациональных. Это свойство — *полнота*, оно отражает то обстоятельство, что вещественная прямая «не содержит дыр», в отличие от множества рациональных чисел. Есть несколько эквивалентных способов формализовать это понятие в случае вещественных чисел, в различных случаях бывают удобны разные определения.

1) Если невозрастающая последовательность действительных чисел ограничена снизу, то она сходится.

2) Если у подмножества $\mathbb{R}$ есть верхняя грань, то есть и точная верхняя грань.

3) Последовательность Коши действительных чисел сходится.

Последовательность a_n является последовательностью Коши, если для любого $\varepsilon > 0$ существует такое $N \in \mathbb{N}$, что для любых $m, n \geq N$ выполняется неравенство $|a_n - a_m| < \varepsilon$.

В первых двух определениях неявно фигурирует упорядоченность множества действительных чисел (понятия верхней грани и невозрастания определены только для упорядоченных множеств). В последнем определении упорядоченность не используется, поэтому именно с его помощью определяют полноту в метрическом пространстве.

Определение А.1.9. Последовательность $(x_i)_{i \in \mathbb{N}}$ называется *последовательностью Коши*, если для любого $\varepsilon > 0$ существует такое $N \in \mathbb{N}$, что для любых $i, j \geq N$ выполняется неравенство $d(x_i, x_j) < \varepsilon$. Метрическое пространство X называется *полным*, если любая последовательность Коши в нем сходится.

Пример А.1.10. Например, $\mathbb{R}$ — полное метрическое пространство, а интервал с обычной метрикой $d(x, y) = |x - y|$ не является полным («не хватает» граничных точек). Однако если мы определим на интервале $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ метрику $d_*(x, y) = |\operatorname{tg} x - \operatorname{tg} y|$, тогда в этой необычной метрике интервал будет полным. Уже не будет «не хватать» граничных точек, потому что последовательности, казалось бы, сходящиеся к граничной точке, не будут последовательностями Коши в новой метрике, так как она растягивает расстояния вблизи граничных точек.

Замечание А.1.11. Это пример *обратного переноса* метрики. Пусть (Y, d) — метрическое пространство и $h: X \rightarrow Y$ — инъективное отображение, тогда $d_*(x, y) = d(h(x), h(y))$ определяет метрику на X . В нашем случае $X = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, $Y = \mathbb{R}$, $h = \operatorname{tg}$.

Лемма А.1.12. *Замкнутое подмножество Y полного метрического пространства X само является полным метрическим пространством.*

Доказательство. Последовательность Коши в Y является последовательностью Коши в X , следовательно, сходится к некоторой точке $x \in X$. При этом $x \in Y$, так как множество Y замкнуто. $\square$

Важным примером является пространство непрерывных функций (см. определение А.1.16).

Теорема А.1.13. Пространство

$$C([0, 1], \mathbb{R}^n) := \{f: [0, 1] \rightarrow \mathbb{R}^n: f \text{ непрерывно}\}$$

— полное метрическое пространство с метрикой, индуцированной нормой $\|f\| := \max_{x \in [0, 1]} \|f(x)\|$ (см. п. А.1.5).

Доказательство. Пусть $(f_n)_{n \in \mathbb{N}}$ — последовательность Коши в $(C[0, 1], \mathbb{R}^n)$. Тогда легко показать, что $(f_n(x))_{n \in \mathbb{N}}$ — последовательность Коши в $\mathbb{R}^n$ для любого $x \in [0, 1]$. Поэтому функция $f(x) := \lim_{n \rightarrow \infty} f_n(x)$ корректно определена ввиду полноты пространства $\mathbb{R}^n$. Для того чтобы доказать, что $f_n \rightarrow f$, зафиксируем $\epsilon > 0$ и найдем такое $N \in \mathbb{N}$, что для любых $k, l \geq N$ выполняется неравенство $\|f_k - f_l\| < \frac{\epsilon}{2}$. Теперь зафиксируем $k \geq N$. Для любого $x \in [0, 1]$ существует такое N_x , что для всех $l \geq N_x$ выполняется неравенство $\|f_l(x) - f(x)\| < \frac{\epsilon}{2}$. Возьмем $l \geq \max(N, N_x)$, тогда $\|f_k(x) - f(x)\| \leq \|f_k(x) - f_l(x)\| + \|f_l(x) - f(x)\| < \epsilon$. Это доказывает утверждение теоремы, потому что k выбрано независимо от x . $\square$

Так же доказывается полнота пространства ограниченных последовательностей.

Теорема А.1.14. Пространство l^∞ ограниченных последовательностей с нормой равномерной сходимости $\|(x_n)_{n \in \mathbb{N}_0}\|_\infty := \sup_{n \in \mathbb{N}_0} |x_n|$ является полным.

Доказательство. Доказательство аналогично, только отрезок $[0, 1]$ заменяется на $\mathbb{N}$. (Ограниченность последовательностей нужна для корректности определения нормы, для непрерывных функций на $[0, 1]$ это свойство выполняется автоматически.) $\square$

Лемма А.1.15 (Теорема Бэра). В полном метрическом пространстве пересечение счетного числа открытых плотных множеств плотно.

Доказательство. Пусть $\{O_i\}_{i \in \mathbb{N}}$ — открытые плотные подмножества X , $B_0 \subset X$ — произвольное непустое открытое подмножество. Построим последовательно шары B_i так, чтобы радиус i -го шара не превышал ϵ/i и выполнялось включение $\bar{B}_{i+1} \subset O_{i+1} \cup B_i$. Центры этих шаров образуют последовательность Коши, которая в силу полноты пространства X имеет предел. Следовательно, множество $\bigcap_i \bar{B}_i$ непусто (поскольку оно содержит этот предел). Поэтому пересечение $B_0 \cap \bigcap_i O_i \supset \bigcap_i \bar{B}_i$ также непусто. $\square$

А.1.3. Непрерывность.

Определение А.1.16. Пусть (X, d) , (Y, d') — метрические пространства. Отображение $f: X \rightarrow Y$ называется *изометрией*, если $d'(f(x), f(y)) = d(x, y)$ для любых $x, y \in X$. Отображение f *непрерывно в точке x* , если для любого $\epsilon > 0$ существует такое $\delta > 0$, что $f(B(x, \delta)) \subset B(f(x), \epsilon)$, или, что эквивалентно, из неравенства $d(x, y) < \delta$ следует, что $d'(f(x), f(y)) < \epsilon$. Отображение f называется *непрерывным*, если оно непрерывно в любой точке $x \in X$. Эквивалентное опреде-

ление: прообраз любого открытого множества открыт. Отображение f называется *равномерно непрерывным*, если выбор δ не зависит от x , т.е. для любого $\epsilon > 0$ существует такое $\delta > 0$, что для любых $x, y \in X$, $d(x, y) < \delta$ выполняется неравенство $d'(f(x), f(y)) < \epsilon$. Отображение f называется *открытым*, если оно отображает открытые множества в открытые.

Непрерывное взаимно однозначное отображение с непрерывным обратным отображением называется *гомеоморфизмом*. Отображение $f: X \rightarrow Y$ называется *липшицевым* с константой Липшица C , если $d'(f(x), f(y)) \leq Cd(x, y)$. Отображение называется *сжатием* (более точно, λ -сжатием), если оно липшицево с константой Липшица $\lambda < 1$.

Из непрерывности не следует, что образ открытого множества открыт. Например, отображение x^2 отображает интервал $(-1, 1)$ и $\mathbb{R}$ в множества, не являющиеся открытыми.

Две метрики могут быть похожи, т.е. эквивалентны, в разных смыслах. Проще всего описать различные понятия эквивалентности, если проследить, как изменяются метрики на одном пространстве, рассмотрев тождественное отображение на X как отображение двух различных метрических пространств.

Определение А.1.17. Мы говорим, что две метрики *изометричны*, если тождественное отображение для них является изометрией. Две метрики *эквивалентны*, если тождественное отображение и его обратное являются липшицевыми отображениями двух метрических пространств. Наконец, две метрики *гомеоморфны*, если тождественное отображение является для них гомеоморфизмом.

А.1.4. Компактность. Важный класс метрических пространств — компактные пространства.

Определение А.1.18. Метрическое пространство (X, d) называется *компактным*, если у любого открытого покрытия множества X есть конечное подпокрытие, т.е. если $\{O_i: i \in I\}$ — такой набор открытых подмножеств множества X , занумерованных элементами множества I , что $X \subset \bigcup_{i \in I} O_i$, то существует такой конечный набор подмножеств $O_{i_1}, O_{i_2}, \dots, O_{i_n}$, что $X \subset \bigcup_{l=1}^n O_{i_l}$.

Предложение А.1.19. *Компактные множества замкнуты и ограничены.*

Доказательство. Пусть X — метрическое пространство и множество $C \subset X$ компактно. Если $x \notin C$, тогда множества $O_n = \left\{ y \in X: d(x, y) > \frac{1}{n} \right\}$ являются открытым покрытием множества $X \setminus \{x\}$, а следовательно, и множества C . Существует конечное подпокрытие $\mathcal{O}$, состоящее из множеств $\{O_n\}_{n \in \mathbb{N}}$. Пусть $n_0 := \max\{n \in \mathbb{N}: O_n \in \mathcal{O}\}$. Тогда для всех $y \in C$ выполняется неравенство $d(x, y) > \frac{1}{n_0}$, следовательно, $x \notin \bar{C}$. Итак, мы доказали, что $\bar{C} \subset C$.

Множество C ограничено, так как открытое покрытие $\{B(x, r): r > 0\}$ содержит конечное подпокрытие. □

Теорема Гейне—Бореля говорит, что в евклидовом пространстве множество компактно тогда и только тогда, когда оно замкнуто и ограничено. В некоторых важных метрических пространствах замкнутые ограниченные множества могут не быть компактными, однако данное выше определение компактности описывает очень важное общее свойство метрического пространства. Метрика используется в этом определении только для того, чтобы определить открытые множества.

При фиксированной метрике компактность эквивалентна полноте и вполне ограниченности.

Определение А.1.20. Метрическое пространство называется *вполне ограниченным*, если для любого $r > 0$ существует такое конечное множество C , что шары радиусов r с центрами в C покрывают пространство.

Предложение А.1.21. *Компактное множество вполне ограничено.*

Доказательство. Если множество C компактно, то для любого $r > 0$ покрытие $\{B(x, r) : x \in C\}$ содержит конечное подпокрытие. $\square$

Предложение А.1.22. Пусть (X, d) , (Y, d') — метрические пространства, множество X компактно и $f: X \rightarrow Y$ — непрерывное отображение. Тогда f равномерно ограничено и множество $f(X) \subset Y$ компактно, а следовательно, замкнуто и ограничено. Если $Y = \mathbb{R}$, то из этого следует, что f достигает максимального и минимального значений.

Последнее замечание о том, что непрерывная вещественнозначная функция достигает максимального и минимального значений, является одним из наиболее часто используемых результатов о компактных пространствах.

Доказательство. Для любого $\epsilon > 0$ существует такое $\delta = \delta(x, \epsilon) > 0$, что если $d(x, y) < \delta$, то $d'(f(x), f(y)) < \frac{\epsilon}{2}$. Шары $B(x, \delta(x, \epsilon)/2)$ покрывают X , поэтому из компактности множества X следует, что существует конечное подпокрытие шарами $B(x_i, \delta(x_i, \epsilon)/2)$. Пусть $\delta_0 = \frac{1}{2} \min\{\delta(x_i, \epsilon)\}$.

Пусть $x, y \in X$, $d(x, y) < \delta_0$. Тогда $d(x, x_i) < \delta_0 < \delta(x_i, \epsilon)$ для некоторого x_i , и по неравенству треугольника $d(x_i, y) \leq d(x_i, x) + d(x, y) < \delta_0 + \delta_0 \leq \delta(x_i, \epsilon)$. Из последних двух неравенств следует, что $d'(f(x), f(y)) \leq d'(f(x), f(x_i)) + d'(f(y), f(x_i)) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$.

Чтобы показать, что множество $f(X) \subset Y$ компактно, рассмотрим открытое покрытие множества $f(X)$: $f(X) \subset \bigcup_{i \in I} O_i$. Тогда множества $f^{-1}(O_i) = \{x : f(x) \in O_i\}$ покрывают X , следовательно, из них можно выбрать конечное подпокрытие $X \subset \bigcup_{l=1}^n f^{-1}(O_{i_l})$. Но тогда $f(X) \subset \bigcup_{l=1}^n O_{i_l}$. $\square$

Предложение А.1.23. Пусть $\{C_i : i \in I\}$ — такой набор компактных множеств в метрическом пространстве X , что $\bigcap_{l=1}^n C_{i_l} \neq \emptyset$ для любого конечного подмножества $\{C_{i_l} : 1 \leq l \leq n\}$. Тогда $\bigcap_{i \in I} C_i \neq \emptyset$.

Доказательство. Проведем доказательство от противного. Пусть $\{C_i: i \in I\}$ — набор компактных множеств, $\bigcap_{i \in I} C_i = \emptyset$. Определим для каждого $i \in I$ множество $O_i = C_1 \setminus C_i$. Из условия $\bigcap_{i \in I} C_i = \emptyset$ следует, что $\bigcup_{i \in I} O_i = C_1$, а значит, множества O_i образуют открытое покрытие множества C_1 . Выберем из него конечное подпокрытие $\bigcup_{i=1}^n O_{i_i} = C_1$. Тогда $\bigcap_{i=1}^n C_{i_i} = \emptyset$. $\square$

Предложение А.1.24. 1) *Замкнутое подмножество компактного множества компактно.*

2) *Пересечение компактных множеств компактно.*

3) *Непрерывное взаимно однозначное отображение между компактными пространствами является гомеоморфизмом.*

4) *Последовательность точек компактного множества содержит сходящуюся подпоследовательность.*

Доказательство. 1) Пусть $C \subset X$ — замкнутое подмножество компактного пространства и $\bigcup_{i \in I} O_i$ — открытое покрытие множества C . Пусть $O = X \setminus C$, тогда $X = O \cup C \subset O \cup \bigcup_{i \in I} O_i$, т.е. $O \cup \bigcup_{i \in I} O_i$ — открытое покрытие пространства X , поэтому оно содержит конечное подпокрытие $O \cup \bigcup_{i=1}^n O_{i_i}$. Так как $O \cap C = \emptyset$, мы получили конечное подпокрытие $\bigcup_{i=1}^n O_{i_i}$ множества C .

2) Пересечение компактных множеств — это пересечение замкнутых подмножеств, следовательно, оно является замкнутым подмножеством любого из компактных множеств. Из утверждения 1) следует, что оно компактно.

3) Нам нужно показать, что образ открытого множества открыт. Используя взаимную однозначность, заметим, что дополнение к образу открытого множества O — это образ дополнения O^c множества O . Но O^c — замкнутое подмножество компактного пространства, следовательно, оно компактно, поэтому его образ компактен, а значит, замкнут. Таким образом, его дополнение, которое является образом множества O , открыто.

4) Пусть дана последовательность $(a_n)_{n \in \mathbb{N}}$ и $A_n := \{a_i: i \geq n\}$ для каждого $n \in \mathbb{N}$. Тогда замыкания $\bar{A}_n$ удовлетворяют условиям предложения А.1.23 и существует $a_0 \in \bigcap_{n \in \mathbb{N}} \bar{A}_n$. Значит, для любого $k \in \mathbb{N}$ существует такое $n_k > n_{k-1}$, что $a_{n_k} \in B(a_0, \frac{1}{k})$, т.е. $a_{n_k} \rightarrow a_0$. $\square$

Интересный пример метрического пространства задается метрикой Хаусдорфа.

Определение А.1.25. Пусть (X, d) — компактное метрическое пространство и $K(X)$ — замкнутые подмножества множества X , тогда *метрикой Хаусдорфа* d_H на $K(X)$ называется функция

$$d_H(A, B) := \sup_{a \in A} d(a, B) + \sup_{b \in B} d(b, A),$$

где $d(x, Y) := \inf_{y \in Y} d(x, y)$ для $Y \subset X$.

Заметим, что функция d_H симметрична по построению и равна 0 тогда и только тогда, когда два множества совпадают (здесь мы используем то, что множества замкнуты, поэтому «sup» на самом деле можно заменить на «max»). Проверка неравенства треугольника потребует немного дополнительной работы. Чтобы показать, что $d_H(A, B) \leq d_H(A, C) + d_H(C, B)$, заметим, что $d(a, b) \leq d(a, c) + d(c, b)$ для $a \in A, b \in B, c \in C$, и, взяв точную нижнюю грань по b , получим $d(a, B) \leq d(a, c) + d(c, B)$ для $a \in A, c \in C$. Поэтому $d(a, B) \leq d(a, C) + \sup_{c \in C} d(c, B)$ и $\sup_{a \in A} d(a, B) \leq \sup_{a \in A} d(a, C) + \sup_{c \in C} d(c, B)$. Аналогично $\sup_{b \in B} d(b, A) \leq \sup_{b \in B} d(b, C) + \sup_{c \in C} d(c, A)$. Складывая два последних неравенства, получаем неравенство треугольника.

Лемма А.1.26. *Метрика Хаусдорфа, определенная на множестве замкнутых подмножеств компактного метрического пространства, задает на нем компактную топологию.*

Доказательство. Нам нужно проверить полноту и вполне ограниченность. Возьмем конечную $\varepsilon/2$ -сеть N . Любое замкнутое множество $A \subset N$ покрывается объединением ε -шаров с центрами в точках из N , и замыкание объединения этих шаров находится на расстоянии не больше ε от A в метрике Хаусдорфа. Поскольку таких объединений конечное число, мы доказали, что эта метрика вполне ограничена. Для доказательства полноты возьмем последовательность Коши (относительно метрики Хаусдорфа) замкнутых множеств $A_n \subset X$. Если мы положим $A := \bigcap_{k \in \mathbb{N}} \bigcup_{n \geq k} A_n$, то легко проверить, что $d(A_n, A) \rightarrow 0$. $\square$

Любой гомеоморфизм компактного метрического пространства f индуцирует естественный гомеоморфизм множества замкнутых подмножеств пространства в метрике Хаусдорфа, откуда выводится следующая лемма.

Лемма А.1.27. *Множество замкнутых подмножеств компактного метрического пространства, инвариантных относительно гомеоморфизма f , замкнуто в метрике Хаусдорфа.*

Доказательство. Это множество есть в точности множество неподвижных точек индуцированного гомеоморфизма, и потому оно замкнуто. $\square$

Определение А.1.28. Метрическое пространство называется *локально компактным*, если для любой точки x и любой ее окрестности O существует компактное множество K в O , содержащее точку x . Пространство называется *сеппарабельным*, если у него есть счетное плотное подмножество (как, например, множество рациональных чисел в $\mathbb{R}$).

А.1.5. Нормы определяют метрики в $\mathbb{R}^n$. Существует особый класс метрик в евклидовом пространстве $\mathbb{R}^n$, которые инвариантны относительно сдвига.

Определение А.1.29. Функция N на линейном пространстве называется *нормой*, если

- 1) $N(\lambda x) = |\lambda|N(x)$ для любого $\lambda \in \mathbb{R}$ (*однородность*),
- 2) $N(x) \geq 0$ и $N(x) = 0 \Leftrightarrow x = 0$ (*положительность*),
- 3) $N(x + y) \leq N(x) + N(y)$ (*выпуклость*).

Линейное пространство с нормой называется *нормированным* линейным пространством.

Любая норма определяет метрику с расстоянием $d(x, y) = N(x - y)$. Для метрики, определенной таким образом, положительность следует из положительности нормы, симметричность из однородности нормы для $\lambda = -1$, неравенство треугольника следует из выпуклости. Для такой метрики сдвиги $T_v: x \rightarrow x + v$ по определению являются изометриями. Кроме того, центральная симметрия $x \rightarrow -x$ является изометрией, а гомотетия $x \rightarrow \lambda x$ увеличивает все расстояния в $|\lambda|$ раз (последнее свойство называется однородностью метрики).

Пример А.1.30. Расстояние в максимум-норме на $\mathbb{R}^n$ задается формулой

$$d(x, y) = \max_{1 \leq i \leq n} |x_i - y_i|. \quad (\text{A.1.1})$$

Конечно, стандартная евклидова метрика также индуцируется нормой (она инвариантна еще и относительно вращений).

Пример А.1.31. Пространство $C[0, 1]$ непрерывных вещественнозначных функций на отрезке $[0, 1]$ — линейное нормированное пространство с нормой $\|f\| := \max\{\|f(x)\|: x \in [0, 1]\}$.

Следующее предложение показывает, почему в динамике так удобно использовать нормы.

Предложение А.1.32. Все метрики в $\mathbb{R}^n$, индуцированные нормами, эквивалентны.

Доказательство. Во-первых, свойство эквивалентности транзитивно, поэтому достаточно показать, что метрика, определенная нормой, эквивалентна стандартной евклидовой метрике.

Во-вторых, так как сдвиги являются изометриями, достаточно рассмотреть расстояния от 0, поэтому мы можем работать непосредственно с нормами.

В-третьих, в силу однородности достаточно рассматривать нормы векторов с евклидовой нормой один, т. е. точки, лежащие на единичной сфере.

Но другая норма — выпуклая функция, следовательно, она непрерывна в евклидовой метрике, поэтому в силу компактности сферы она ограничена сверху. Она также достигает своего минимального значения на единичной сфере. Минимум не может быть нулем, так как из этого следовало бы, что существует ненулевой вектор с нулевой нормой. Поэтому отношение норм ограничено сверху и снизу положительными константами. $\square$

А.1.6. Прямое произведение пространств. Конструкция тора как произведения окружностей показывает, как полезно рассматривать произведения метрических пространств. Чтобы определить произведение метрических пространств (X, d_X) и (Y, d_Y) , нужно ввести метрику на декартовом произведении $X \times Y$, например,

$$d_{X \times Y}((x_1, y_1), (x_2, y_2)) := \sqrt{d_X(x_1, x_2)^2 + d_Y(y_1, y_2)^2}.$$

То, что эта формула определяет метрику, проверяется аналогично тому, что евклидова норма в $\mathbb{R}^2$ определяет метрику.

Существует другие эквивалентные метрики на произведении. Вот две из них:

$$d'_{X \times Y}((x_1, y_1), (x_2, y_2)) := d_X(x_1, x_2) + d_Y(y_1, y_2)$$

и

$$d''_{X \times Y}((x_1, y_1), (x_2, y_2)) := \max(d_X(x_1, x_2), d_Y(y_1, y_2)).$$

То, что эти метрики попарно эквивалентны, доказывается аналогично тому, что евклидова норма, норма $\|(x, y)\|_1 := |x| + |y|$ и максимум-норма $\|(x, y)\|_\infty := \max(|x|, |y|)$ определяют попарно эквивалентные метрики.

Для произведения конечного числа пространств (X_i, d_i) ($i = 1, \dots, n$) можно определить эквивалентные метрики следующим образом: зафиксируем норму $\|\cdot\|$ в $\mathbb{R}^n$ и для любых двух точек $(x_1, x_2, \dots, x_n)$ и $(x'_1, x'_2, \dots, x'_n)$ определим расстояние как норму вектора в $\mathbb{R}^n$ с координатами $d_X(x_i, x'_i)$. То, что полученные метрики эквивалентны, следует из эквивалентности любых двух норм в $\mathbb{R}^n$ (см. предложение A.1.32).

Также рассматривают произведение бесконечного количества метрических пространств (или обычно произведение бесконечного числа копий одного и того же метрического пространства). В бесконечном декартовом произведении множества X на себя каждый элемент определяется своими компонентами, т. е. если копии множества X пронумерованы индексом i , который пробегает множество I , то каждый элемент произведения определяется приписыванием каждому значению i элемента из X , который служит i -й координатой. Это приводит к формальному определению бесконечного произведения $\prod_{i \in I} X =: X^I$ как множества всех функций на I со значениями в X .

В отличие от конечных произведений, надо очень осторожно выбирать метрику. Необходимо не только помнить о сходимости, но и иметь в виду, что можно получить не эквивалентные с точностью до гомеоморфизма метрики. Чтобы определить метрику на произведении, предположим, что множество I счетно. В случае если $I = \mathbb{N}$ и метрика на X ограничена, т. е., например, $d(x, y) < 1$ для любых $x, y \in X$, можно определить гомеоморфные метрики, положив

$$d_\lambda(x, y) := \sum_{i=1}^n \frac{d(x_i, y_i)}{\lambda^{|i|}}. \quad (\text{A.1.2})$$

Для любого $\lambda > 1$ ряд сходится, так как он мажорируется геометрической прогрессией.

Если $I = \mathbb{Z}$, можно дать такое же определение, только суммирование производить по $\mathbb{Z}$ (поэтому в (A.1.2) пишем в показателе степени $|i|$, а не просто i).

Теорема A.1.33 (Тихонов). *Произведение компактных пространств с введенной метрикой компактно.*

В качестве частного случая можно рассмотреть эту конструкцию для единичного отрезка $X = [0, 1]$. Получающееся произведение называется *гильбертовым кирпичом*. Это еще один способ думать о множестве всех последовательностей, элементы которых принадлежат единичному отрезку.

А.1.7. Пространства последовательностей. Обобщая стандартное трюичное канторово множество, введенное в п. 2.7.1, определим более общий класс метрических пространств, включающий много важных пространств.

Определение А.1.34. *Канторовым множеством* называется метрическое пространство, гомеоморфное стандартному канторову множеству.

Простой и естественный пример — пространство Ω_2^R последовательностей $\omega = (\omega_i)_{i=0}^\infty$ с элементами 0 и 1. Это произведение $\{0, 1\}^{N_0}$ счетного числа копий множества $\{0, 1\}$ из двух элементов, поэтому естественно ввести на нем метрику произведения. С точностью до умножения на константу есть только одна метрика на $\{0, 1\}$. Мы определим ее, положив $d(0, 1) = 1$. Возвращаясь к (А.1.2), введем на Ω_2^R метрику произведения

$$d(\omega, \omega') := \sum_{i=0}^{\infty} \frac{d(\omega_i, \omega'_i)}{3^{i+1}}.$$

Предложение А.1.35. *Пространство $\Omega_2^R = \{0, 1\}^{N_0}$, наделенное метрикой произведения $d(\omega, \omega') := \sum_{i=0}^n d(\omega_i, \omega'_i) 3^{-(i+1)}$, является канторовым множеством.*

Чтобы это доказать, нам нужно построить гомеоморфизм между стандартным канторовым множеством C и Ω_2^R .

Лемма А.1.36. *Взаимно однозначное соответствие между стандартным канторовым множеством C и Ω_2^R , устанавливаемое с помощью отображения, которое каждой точке $x = 0, \alpha_1 \alpha_2 \alpha_3 \dots = \sum_{i=1}^{\infty} \frac{\alpha_i}{3^i} \in C$ ($\alpha_i \neq 1$) сопоставляет последовательность $f(x) := \{\alpha_i/2\}_{i=1}^\infty$, является гомеоморфизмом.*

Доказательство. Если точки $x = 0, \alpha_0 \alpha_1 \alpha_2 \dots = \sum_{i=0}^{\infty} \frac{\alpha_i}{3^{i+1}}$ ($\alpha_i \neq 1$) и $y = 0, \beta_0 \beta_1 \beta_2 \dots = \sum_{i=0}^{\infty} \frac{\beta_i}{3^{i+1}}$ ($\beta_i \neq 1$) принадлежат C , то

$$\begin{aligned} d(x, y) = |x - y| &= \left\| \sum_{i=0}^n \frac{\alpha_i}{3^{i+1}} + \sum_{i=0}^n \frac{\alpha_i}{3^{i+1}} \right\| = \\ &= \left\| \sum_{i=0}^n \frac{\alpha_i - \beta_i}{3^{i+1}} \right\| \leq \sum_{i=0}^n \frac{|\alpha_i - \beta_i|}{3^{i+1}} = 2d(f(x), f(y)). \end{aligned}$$

Пусть теперь $\alpha = f(x)$, $\beta = f(y)$. Тогда $d(f^{-1}(\alpha), f^{-1}(\beta)) = d(x, y) \leq 2d(\alpha, \beta)$, поэтому f^{-1} — липшицево отображение с константой Липшица 2.

Если ω, ω' — две такие последовательности, что $d(\omega, \omega') \geq 3^{-n}$, то $\omega_i \neq \omega'_i$ для некоторого $i \leq n$, так как иначе выполнялись бы неравенства

$$d(\omega, \omega') \leq \sum_{i=n+1}^{\infty} 3^{-i-1} = \frac{3^{-n-2}}{1 - \frac{1}{3}} = \frac{3^{-n-1}}{2} < 3^{-n}.$$

Таким образом, $f^{-1}(\omega)$ и $f^{-1}(\omega')$ отличаются в i -м знаке для некоторого $i \leq n$. Отсюда следует, что $d(f^{-1}(\omega), f^{-1}(\omega')) \geq 3^{-(n+1)}$, потому что две точки лежат в различных частях множества C_{n+1} . Взяв $x = f^{-1}(\omega)$, $x' = f^{-1}(\omega')$, получим $d(x, x') < 3^{-(n+1)} \Rightarrow d(f(x), f(y)) < 3^{-n}$. Это показывает, что отображение f липшицево. $\square$

Мы показали, в частности, что пространство Ω_2^R компактно и вполне несвязно. Заметим, кроме того, что каждую последовательность в Ω_2^R можно сколь угодно точно приблизить различными последовательностями из Ω_2^R , меняя лишь очень удаленные элементы. Таким образом, каждая точка из Ω_2^R является предельной, и Ω_2^R — совершенное множество.

Предложение А.1.37. *Канторовы множества компактны, вполне несвязны и совершенны.*

Нетрудно заметить, что пространство $\Omega_2 = \{0, 1\}^{\mathbb{Z}}$ с метрикой произведения тоже гомеоморфно пространству Ω_2^R и поэтому является канторовым множеством. Чтобы это доказать, положим

$$\alpha: \mathbb{Z} \rightarrow N_0, \quad n \mapsto \begin{cases} 2n, & \text{если } n \geq 0, \\ 1 - 2n, & \text{если } n < 0, \end{cases}$$

и $f: \Omega_2^R \rightarrow \Omega_2$, $\omega \mapsto \omega \circ \alpha = (\dots \omega_3 \omega_1 \omega_0 \omega_2 \omega_4 \dots)$. Введя на Ω_2^R и Ω_2 метрики произведения (А.1.2), получаем, что f — гомеоморфизм, так как две последовательности α и α' близки в том и только том случае, если они совпадают в достаточно большом числе первых элементов. Но тогда полученные последовательности $\omega = f(\alpha)$ и $\omega' = f(\alpha')$ совпадают в большом числе элементов, близких к нулю, и поэтому близки. Таким образом, отображение f — непрерывное взаимно однозначное соответствие между двумя компактными пространствами, следовательно, по предложению А.1.24 оно является гомеоморфизмом. (Несложно непосредственно показать, что отображение f^{-1} непрерывно.)

А.1.8. Основные свойства канторовых множеств.

Теорема А.1.38. *Каждое совершенное, вполне несвязное, компактное метрическое пространство является канторовым множеством.*

Мы видели, что пространства последовательностей совершенны и компактны; легко показать, что они вполне несвязны: если $\alpha \neq \beta$ — последовательности, то $\alpha_i \neq \beta_i$ для некоторого индекса i . Пространство таких последовательностей ω , что $\omega_i = \alpha_i$, открыто, и то же самое верно для пространства таких последовательностей, что $\omega_i = \beta_i$. Но эти множества не пересекаются, и объединение есть пространство всех последовательностей.

Следствие А.1.39. *Непустое, совершенное, ограниченное, нигде не плотное множество на прямой является канторовым множеством.*

Доказательство. Совершенное ограниченное множество на прямой компактно по теореме Гейне—Бореля. (Замкнутое подмножество в $\mathbb{R}^n$ компактно.)

Так как оно совершенно, оно содержит более одной точки. Если оно не вполне несвязно, то оно содержит компоненту связности с более чем одной точкой, а следовательно, содержит невырожденный интервал, что противоречит нигде не плотности. $\square$

А.1.9. Двадические целые числа. Определим метрику d_2 на группе $\mathbb{Z}$ целых чисел: $d_2(n, n) = 0$ и $d_2(m, n) = \|m - n\|_2$ для $m \neq n$, где $\|n\|_2 = 2^{-k}$, если $n = 2^k l$, l — нечетное число. Пополнение группы $\mathbb{Z}$ в этой метрике называется группой *двадических целых чисел* и обычно обозначается $\mathbb{Z}_2$. Это компактная топологическая группа.

§ А.2. Дифференцируемость

А.2.1. Производная. Отображение дифференцируемо, если оно допускает хорошую линейную аппроксимацию. Необходимо, чтобы для каждой точки существовало линейное преобразование, которое, как функция от расстояния до рассматриваемой точки, отличалось бы от отображения на величину, меньшую чем линейная.

Определение А.2.1. Пусть V, W — нормированные пространства, множество $U \subset V$ открыто и $x \in U$. Отображение $f: U \rightarrow W$ называется *дифференцируемым* в точке x , если существует такое линейное отображение $A: V \rightarrow W$, что

$$\lim_{h \rightarrow 0} \frac{\|f(x+h) - f(x) - Ah\|}{\|h\|} = 0.$$

В этом случае отображение A называется *производной отображения f* в точке x и мы пишем $Df(x) := A$.

Если отображение $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ дифференцируемо в точке x , то $Df(x)$ — это матрица частных производных в точке x (см. п. 2.2.4.1), но в то же время из существования всех частных производных дифференцируемость не следует.

А.2.2. C^r -топология. Последовательность функций $f_n(x) := \sin(nx)/n$ сходится к 0 равномерно, но последовательность их производных к нулю не сходится. Поэтому если мы хотим гарантировать сходимость производных последовательности функций, необходимо дополнительно потребовать, чтобы выполнялось это условие. Это требование можно красиво формализовать с помощью C^1 -топологии. На пространстве ограниченных функций с ограниченной производной введем метрику

$$d(f, g) := \max_x (d(f(x), g(x)), \sup_x d(Df(x), Dg(x))).$$

Тогда условие $d(f_n, g) \rightarrow 0$ означает, что $f_n \rightarrow g$ и $Df_n \rightarrow Dg$ равномерно. Аналогично C^r -топология определяется метрикой

$$d(f, g) := \max_{0 \leq i \leq r} \sup_x d(D^i f(x), D^i g(x)).$$

Теорема А.2.2. *Пространство ограниченных непрерывных функций со значениями в полном метрическом пространстве с метрикой равномер-*

ной сходимости является полным метрическим пространством. Пространство ограниченных функций с ограниченной производной (и значениями в полном пространстве) также полно в C^1 -топологии. Аналогичное утверждение верно для C' -топологии.

Это обобщение теорем А.1.13 и А.1.14 показывает, почему эти топологии так важны.

А.2.3. Теорема о конечных приращениях и остаточный член для разложения в ряд Тейлора. Теорема о конечных приращениях — один из основных результатов дифференциального исчисления. Она связывает поведение функции на отрезке с ее производной.

Теорема А.2.3. Если функция $f: [a, b] \rightarrow \mathbb{R}$ непрерывна, и дифференцируема на (a, b) , то существует такая точка $x \in (a, b)$, что $f(b) - f(a) = (b - a)f'(x)$.

Доказательство. Функция $g(t) := t(f(b) - f(a)) - f(t)(b - a)$ непрерывна на $[a, b]$ и дифференцируема на (a, b) и $g(a) = af(b) - bf(a) = g(b)$. Если g — константа, то все доказано. В противном случае из непрерывности следует, что функция g имеет экстремум $g(x)$ в некоторой точке $x \in (a, b)$. Функция g дифференцируема в точке x , следовательно, $0 = g'(x) = f(b) - f(a) - f'(x)(b - a)$. $\square$

Усиленный вариант этого результата используется для доказательства возможности разложения в ряд Тейлора.

Теорема А.2.4. Если функция $f: (a, b) \rightarrow \mathbb{R}$ имеет производные до порядка $k + 1$ и $x_0 \in (a, b)$, то для любой точки $x \in (a, b)$ существует такая точка c , лежащая между x и x_0 , что

$$f(x) = \sum_{i=0}^k \frac{f^{(i)}(x_0)}{i!} (x - x_0)^i + \frac{f^{(k+1)}(c)}{(k+1)!} (x - x_0)^{k+1},$$

где $f^{(i)}$ — производная порядка i .

Доказательство. Пусть $f_k := \sum_{i=0}^k \frac{f^{(i)}(x_0)}{i!} (x - x_0)^i$, $z := (f(x) - f_k(x)) / (x - x_0)^{k+1}$ и $g(t) := f(t) - f_k(t) - z(t - x_0)^{k+1}$ на (a, b) .

Покажем, что $g^{(k+1)}(c) = 0$ для некоторой точки c , лежащей между x и x_0 . Так как $g^{(k+1)}(t) = f^{(k+1)}(t) - (k+1)!z$, мы получаем $f^{(k+1)}(c) = (k+1)!z$, что и требовалось.

Будем использовать равенство $g^{(i)}(x_0) = 0$ для $0 \leq i \leq k$, которое выполнено, так как $f^{(i)}(x_0) = f_k^{(i)}(x_0)$ по определению. Так как и $g(x) = 0$ (благодаря выбору z), существует такая точка c_1 , лежащая между x и x_0 , что $g'(c_1) = 0$. Так как и $g'(x_0) = 0$, существует такая точка c_2 , лежащая между c_1 и x_0 , что $g''(c_2) = 0$. Повторяя аналогичные рассуждения k раз, получим искомую точку c . $\square$

А.2.4. Диффеоморфизмы и вложения. Обратное отображение к дифференцируемому отображению не обязательно дифференцируемо, как показывает

пример отображения x^3 . Дифференцируемость обратного отображения — полезное свойство, поэтому у отображений, обладающих этим свойством, есть название: дифференцируемое отображение с дифференцируемым обратным называется диффеоморфизмом.

Для наших целей полезно расширить это понятие на отображения, не являющиеся сюръективными (отображениями «на»). Мы хотим рассматривать такие отображения, как отображение $(x, y) \rightarrow (x, y, x^2 + y^2)$ из единичного диска в $\mathbb{R}^3$, но исключить, скажем, отображение $t \rightarrow (t, \pi t) \pmod{1}$ из $\mathbb{R}$ в тор (см. п. 2.6.4), потому что «обратное» к нему отображение не непрерывно.

Определение А.2.5. Пусть $U \subset \mathbb{R}^n$. Отображение $f: U \rightarrow \mathbb{R}^m$ называется *вложением*, если оно дифференцируемо, его производная имеет ранг n в каждой точке и $f: U \rightarrow f(U)$ — гомеоморфизм.

В этом определении можно заменить одно или оба евклидовых пространства на тор, цилиндр или сферу соответствующей размерности.

§ А.3. Интегрирование по Риману в метрических пространствах

Понятие интегрирования по «мере» неоднократно появляется в этой книге.

А.3.1. Интеграл Римана. Первоначальное понятие, которое мы используем, — это интегрирование по Риману по длине, площади или объему в пространстве и его подпространствах и таких связанных с ним пространствах, как сферы, цилиндры и торы. Важный вопрос состоит в том, какие функции интегрируемы по Риману. При стандартном определении через верхние и нижние суммы, очевидно, необходима ограниченность. Функция также должна иметь *компактный носитель*, т. е. она должна обращаться в нуль вне компактного множества. Это ограничение несущественно, если само объемлющее пространство компактно, как, например, отрезок, прямоугольник, сфера или тор. В этих предположениях каждая непрерывная функция интегрируема. Однако многие функции, возникающие в связи с интегрированием, разрывны. Во-первых, это характеристические функции отрезков, прямоугольников и других «хороших» множеств, для которых интеграл эквивалентен длине, площади или объему, в зависимости от размерности.

Оказывается, есть мощное необходимое и достаточное условие интегрируемости.

Теорема А.3.1 (Лебег). *Функция, определенная в компактной области евклидова пространства, на сфере, торе или на аналогичном компактном дифференцируемом многообразии, интегрируема по Риману тогда и только тогда, когда она ограничена и множество точек разрыва имеет меру нуль.*

Главная идея доказательства состоит в том, чтобы связать счетное число множеств в определении множества меры нуль (см. определение 7.5.3) с конечным числом прямоугольников в верхней и нижней суммах. Для этого нужно заметить,

что множество точек, вблизи которых значения функции f отличаются по крайней мере на ϵ , может быть покрыто счетным числом прямоугольников, сумма объемов которых может быть сколь угодно малой, и в силу компактности можно выбрать конечное подпокрытие, которое будет частью допустимого разбиения¹.

Применяя этот критерий к характеристической функции компактного множества A (точки разрыва которой — это граничные точки множества A), сразу получаем такой результат.

Следствие А.3.2. *Для компактного множества A длина, площадь или объем определены тогда и только тогда, когда граница множества A имеет меру нуль.*

Это следствие легко распространяется на множества с компактным замыканием.

А.3.2. Интегрирование с весом. Естественным обобщением интегрирования по Риману является интегрирование с весом с ненулевой плотностью ρ , которое можно свести к стандартному случаю просто умножением подынтегрального выражения на ρ . Конечно, это можно сделать, только если сама функция ρ интегрируема по Риману. Это условие также является достаточным, как видно из следующего результата.

Предложение А.3.3. *Сумма, произведение и равномерный предел интегрируемых по Риману функций интегрируемы по Риману. Кроме того, интеграл линейной комбинации (или предела) — это соответствующая линейная комбинация (или предел) интегралов.*

Понятие интеграла может быть расширено. Во многих учебниках по математическому анализу и элементарной теории функций вещественной переменной рассматриваются ситуации (которые часто возникают в конкретных задачах), когда или носитель функции не компактен (например, функция $f(x) = \frac{1}{1+x^2}$ на действительной прямой), или функция не ограничена (например, функция $f(x) = \ln x$ на полуинтервале $(0, 1]$), или и то, и другое. В таких случаях естественная процедура аппроксимации часто позволяет определить интеграл, называемый *несобственным интегралом*.

А.3.3. Интеграл Римана—Стилтьеса на прямой. Интегрирование по мере Бернулли (см. п. 7.6.4) естественно определено, но его нельзя свести ни к одной из ситуаций, описанных выше. Интегрирование по любой несимметрической мере Бернулли нельзя свести к интегрированию по Риману с весом, так как множества меры нуль могут иметь положительную меру Бернулли (это свойство называется *сингулярностью меры*). Однако интегрирование по мере Бернулли можно определить с помощью знакомой процедуры нахождения верхних и нижних сумм Римана по разбиению отрезка $[0, 1]$ на маленькие отрезки и взятия их общего предела при длине наибольшего элемента разбиения, стремящейся к нулю.

¹ См.: J. E. Marsden, M. J. Hoffman. Elementary Classical Analysis. New York: W. H. Freeman, 1993.

Общая конструкция такого вида называется *интегралом Римана—Стилтьеса*. Он определяется для функций, заданных на отрезке $I \subset \mathbb{R}$. Он зависит от функции распределения F на I . Это монотонная, ограниченная сверху и снизу и непрерывная слева функция, т. е. для возрастающей последовательности $x_n \in I$ выполняется соотношение $\lim_{n \rightarrow \infty} F(x_n) = F(\lim_{n \rightarrow \infty} x_n)$.

Сначала рассмотрим случай непрерывной функции F ; это не является существенным ограничением для динамических систем. Определим меру отрезка $[a, b] \subset I$ как $F(b) - F(a)$. Используя эту меру вместо длины, определим верхние и нижние суммы Римана для функции по конечному разбиению отрезка I . В этом случае нет отличия между интервалами, отрезками или полуинтервалами.

В общем случае, т. е. если есть точки разрыва, сначала рассматривают интервалы, концы которых — это точки непрерывности функции F , и определяют меру аналогичным образом. Чтобы избежать двусмысленности при определении сумм Римана, необходимо ограничиться разбиениями, в которых точки деления — точки непрерывности функции F , потому что точки разрыва не являются множествами меры нуль: мера каждой такой точки x определяется скачком функции распределения в точке x . Доказательство того, что непрерывные функции интегрируемы (т. е. верхние и нижние суммы Римана имеют одинаковый предел для любой последовательности разбиений со стремящейся к нулю длиной максимального отрезка), совершенно аналогично доказательству для стандартного интеграла Римана.

Конструкция с функцией распределения обеспечивает наиболее общий подход к интегрированию в размерности один; в действительности она может быть расширена на случай всей прямой или полупрямой. В частности, она подходит для «хороших» неограниченных функций плотности, которые появляются, например, как инвариантные меры некоторых квадратичных отображений, таких как отображения, рассмотренные в п. 11.4.3.

А.3.4. Интегралы как положительные функционалы. Однако при рассмотрении гиперболических систем и странных аттракторов мы сталкиваемся со случаями размерности более чем один, в которых асимптотическое распределение существует, но не может быть сведено к интегрированию по плотности. Кроме того, некоторые естественные системы, рассматриваемые в этой книге, такие как символические системы, действуют в пространствах, сильно отличающихся от евклидовых. В оставшейся части параграфа мы обсудим структуру интегрирования в метрических пространствах, которая, в частности, включает эти ситуации. В сущности, мы опишем переход от интеграла, определенного на непрерывных функциях, к мере «хороших» множеств и снова к интегралу, как мы делали для равномерного распределения на окружности в п. 4.1.4 и 4.1.6.

Определение А.3.4. Пусть X — компактное метрическое пространство. *Интеграл Рисса* — это ненулевой линейный функционал I на пространстве $C(X)$ непрерывных вещественнозначных функций на X , непрерывный в равномерной топологии и неотрицательный, т. е. $I(f) \geq 0$, если $f \geq 0$.

Интеграл Римана с весом $J(f) = \int \rho f dx$, а также любой интеграл Римана—Стилтьеса удовлетворяют этим условиям.

Интеграл Рисса определен на характеристической функции χ_A , только если функция χ_A непрерывна, т.е. множество A одновременно открыто и замкнуто. Это верно для цилиндрических множеств в пространстве последовательностей, но невозможно для связных пространств. Мы расширим определение интеграла Рисса на некоторые характеристические функции и многие другие функции с помощью аппроксимации.

Определение А.3.5. Для функции $f: X \rightarrow \mathbb{R}$ определим *верхний интеграл*

$$J^+(f) := \inf\{J(g): g \in C(X), g \geq f\}$$

и аналогично *нижний интеграл*

$$J^-(f) := \inf\{J(g): g \in C(X), g \leq f\}.$$

Функция f *интегрируема*, если $J^+(f) = J^-(f)$. В этом случае общее значение $J^+(f)$ и $J^-(f)$ обозначается $J(f)$.

Очевидно, что линейная комбинация интегрируемых функций интегрируема. Не столь очевидно, но все же легко доказывается, что произведение интегрируемых функций интегрируемо и равномерный предел интегрируемых функций интегрируем. Следующее предложение показывает, насколько интегрируемые функции распространены среди характеристических функций.

Предложение А.3.6. Если $x \in X$, то для любого значения r , кроме, быть может, счетного числа, характеристические функции и открытого, и замкнутого шаров радиуса r с центром в точке x интегрируемы.

Определение А.3.7. Множества, характеристические функции которых интегрируемы, называются *измеримыми (по Риману)* в соответствии с J . Величина $J(\chi_A)$ называется *мерой* множества A .

Предложение А.3.8. Любое конечное объединение или конечное пересечение измеримых множеств измеримо.

Замечание А.3.9. В этом определении есть явный недостаток, который трудно заметить в случае стандартного интегрирования по Риману. А именно, некоторые «хорошие» множества могут не быть измеримыми. Простой пример появляется для δ -меры: $\delta_{x_0}(f) = f(x_0)$ для фиксированной точки $x_0 \in X$. Очевидно, что это интеграл Рисса и что он «сосредоточен» в точке x_0 , так как он обращается в нуль на каждой функции, носитель которой не содержит точку x_0 . Однако множество $\{x_0\}$, состоящее из одной точки, не измеримо по нашему определению. Чтобы решить эту проблему, нужно расширить понятие интегрируемости и измеримости. Это приводит к теории интегрирования *по Лебегу*.

А.3.5. Разбиения и суммы Римана. Теперь мы покажем, как переопределить интеграл Рисса, используя процедуру, сходную с построением стандартного интеграла Римана с помощью разбиений на прямоугольники и сумм Римана.

Определение А.3.10. Пусть дан интеграл Рисса I ; *измеримое разбиение множества X* — это разбиение множества X в объединение конечного числа измеримых множеств. *Диаметр* измеримого разбиения — это точная верхняя грань расстояний между двумя точками в одном и том же множестве.

Предложение А.3.11. Для любого ϵ существует измеримое разбиение с диаметром меньше ϵ .

Доказательство. По предложению А.3.6 для любой точки существует шар радиуса меньше чем $\epsilon/3$ с центром в этой точке, характеристическая функция которого измерима. Возьмем конечное покрытие $B_1, \dots, B_n$ такими шарами и положим $C_k = B_k \setminus \bigcup_{i=1}^{k-1} B_i$, $k = 1, \dots, n$. Множества $C_1, \dots, C_n$ образуют разбиение с нужным свойством. $\square$

Пусть даны ограниченная функция f на X и измеримое разбиение $\xi = (C_1, \dots, C_n)$ на X . Определим *верхнюю и нижнюю суммы Римана* как

$$U(f, \xi) := \sum_{i=1}^n \text{mes}(C_i) \sup\{f(x) : x \in C_i\}$$

и

$$L(f, \xi) := \sum_{i=1}^n \text{mes}(C_i) \inf\{f(x) : x \in C_i\}$$

соответственно.

Теорема А.3.12. Если функция f интегрируема и ξ_m — последовательность разбиений с диаметром, стремящимся к нулю, то

$$\lim_{m \rightarrow \infty} U(f, \xi_m) = \lim_{m \rightarrow \infty} L(f, \xi_m) = I(f).$$

Доказательство. Сначала рассмотрим непрерывную функцию f , и пусть $\epsilon > 0$. Так как функция f равномерно непрерывна, существует такое число $N \in \mathbb{N}$, что для любого $m > N$ и любого элемента C разбиения ξ_m выполняется неравенство $\sup\{f(x) : x \in C\} - \inf\{f(x) : x \in C\} < \epsilon$. Так как мера аддитивна, $U(f, \xi_m) - L(f, \xi_m) < \epsilon$.

Для произвольной интегрируемой по Риману функции f существуют такие непрерывные функции f^+ и f_- , что $f_- \leq f \leq f^+$ и $I(f^+) - I(f_-) < \epsilon$. Применяя предыдущее рассуждение к этим функциям, получаем утверждение теоремы. $\square$

А.3.6. Общий интеграл Римана. В заключение покажем, что интеграл Рисса возникает из меры, определенной на достаточном наборе множеств, с помощью конструкции сумм Римана. По аналогии с конструкцией классического интеграла Римана, будем называть эти множества «прямоугольниками». Мы предполагаем, что задан набор прямоугольников, а также и мера каждого прямоугольника, и считаем X компактным метрическим пространством.

Достаточность. Все пространство — прямоугольник.

Продолжение. Для любого $\epsilon > 0$ прямоугольник можно разбить на конечное число прямоугольников, каждый из которых покрывается шаром радиуса ϵ .

Пересечение. Пересечение двух прямоугольников — прямоугольник.

Аддитивность. Если прямоугольник R разбит на конечное число прямоугольников R_i ($i = 1, \dots, k$), то $\text{mes}(R) = \sum_{i=1}^k \text{mes}(R_i)$.

Пусть f — ограниченная вещественнозначная функция, определенная на ограниченном прямоугольнике A . Для каждого разбиения $\mathcal{R} = \{R_1, \dots, R_k\}$ (т. е. $A = \bigcup_{i=1}^k R_i$) на прямоугольники определим верхнюю сумму

$$U(f, \mathcal{R}) := \sum_{i=1}^k \text{mes}(R_i) \sup\{f(x) : x \in R_i\}$$

и нижнюю сумму

$$L(f, \mathcal{R}) := \sum_{i=1}^k \text{mes}(R_i) \inf\{f(x) : x \in R_i\}.$$

Лемма А.3.13. Пусть $\mathcal{R}, \mathcal{R}'$ — разбиения, тогда выполняется неравенство $L(f, \mathcal{R}') \leq U(f, \mathcal{R})$.

Доказательство. Если $\mathcal{R} = \mathcal{R}'$, то утверждение очевидно. Чтобы свести доказательство к этому случаю, воспользуемся *общим измельчением* $\bar{\mathcal{R}} := \{R \cap R' : R \in \mathcal{R}, R' \in \mathcal{R}'\}$. Это разбиение на прямоугольники, и, используя аддитивность, легко проверить, что

$$L(f, \mathcal{R}') \leq L(f, \bar{\mathcal{R}}) \leq U(f, \bar{\mathcal{R}}) \leq U(f, \mathcal{R}). \quad \square$$

Из леммы следует, что величины

$$\int_A^{\bar{}} f = \inf_{\mathcal{R}} U(f, \mathcal{R}) \quad \text{и} \quad \int_A f = \sup_{\mathcal{R}} L(f, \mathcal{R})$$

корректно определены и конечны и $\int_A^{\bar{}} f \geq \int_A f$.

Определение А.3.14. Функция f , определенная на прямоугольнике A , называется интегрируемой по Риману по A , если $\int_A^{\bar{}} f = \int_A f$. В этом случае величина $\int_A f := \int_A f$ называется интегралом Римана функции f по A .

Используя свойство продолжения и рассуждения, напоминающие доказательство теоремы А.3.12, мы получим, что непрерывные функции интегрируемы по Риману и интеграл Римана, определенный таким образом на $C(X)$, — это интеграл Рисса (см. определение А.3.4).

ОТВЕТЫ И УКАЗАНИЯ

Упражнение 1.2.5. $kT = -\ln 2$.

Упражнение 1.3.3. Семь шагов. После шестого шага получится приближение, довольно близкое к значению 27, с которого начинал Герон.

Упражнение 1.3.8. В упражнении 1.1.5 решается уравнение $\cos x = x$, в упражнении 1.1.8 находится $\sqrt{5}$, а в упражнении 1.1.9 решается уравнение $\sin x = x$.

Упражнение 1.3.23. Поскольку числа, образованные последними двумя цифрами, повторяются с периодом 20, достаточно умножать 8 на 2^{20} , пока опять не встретится 008. По дороге полезно избавляться от больших чисел (поскольку нас интересуют только последние три цифры).

Упражнение 2.2.6. Используйте «графические вычисления» (см. замечание 2.3.3, и рис. 2.3.1) для среднего из рисунков 2.3.2.

Упражнение 2.2.7. Последние два.

Упражнение 2.2.12. Рассмотрите в качестве U кольцо с центром в нуле и небольшим разрезом (что-то вроде буквы С). Отображение задайте в полярных координатах так, чтобы оно «сжимало» и радиальную, и угловую составляющие.

Упражнение 2.2.13. Докажите, что функция $d(f(x), x)$ достигает минимума и точка минимума является неподвижной.

Упражнение 2.2.14. $f(x) = x + e^{-x}$.

Упражнение 2.3.4. Определите отображение на интервалах, составляющих дополнение к E , таким образом, чтобы каждая точка сдвигалась вправо.

Упражнение 2.3.5. Можно использовать конструкцию из предыдущего упражнения, но нужно проявлять большую аккуратность. Во-первых, на каждом интервале функция f должна быть бесконечно дифференцируемой, причем все производные разности $(f - \text{Id})$ должны стремиться к нулю при приближении к концам интервалов. Во-вторых, нужно следить за поведением производных при уменьшении длин интервалов. Другими словами, функция, равная разности f и тождественного отображения, должна быть очень «плоской» вблизи множества E .

Упражнение 2.4.2. После деления переменных и интегрирования получается $s = k - 1$.

Упражнение 2.4.3. В этом примере должно нарушаться условие Липшица для функции f . Рассмотрите уравнение $\dot{x} = \sqrt{x}$ и решение $x(t) = 4t^2$.

Упражнение 2.4.4. $\dot{x} = x^2$.

Упражнение 2.4.7. Обратное отображение к сечению может иметь вид $x \rightarrow x^3$.

Упражнение 2.5.1. Докажите, что функция f_λ^2 выпукла на рассматриваемом интервале, чтобы ограничить производную сверху ее значением в точке x_λ .

Упражнение 2.6.1. Докажите, что $B(y, r - d(x, y)) \subset B(x, r)$ при $y \in B(x, r)$.

Упражнение 2.6.8. См. лемму A.1.12.

Упражнение 2.6.9. В силу компактности X функция $d(x, f(x))$ достигает минимума в некоторой точке x_0 . Выведите из указанного в условии свойства отображения f , что минимум равен нулю и неподвижная точка единственна. *Сходимость:* для любой точки $x \in X$ последовательность $(f^n(x))_{n \in \mathbb{N}}$ имеет предельную точку x' . Докажите, что точка $f(x')$ тоже является предельной и что случай $x' \neq x_0$ противоречит тому, что $d(f^n(x), x_0)$ уменьшается с ростом n .

Упражнение 2.7.5. Проверьте, что отображение $0, \alpha_1 \alpha_2 \alpha_3 \dots \mapsto (0, \alpha_1 \alpha_3 \dots, 0, \alpha_2 \alpha_4 \dots)$ (все разложения троичные, $\alpha_i \in \{0, 2\}$) является гомеоморфизмом.

Упражнение 2.7.6. См. п. 4.4.1.

Упражнение 2.7.7. Постройте гомеоморфизм отрезка $[0, 1]$ в себя, переводящий C' в C . Для этого нужно установить соответствие между интервалами, из которых состоят дополнения к C и C' , сохраняя порядок и выбирая каждый раз самый длинный из еще не использованных интервалов.

Упражнение 3.1.4. Решения имеют вид $x\lambda^n + yn\lambda^n$.

Упражнение 3.1.5. Обозначим количество способов через a_n . Рассматривая самый левый прямоугольник в покрытии, получим рекуррентное соотношение $a_n = a_{n-1} + a_{n-2}$.

Упражнение 3.2.3. Перепишите дифференциальное уравнение в полярных координатах и разделите переменные.

Упражнение 3.3.2. Используйте предложение A.1.32.

Упражнение 3.3.3. Используйте предложения 3.3.3 и A.1.32.

Упражнение 3.3.5. Рассмотрите модули собственных значений. Их сумма не меньше 2,7, поэтому если все они не превосходят единицы, то произведение не меньше 0,7.

Упражнение 3.3.6. Используйте жорданову нормальную форму.

Упражнение 3.3.7. Рассмотрите сначала жордановы клетки.

Упражнение 4.1.5. Докажите, что промежуток времени между восходом Солнца и восходом Луны каждый раз изменяется на одно и то же число. Таким образом, продолжительности этих промежутков образуют орбиту поворота. Выведите из условия, что угол поворота должен быть иррационален.

Упражнение 4.1.6. Рассмотрите множество $\mathbb{Z}$ в качестве пространства.

Упражнение 4.1.7. Рассмотрите пространство $\{-1, 1\} \cup \{\frac{1}{n} - 1 : n \in \mathbb{N}\} \cup \{1 - \frac{1}{n} : n \in \mathbb{N}\}$ и модифицируйте конструкцию из предыдущего упражнения.

Упражнение 4.1.9. Временные средние сходятся к значению в неподвижной точке.

Упражнение 4.1.11. 292; см. предложение 15.2.7.

Упражнение 4.1.14. Временные средние сходятся к значению в нуле.

Упражнение 4.2.5. Угол поворота равен $1/\gamma$.

Упражнение 4.2.10. Рассмотрите отдельно «рациональный» и «иррациональный» случаи. В первом случае имеется конечное число наиболее освещенных точек; во втором случае эти точки заполняют окружность.

Упражнение 4.3.4. F является поднятием гомеоморфизма окружности, сохраняющего ориентацию. Поскольку $F(0) = 0$, получаем $\rho(F) = 0$.

Упражнение 4.3.5. Заметьте, что при доказательствах для гомеоморфизмов использовались только непрерывность и монотонность.

Упражнение 4.3.6. Число вращения изменит знак.

Упражнение 4.3.7. Воспользуйтесь леммой 4.3.7.

Упражнение 4.3.8. Рассмотрите точку, где пределы $\overline{\lim}$ и $\underline{\lim}$ не совпадают, выберите между ними рациональное число и возьмите соответствующую периодическую точку.

Упражнение 4.3.9. Рассмотрите интервалы между соседними неподвижными точками и получите противоречие, предположив, что все точки устойчивы или полуустойчивы.

Упражнение 4.4.5. Да. Пусть $A_{p/q}$ является объединением непересекающихся открытых множеств O_1 и O_2 . Покажите, что можно считать замыкания $\bar{O}_1$ и $\bar{O}_2$ непересекающимися. Используя компактность, получите противоречие.

Упражнение 5.1.2. Либо хотя бы одно из чисел является точным квадратом, либо их отношение равно m^2/n^2 для некоторых $m, n \in \mathbb{N}$. Для доказательства рассмотрите отдельно случаи с разным количеством ненулевых коэффициентов в линейной зависимости.

Упражнение 5.1.5. Используйте китайскую теорему об остатках.

Упражнение 5.1.7. Если все элементы подгруппы Γ линейно зависимы над $\mathbb{R}$ (т.е. Γ лежит на некоторой прямой), то утверждение получается из упражнения 4.2.8 с помощью замены координат. В противном случае Γ содержит два линейно независимых вектора. Фактор группы Γ по решетке, порожденной этими векторами, является замкнутой подгруппой в $\mathbb{T}^2$. Предыдущее упражнение позволяет описать такие подгруппы.

Упражнение 5.1.9. $\mathbb{R}^k \times \mathbb{Z}^l$, где $0 \leq k + l \leq n$ и $k < n$.

Упражнение 5.1.11. Устройство группы $\mathbb{Z}_2$ похоже на устройство канторова множества. На n -м уровне имеется 2^n множеств (два элемента лежат в одном множестве, если их разность является степенью двойки). Равномерность распределения означает, что предельные частоты попаданий во все эти множества равны.

Упражнение 5.2.1. Восемь, если исходное направление не параллельно граням.

Упражнение 5.2.2. Рассмотрим группу S , порожденную отражениями от трех плоскостей, параллельных граням куба I и проходящих через центр куба.

Орбита единичного куба под действием этой группы — это куб S вдвое большего размера с центром в нуле. Используйте группу S для построения частичной развертки. Каждая билиардная орбита внутри I после этого разворачивается в прямолинейное движение по тору, который получается из S отождествлением противоположных сторон.

Упражнение 6.1.3. Да.

Упражнение 6.1.9. Используя теорему Бэра о категории, найдите рекуррентные точки внутри замкнутого шара, содержащегося в области.

Упражнение 6.2.6. Можно или произвести вычисление в координатах, аналогичное (6.2.4), либо использовать инвариантность потенциала относительно сдвигов (для центра тяжести) и отражений (для углового момента).

Упражнение 6.2.7. В системе отчета, связанной с центром тяжести, получается два независимых движения в поле центральных сил, так что орбиты являются эллипсами.

Упражнение 6.2.9. См. В. И. Арнольд: *Математические методы в классической механике*, М: УРСС, 2000.

Упражнение 6.3.3. Рассмотрите отдельно случай, когда происходят соударения с внутренней окружностью, и когда они не происходят. В обоих случаях фазовое пространство расслаивается на инвариантные окружности.

Упражнение 6.3.4. Рассмотрите развертку с помощью отражений относительно осей координат.

Упражнение 6.3.6. Постройте линейную комбинацию квадратов трех интегралов свободного движения (двух компонент скорости и углового момента относительно начала координат), инвариантную относительно отражений от стенок.

Упражнение 6.4.3. Они состоят из равных эллиптических дуг, количество которых равно количеству сторон.

Упражнение 6.4.4. $H(S_1, S_2) + \dots + H(S_n, S_1)$.

Упражнение 6.4.5. Деформируйте небольшую дугу граничной окружности.

Упражнение 6.4.6. Такие кривые называются кривыми постоянной ширины. Их можно получить, вращая отрезок вокруг точки, которая перемещается по этому отрезку.

Упражнение 6.4.9. Модифицируйте конструкцию шнура.

Упражнение 7.1.2. Точка должна иметь вид p/q , где q и m взаимно просты.

Упражнение 7.1.4. Рассмотрите поднятие, примените линейную интерполяцию и спроектируйте.

Упражнение 7.1.5. Если 0 — притягивающая неподвижная точка, то точка

$$x_0 := \sup\{x \in [0, 1] : f(y) \leq y \text{ при } y \in [0, x]\}$$

также является неподвижной.

Упражнение 7.1.9. См. указание к упражнению 7.1.6.

Упражнение 7.2.1. $1/4$.

Упражнение 7.2.7. Рассмотрите два случая: 1) оба собственных значения по модулю больше 1 и 2) одно из собственных значений по модулю меньше 1. В случае 1) отображение является растягивающим относительно некоторой нормы, и можно рассуждать как в предложении 7.2.7; в случае 2) рассуждения аналогичны предложению 7.2.9.

Упражнение 7.3.4.

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

Упражнение 7.3.8. 1.

Упражнение 7.3.11. Рассмотрите прообразы второго порядка точки 0.

Упражнение 7.3.13. $\frac{1+\sqrt{5}}{2}$.

Упражнение 7.3.14. Каждый фактор задается отображением E_n .

Упражнение 7.3.15. Наименьшее значение достигается при

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Упражнение 7.4.8. Диагонализуйте матрицу, продолжите прямые, порожденные собственными векторами, пока не получится достаточное количество пересечений, и рассмотрите связанные компоненты. Полезно для начала рассмотреть случай матрицы $\begin{pmatrix} 13 & 8 \\ 8 & 5 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^3$.

Упражнение 7.5.2. В правой части ключевой формулы имеются выражения, содержащие факториалы. Общее количество слагаемых растет как n^3 . Нужно найти наибольшие «плохие» слагаемые и оценить факториалы с помощью формулы Стирлинга. Как и в (7.5.3), получается, что количество слагаемых растет полиномиально, а каждое из них экспоненциально мало, поэтому сумма также экспоненциально убывает.

Упражнение 7.6.1. Для одностороннего сдвига полусопряжение с E_2 обратимо почти всюду, так что доказательство остается тем же. Для двустороннего сдвига достаточно рассматривать цилиндры с положительными индексами, что эквивалентно рассмотрению одностороннего сдвига.

Упражнение 7.6.1. Если ϵ достаточно мало, то сумма (7.6.1) для p и аналогичная сумма для q соответствуют непересекающимся множествам.

Упражнение 8.1.2. 1.

Упражнение 8.1.5. Для построения минимального покрытия воспользуйтесь тем, что цилиндры являются шарами. Чтобы показать, что покрытие цилиндрами является оптимальным, используйте то, что d''_λ — ультраметрика (т. е. в любом треугольнике хотя бы две стороны равны).

Упражнение 8.1.8. Действуйте так же, как и при построении канторова множества. Чтобы получить энтропийную размерность 0, нужно на n -м шаге удалять $1 - 1/2^n$ -ю часть из каждого из оставшихся отрезков, а чтобы получить энтропийную размерность 1 — удалять $1/2^n$ -часть.

Упражнение 8.1.8. $\frac{\ln \frac{1 + \sqrt{5}}{2}}{\ln 2}$.

Упражнение 8.2.1. 0.

Упражнение 8.2.2. $\ln 2$.

Упражнение 8.2.3. $h_{\text{top}}(f) = 0$. Используйте полусопряжение с иррациональным поворотом и покажите, что блуждающие интервалы добавляют лишь ограниченное число (n, ϵ) -отделенных точек для любого $\epsilon > 0$.

Упражнение 8.2.4. $h_{\text{top}}(f) = 0$. Действительно, (n, ϵ) -отделенные множества растут квадратично с ростом n .

Упражнение 8.3.3. Можно рассмотреть, например, *гильбертов кирпич* H , то есть произведение счетного числа единичных отрезков (занумерованных натуральными числами). Это компактное пространство, которое можно наделить метрикой. Тогда сдвиг на пространстве H топологически транзитивен и имеет бесконечную топологическую энтропию.

Вот более простой пример, в котором отображение не топологически транзитивно. Рассмотрим несвязное объединение пространств Ω_N , $n = 2, 3, \dots$, причем метрика на N -ом множестве получается из метрики d_2 умножением на 2^{-N} . Чтобы пространство стало компактным, необходимо добавить точку p , так что для $x \in \Omega_N$ расстояние между p и x равно $10/2^n$.

Задача 8.3.8. Для целых t утверждение непосредственно следует из предложений 8.3.6 и 8.2.9, 3). Последнее из них позволяет сразу получить утверждение и для рациональных t . Для иррациональных t используйте рациональные приближения и рассуждения из доказательства предложения 8.3.6, чтобы получить неравенства в обе стороны.

РЕШЕНИЯ

Упражнение 1.2.3. См. пример 2.2.9.

Упражнение 1.2.4. Обозначим рассматриваемую сумму через x_n . Записав ее второй раз со «сдвигом» на единицу и сложив почленно, получим $2x_n = x_n - 1 + b_{n+1} + b_n$, откуда $x_n = b_{n+2} - 1$.

Упражнение 1.2.6. См. предложение 2.5.1.

Упражнение 1.2.19. $1/3$. Докажите, что число $a_n + 2a_{n+1}$ не зависит от n .

Упражнение 1.3.1. $(1, 4) \mapsto (5/2, 8/5) \mapsto (41/20, 80/41) \mapsto ((41^2 + 40^2)/41 \cdot 40, 41 \cdot 160/(41^2 + 40^2))$.

Упражнение 1.3.6. См. п. 2.2.8.

Упражнение 1.3.9. $(n + 10)^2 = n^2 + 20n + 100 = n^2 \pmod{10}$.

Упражнение 1.3.10. $(10 - n)^2 = 100 - 10n + n^2 = n^2 \pmod{10}$.

Упражнение 1.3.11. Число $a_{10q-n} - a_n = (10q - n)^2 p/q - n^2 p/q = 10(10pq - 2np)$ кратно 10.

Упражнение 1.3.19. См. предложение 2.2.27.

Упражнение 1.3.20. См. предложение 2.2.27.

Упражнение 1.3.24. Для $n \geq 3$ число $2^{n+50} + 2^n = 2^n(2^{50} + 1) = 2^n(\dots 625)$ делится на 125 и на 8, поэтому оно делится на 1000.

Упражнение 2.1.1. $y_{i+1} = x_{i+1} - \frac{b}{1-k} = kx_i + b - \frac{b}{1-k} = k\left(y_i + \frac{b}{1-k}\right) + b - \frac{b}{1-k} = ky_i$.

Упражнение 2.2.1. Если используются радианы, наша процедура состоит в многократном применении к исходному значению функции $\sin x$; поскольку $|\sin x| = \sin |x| < |x|$, последовательность $a_n := \sin^n |x|$ убывает. Поскольку эта последовательность ограничена снизу нулем, она сходится (п. А.1.2), причем предел является неподвижной точкой (см. (2.3.1)); единственная неподвижная точка — это точка 0. Поскольку производная функции $\sin x$ в нуле равна 1, это отображение не является сжимающим и, более того, отношение соседних членов стремится к 1. Поэтому сходимость не экспоненциальная.

При использовании градусов мы применяем функцию $\sin(\pi x/180)$, которая в силу предложения 2.2.3 задает сжимающее отображение, причем отношение соседних членов (быстро) приближается снизу к $a := \pi/180$. Таким образом, чтобы получить значение, в 10^{-10} меньшее исходного, нужно примерно $-10/\lg a < 6$ шагов.

Упражнение 2.2.2. Функция $\sqrt{x}$ является сжатием на $[a, \infty)$ при $a > 1/4$, причем точка 1 является единственной неподвижной точкой, и, следовательно,

пределом любой рассматриваемой последовательности. Значение производной $1/2\sqrt{x}$ в точке $x = 1$ равно $1/2$ поэтому расстояние до точки 1 на каждом шаге уменьшается примерно вдвое; следовательно, после примерно k шагов расстояние до 1 становится меньше предела точности калькулятора, если начальное значение невелико (для значения порядка 2^l потребуется $k + l$ шагов).

Упражнение 2.2.3. Для полуинтервала $(0, 1]$ нельзя применить принцип сжимающих отображений, но можно свести задачу к предыдущей, рассмотрев последовательность обратных величин: если начальное значение равно x , то $\sqrt[n]{x} = 1/\sqrt[n]{1/x}$ находится примерно на том же расстоянии от точки 1, что и $\sqrt[n]{1/x}$. Поэтому сходимость также является экспоненциальной, и при начальном значении порядка 2^{-l} и точности k двоичных разрядов требуется $k + l$ итераций, чтобы достигнуть предела точности.

Упражнение 2.2.4. $|x^2 - y^2| = |x - y||x + y| \leq \lambda|x - y|$ для $x, y \in [-\lambda/2, \lambda/2]$.

Упражнение 2.2.5. Каждое лето популяция утраивается, после чего все лемминги, которые существовали к началу предыдущего лета, умирают. Таким образом, $b_{n+1} = 3b_n - b_{n-1}$. После деления на b_n получаем

$$a_n := \frac{a_{n+1}}{a_n} = \frac{3a_n - a_{n-1}}{a_n} = 3 - \frac{1}{b_{n-1}} =: g(b_{n-1}).$$

При этом функция g возрастает и $g(2) = 5/2 > 2$, $g(4) = 11/4 < 4$, поэтому $g([2, 4]) \subset [2, 4]$. На этом интервале $g'(x) = 1/x^2 \leq 1/4 < 1$, поэтому g является сжатием и имеет единственную неподвижную точку $\omega \in [2, 4]$. Для нее получаем уравнение $\omega = 3 + 1/\omega$, из которого находим $\omega = (3 + \sqrt{5})/2$.

Упражнение 2.3.2. Вторая итерация f^2 является неубывающим отображением, поэтому возможны только периоды 1 и 2. С другой стороны, отображение $f(x) = -x$ имеет периодические точки периода 2.

Упражнение 2.5.3. Функция $f(x) - x$ невозрастает и имеет нуль (по теореме о промежуточном значении), поэтому множество нулей — точка или отрезок.

Упражнение 2.6.2. Если $x \in \bigcup_{\alpha \in A} O_\alpha$, то $x \in O_\alpha$ при некотором $\alpha \in A$, поэтому существует число $r > 0$, для которого $B(x, r) \subset O_\alpha \subset \bigcup_{\alpha \in A} O_\alpha$. Осталось воспользоваться тем, что замкнутые множества — это дополнения к открытым.

Упражнение 2.6.3. Эти шары имеют такой вид: $\{n \in \mathbb{Z} : d(n, 0) < 1\} = \{0\}$, $\{n \in \mathbb{Z} : d(n, 0) \leq 1\} = \{-1, 0, 1\}$. Оба шара открыты и замкнуты. Действительно шар $\{n \in \mathbb{Z} : d(n, 0) < 1\}$ (состоящий из одной точки) открыт согласно упражнению 2.6.1; следовательно, любая точка из $\mathbb{Z}$ является открытым подмножеством, а тогда вообще любое подмножество открыто (упражнение 2.6.2). Отсюда следует, что любое подмножество замкнуто (так как его дополнение открыто).

Упражнение 2.6.4. Для любого числа m множество $\{n \in \mathbb{Z} : d(n, m) < 1\} = \{m\}$ открыто (в силу упражнения 2.6.1), поэтому любое подмножество в $\mathbb{Z}$ открыто (упражнение 2.6.2).

Упражнение 2.6.5. То, что внутренность любого множества открыта, следует из определения A.1.4 и упражнения 2.6.2. Покажем теперь, что $\bar{A} \subset \bar{A}$.

Пусть $x \in \bar{A}$ и $r > 0$. Тогда, согласно определению A.1.4 существуют точки $y \in B(x, r/2) \cap \bar{A}$ и $z \in B(y, r/2) \cap A$. Поэтому $z \in A \cap B(x, r)$ и $x \in \bar{A}$ в силу определения A.1.4.

Упражнение 2.6.6. Из определения A.1.4 получаем, что $\partial S = \bar{S} \setminus \text{Int } S$ и поэтому граница ∂S замкнута. Если S открыто, $x \in \partial S$ и $r > 0$, то $B(x, r) \cap S \neq \emptyset$ и $x \notin \text{Int } \partial S$, поскольку $S \cap \partial S = \emptyset$ в силу определения A.1.4. Теперь применим оба утверждения: граница замкнута, а ее граница совпадает с границей ее дополнения, которое открыто.

Упражнение 2.6.7. Пространство $\mathbb{R}$ полно (это хорошо известное свойство), $\mathbb{Q}$ не полно, поскольку последовательность из упражнения 1.1.8 является фундаментальной и ее члены рациональны, но она не сходится (в $\mathbb{Q}$). Пространства $\mathbb{Z}$ и $[0, 1]$ полны в силу упражнения 2.6.8. (Полноту $\mathbb{Z}$ можно доказать и проще, заметив, что любая фундаментальная последовательность постоянна начиная с некоторого места.)

Упражнение 2.7.1. Воспользуйтесь упражнением 2.6.8.

Упражнение 3.1.2. $0 = \langle Av, w \rangle - \langle v, Aw \rangle = (\lambda - \mu)\langle v, w \rangle$.

Упражнение 3.2.1. а) Неустойчивый узел, б) седло, в) неустойчивый фокус, г) неустойчивый вырожденный узел.

Упражнение 3.3.1. Характеристические многочлены имеют следующий вид:

а) $-\lambda((3-\lambda)(-3-\lambda)+8) + (2(-3-\lambda)+8) + 2(4-2(3-\lambda)) = -\lambda(\lambda^2-1) + 2\lambda-2 = (\lambda-1)[-\lambda(\lambda+1)+2] = (1-\lambda)[\lambda^2+\lambda-2] = (1-\lambda)(\lambda-1)(\lambda+2);$

б) $(-1-\lambda)[(-1-\lambda)^2-1] + 2(-1-\lambda) = -(1+\lambda)[(1+\lambda)^2+1];$

в) $(2-\lambda)[(1+\lambda)^2(2-\lambda) + (2-\lambda)] = (2-\lambda)^2[(1-\lambda)^2+1].$

Упражнение 4.1.8. Пересечение двух таких множеств — замкнутое инвариантное подмножество, содержащееся в каждом из них.

Упражнение 4.1.10. Например, $X = \{z \in \mathbb{C} : |z| \in \{1, 2\}\}$ и $f(z) = e^{2\pi i \alpha}$, где $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Упражнение 4.1.15. Чтобы доказать транзитивность, достаточно показать, что точка 1 лежит в замыкании орбиты. Любой элемент $g \in \mathbb{Z}_2 \setminus \mathbb{Z}_2^+$ — это предел некоторой последовательности нечетных чисел. Если m нечетно и $n \in \mathbb{N}$, то найдется число $k \in \mathbb{N}$, для которого $mk \equiv 1 \pmod{2^n}$.

Упражнение 4.2.8. Если точная нижняя грань множества положительных элементов подгруппы Γ — положительное число a , то любой элемент подгруппы есть целое кратное a . В противном случае в Γ имеются положительные и отрицательные числа, сколь угодно близкие к нулю, и потому эта подгруппа плотна.

Упражнение 4.3.3. $F(x+1) - F(x) = 1/2(\sin(x+1) - \sin x)$. Это непостоянная функция, поэтому F не является поднятием никакого отображения окружности.

Упражнение 4.4.4. В силу монотонности всякое пересечение является отрезком. Чтобы показать, что пересечение непусто, заметим, что $\rho_{0,b} = 0$, $\rho_{1,b} = 1$ и воспользуемся непрерывностью. Наконец, то, что длина положительна, следует из предложения 4.4.10 (которое применимо, поскольку $f_{a,b}$ — целая функция).

Упражнение 4.4.6. Используйте предложения 4.4.9 и 4.4.10 (как при доказательстве предложения 4.4.13).

Упражнение 5.1.1. В противном случае $\sqrt{3}$ является линейной комбинацией (с рациональными коэффициентами) чисел 1 и $\sqrt{5}$. Возводя обе части в квадрат, получаем, что число 3 иррационально.

Упражнение 6.1.1. Поскольку длина и ориентация сохраняются, $f(x) - f(0) = \text{length}([f(0)/f(x)]) = \text{length}([0, x]) = x - 0$, откуда $f(x) = x + f(0)$.

Упражнение 6.1.8. Например, $\mathbb{Q}$ и $\{O_q := \mathbb{Q} \setminus \{q\} : q \in \mathbb{Q}\}$.

Упражнение 6.1.10. Множество точек, орбиты которых пересекают данное открытое множество, открыто (по определению) и плотно (поскольку содержит любую плотную орбиту). Рассмотрим плотную орбиту и всевозможные шары с рациональными радиусами, центры которых находятся в точках этой орбиты. Множество точек, орбиты которых пересекают все эти шары (и тем самым плотны), является объединением счетного количества открытых всюду плотных множеств. Переходя к дополнениям, получаем требуемое утверждение.

Упражнение 6.2.5. Уравнения Лагранжа сохраняются при отражениях относительно больших кругов. Поэтому их решения (т. е. геодезические) инвариантны в том смысле, что образ геодезической при отражении также является геодезической. Если начальное направление является касательным к большому кругу, то (в силу теоремы единственности) соответствующая геодезическая совпадает с большим кругом. Поскольку любой вектор является касательным к некоторому большому кругу, никаких других геодезических нет. Поэтому геодезический поток является периодическим.

Упражнение 6.4.1. Рассмотрим точки пересечения границ стола с осями симметрии и соединим их последовательно (в направлении против часовой стрелки). Полученная фигура инвариантна относительно обоих отражений, поэтому в любой точке пересечения угол падения равен углу отражения.

Упражнение 7.1.4. Периодическая точка x однозначно определяется (кодируется) последовательностью своих посещений отрезков Δ_0 и Δ_1 . Наоборот, любая последовательность нулей и единиц длины n задает (единственную) периодическую n -орбиту. В частности, если последовательность не имеет более короткого периода, то период соответствующая орбита в точности равен n .

Упражнение 7.1.10. Воспользуемся описанием рекуррентного соотношения, данным в п. 3.1.9, и спроектируем на $\mathbb{R}^2/(\mathbb{10}\mathbb{Z})^2$, т. е. рассмотрим остатки по модулю 10 (другой вариант: рассмотрим орбиту точки $(1/10, 1/10)$ на стандартном торе). Каков период орбиты, начинающейся в точке $(1, 1)$? Заметим, что здесь мы формально описали тот факт, что последняя цифра любого числа Фибоначчи однозначно определяется последними цифрами двух предыдущих.

Упражнение 7.2.5. Пусть $p \in U$, $q \in V$ — периодические точки (мы используем предложение 7.1.10), n — их общий период. Прямая из первого семейства, проходящая через точку p , плотна, как и прямая из второго семейства, проходящая через точку q . Следовательно, множество точек пересечения плотно и эти точки являются гетероклиническими.

Упражнение 7.3.2. Докажите, что множество двоичных разложений $0, a_1 a_2 \dots$, в которых $a_i = 1$, когда i не кратно 20, несчетно.

Упражнение 7.3.13. Нарисуем марковский граф для матрицы

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

Рассмотрим теперь другой граф, точки которого соответствуют ребрам первого графа, а ребра соединяют пары последовательных ребер первого графа. Сравните его с графом Маркова для матрицы

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

Упражнение 7.4.1. Если точка не лежит на отрезке $[0, 1]$, то ее вторая итерация будет отрицательным числом, а орбита любого отрицательного числа уходит в $-\infty$.

Упражнение 7.4.3. Убедитесь, что $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^2$, и проверьте, что разбиение, показанное на рис. 7.4.4, является марковским разбиением для матрицы

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

Упражнение 7.4.7. Предположим, что полусопряжение не является взаимно однозначным. Тогда некоторый интервал при полусопряжении переходит в точку. Значит, любая итерация этого интервала при полусопряжении переходит в точку. Из теоремы о конечных приращениях следует, что поднятие отображения f увеличивает длину любого интервала. В силу компактности имеется наиболее длинный интервал, который переходит в точку — противоречие.

Упражнение 8.1.1. Отрезок $[0, 3]$ можно покрыть двумя шарами радиуса 1 (например, с центрами в точках $2/3$ и $8/3$); с другой стороны, покрытие шарами радиуса 1 с центрами в точках $1/2$, $3/2$ и $5/2$ является минимальным.

Упражнение 8.1.3. $\ln \mu < \ln 4 + \ln \mu < \ln 4 - \ln \lambda < \ln 4 - \ln 2 = \ln 2 < -\ln \mu$.

Упражнение 8.3.2. Рассмотрим прямое произведение произвольного отображения f с положительной топологической энтропией и иррационального поворота R_α . В силу предложения 8.2.9, 5) имеем $h_{\text{top}}(f \times R_\alpha) = h_{\text{top}}(f) + h_{\text{top}}(R_\alpha) = h_{\text{top}}(f) > 0$.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- 0-1 матрица 232
2-адические целые числа 121
 ω -предельное множество 146, 174
 $B_f(x, r, n)$ 265
 C^1 -топология 438
 C' -топология 438
 $NW(f)$ 319
 $P_n(f)$ 210
- Абсолютно непрерывная мера 321
Автоморфизм тора 215–218, 224, 243, 258, 272, 302, 306, 316
Автономное дифференциальное уравнение 292
Адаптированная метрика 80
Алфавит 231, 232
Асимптотическое распределение 121
Астроида 209
Аттрактор 322, 362
— Лоренца 303, 365–372
— Смейла 363–365
- Бабочка 20, 64, 225
Базисное множество 303
Безразмерные координаты 180
Биллиардное отображение 193
Биллиардный поток 129, 209
Бинарный поиск 29
Биркгофова периодическая орбита 374, 375, 381–384, 392, 393
Бифуркация 60, 332
— седлоузла 60
— удвоения периода 332
Блуждающая точка 150
- Векторное поле 292
Взаимно однозначное отображение 75
Вложение 440
Внешний билиард 380
Внутренность 426
Возрастающая функция 57
Вполне интегрируемая система 135
- Вполне несвязное пространство 427, 437
— ограниченное пространство 431
Временное среднее 118
Второй закон Кеплера 187
Вынужденные колебания 380
Выпуклая каустика 207–209
Выпуклое множество 49, 56
Выпуклость 49, 56, 200, 204
Выврожденный узел 93, 100
- Гармонический осциллятор 127, 154, 178, 183, 188
Геодезическая 32, 179, 190
Геодезический поток 167, 179, 190, 356, 414, 419
Геометрическая оптика 206
Гетероклиническая картина 355
— орбита 236, 393
— точка 59, 227, 351
Гильбертов кирпич 435
Гиперболическая динамическая система 314
— метрика 414
— неподвижная точка 295
Гиперболический аттрактор 365
— репеллер 303
Гиперболическое квадратичное отображение 329, 343
— линейное отображение 89, 223
— множество 302–303, 314–319
— локально максимальное 303
— отображение 106
— отталкивающее множество 303
Гиперболичность 417
Гипотеза Оппенгейма 400
Гладкая зависимость от начальных условий 291
Гомеоморфизм 75, 430
Гомоклиническая картина 353, 355, 394
— петля 174, 181, 184

- Гомоклиническая точка 59, 349, 351, 354, 393
- Гомоклинические касания 359
- пересечения 356
- Гомотетия 89
- Граница 426
- Графические вычисления 57
- Группа 119, 157
- Губка Серпинского 85
- Движение свободной частицы 167, 179
- двадцатые целые числа 438
- Детерминант 171
- Диагонализация 89
- Диаметр 196, 204
- Дивергенция 172
- Диофантово число 410
- Диффеоморфизм 440
- Дифференциал 48
- Допустимый путь (допустимая последовательность) 234
- Емкость 261, 265
- Жорданова нормальная форма 104
- Задача Кеплера 135
- Штермера 355
- Закон больших чисел 252, 345
- Ньютона 177
- Закручивающее отображение 193, 355, 377
- Замена времени 294
- Замкнутая геодезическая 32, 356, 398
- Замкнутое множество 426
- Замощение 130
- Замыкание 426
- Захват 356
- Игра «жизнь» 36
- Измеримое множество 443
- разбиение 444
- Изометрия 75, 169, 223, 429
- Импульс 179
- Инвариант 264
- Инвариантная мера 320–321
- окружность 199, 207, 390–392
- плотность 320
- Индуктированная метрика 426
- Интеграл Римана 445
- Рисса 442
- Интегрируемая функция 440
- Интегрируемое закручивание 167
- Иррациональный поворот 111
- Итерационный процесс Пикара 289
- Канторова лестница 84, 149, 153
- Канторово множество 82–84, 146, 149, 151, 230–231, 233, 243, 262, 340, 373, 427, 436–438
- Картины 351
- Категория Люстерника—Шнирельмана 398
- Каустика 195, 197, 198, 205, 207
- Квадратичная сходимости 72
- форма 418, 420–422
- Квадратичное семейство 22, 69–73, 212–214, 240, 325–346
- Квадратный корень 28
- Квазипериодичность 108
- Кинетическая энергия 179
- Классификация Пуанкаре 147
- Клеточный автомат 36
- Климат 345
- Ковер Серпинского 84, 262
- Кодирование 227, 228, 231, 232, 239, 243, 272, 313, 315–318
- Компактное пространство 430, 437
- Консервативная система 179
- Конструкция шнура 209
- Конус 304
- Конфигурационное пространство 177
- Корневое подпространство 104, 105
- Кривая, заполняющая все пространство 86
- Менгера 85
- Пеано 86
- Кривизна 206
- Критерий Малера 418
- Кролики 11, 12, 46
- и хищники 18
- Круговое кольцо 49
- Кусочно монотонная функция 122
- Лемма Аносова о замыкании 308
- об ϵ -траекториях 310
- Лемминги 56
- Лестница Лемерея 57
- Линеаризация 126, 183
- Линейная аппроксимация 42

- Линейное закручивание 170
Линейный поток 159
— — на торе 125, 135
Липшицева функция 62, 151, 293, 386, 391
Липшицево отображение 44, 430, 436
Логистическое дифференциальное уравнение 64–67
— уравнение 22, 69–73, 99
Локализация 298
Локальная трубка тока 185
— энтропия 274
Локально компактное пространство 433
Малые колебания 183
Марковский граф 234, 235, 337
Марковское разбиение 316, 317, 326
Маршрут 227
Математический маятник 135, 174, 180–184
Мера 443
— Бернулли 256, 258
— Гиббса 321
— Лиувилля 415
— Синяя—Рюэлля—Боуэна 323
— Хаара 415
Меркурий 188
Метод исключения параметра 346
— Кронекера—Вейля 118–119, 164, 401
— Ньютона 30, 53, 281
— Пуанкаре—Мельникова 358
Метрика 74, 109, 231, 426
— адаптированная 80
— Ляпунова 80
— на произведении пространств 434–435
— Хаусдорфа 432, 433
Метрическое пространство 426
Минимаксная орбита 377, 384, 393
Минимальное множество 111, 120
Минимальность 111, 126, 158, 160, 162, 167
Минимальный гомеоморфизм 111
— поток 126
Множества Обри—Мазера 385–389
Множество вращения 148
— меры нуль 250, 321, 329, 440
— ограниченное 426
Модифицированная норма 107
Модифицированное скалярное произведение 107
Модулярная поверхность 416, 419
Наименьший период 45
Неблуждающая точка 319
Невозрастающая функция 57
Неподвижная точка 45
— — притягивающая 53, 361
Непрерывное отображение 75, 429
Неравенство треугольника 74, 426, 434
Несжимаемость 169, 172
Неубывающая функция 57
Неустойчивое отображение 224, 226, 233, 235, 274, 307
Нигде не плотное множество 427
Норма 91, 104, 433
— Ляпунова 106
— матрицы 48, 56, 103
Нормальная форма 371
Область 173
— неустойчивости 392
— притяжения 327, 331
Обратный предел 218–219, 365
Общая теория относительности 188
Огибающая 205–207
Однородное действие 414
— пространство 414
Одометр 121, 340
Окрестность 426
Окружность 75, 108, 136
Омары 14
Орбита 45
Орицикл 417
Орициклический поток 417
Открытое множество 426
— отображение 430
Отображение возвращения 68, 126
— Гаусса 410, 419
— Кремоны 350
— окружности 136
— сечения 193
—, сохраняющее ориентацию 139
Отрезок закручивания 378, 395
Отталкивающая неподвижная точка 60
Параболическое линейное отображение 90

- Параметр Лазуткина 209
Первый интеграл 186
Перевал 375
Перемешивание 256–259
Перемешивающее отображение 257
Перенос обратный 428
Период 45
Периодическая орбита 210–212
— точка 45, 327
— — притягивающая 327
Периодические коэффициенты 102
— орбиты 233, 343
— точки 210–220, 234, 243, 252, 273, 308–310, 312, 338–339, 342, 394, 396
Периодическое движение 129
Плотное множество 426
Поворот окружности 108–121, 175
Подбрасывание монеты 253
Подкова 230, 241, 263, 303, 314, 347, 351, 353–357, 359, 365, 392, 393
Поднятие 137, 377
Подходящая дробь 406
Поле конусов 305
Полная энергия 179, 181
Полнота 75, 428
Положение равновесия 61, 62
Положительная матрица 235
Полусопряжение 233, 245
Полустойчивая орбита 144, 153
— точка 61
Последние цифры значений
многочленов 34–36
Последняя геометрическая теорема
Пуанкаре 395, 399
Последовательность Коши 46, 75, 428
Потенциальная энергия 179, 188
Поток 66–67, 294
Почти всюду 250
— сжимающее отображение 80, 91, 104–106
Правила Хопфмейстера 16
Предельная точка 427
Предельный цикл 19, 67, 362
Преобразование Мёбиуса 415
Прессессия 188
Приближение 310, 321
Пример Данжуа 150
Принцип сжимающих отображений 45, 47, 79, 279–283, 286, 289, 307–308
Принцип Ферма 196
Притягивающая неподвижная точка 53, 361
Притягивающая точка (орбита) 327
Производная 438
— Шварца 333
Производящая функция 196, 201, 378–380
Пространственное среднее 118
Пространство двусторонних последовательностей 263
— непрерывных функций 429, 434
— последовательностей 231, 436
— состояний 177
Прямоугольник 316
Псевдоорбита 308
Псевдосфера 416
Равномерная рекуррентность 174–175
Равномерно рекуррентная точка 175
Равномерное распределение 36, 115, 119, 124, 129, 133, 159–162, 247–253, 320, 344
— — для квадратов 403
— — для многочленов 404
Радиоактивный распад 17
Развертка 130
Разделение 307, 310, 312
Растягивающее отображение 210–212, 214–218, 223, 227–228, 237–240, 247, 270–271, 315, 319
Растягивающееся подпространство 106
Рациональная независимость 158, 165
Резонанс 93
Рекуррентная точка 173
— формула 97
Рекуррентное поведение 169
— соотношение 98
Рекуррентность 108, 150, 169
Ренормализация 332, 413, 414
Решетка 417, 421
Самоподобие 83, 150
Сахарная вата 364
Сведение к первому порядку 97, 178, 181
Сверхпритягивающая неподвижная точка 54
Светлячки 26, 154

- Свойство перенормировки 83
— спецификации 311, 312
Связное пространство 427
Сдвиг 37, 232, 233, 258
— конечного типа 232
Седло 95, 100, 101
Седловая точка 196
Семейство конусов 348
Сепарабельное пространство 433
Сжатие 44
Сжимающее отображение 42–56, 75, 430
Сжимающееся подпространство 106
Символическая динамика 37
Синдетическое подмножество 175
Синхронизация 27, 154
— частот 153–155
Скользющие блочные коды 37
Скорость роста 211
Случайность 253
Снежинка Коха 84, 262
Собственная функция 164
Собственное значение 88, 165
— подпространство 88
Собственный вектор 88
Совершенное множество 146, 427, 437
Соленоид 218, 322, 363
Сопряжение 148, 233, 239, 264
Сопряженные отображения 233
Софическая система 312
Софокусная гипербола 198
Софокусный эллипс 197
Спектр 103
Спектральное разложение 319
Спектральный радиус 103, 234
Спецификация 311
Среднее Биркгофа 117
Статистические свойства 319
Стационарное состояние 42
Степени двойки 32, 123
Степень отображения 137, 215
Стохастичность 344
Строгая дифференциальная выпуклость 201
— эргодичность 118, 247, 252, 320, 341, 400, 402, 417, 421
Строго эргодичное отображение 118
Структурная устойчивость 299, 313–315, 326
Субаддитивная последовательность 139
Субаддитивность 161
Суммы Римана 445
Сфера 179
Сферический маятник 126, 188
Схема Бернулли 254
Сходимость 427
Тентообразное отображение 344
Теорема Биркгофа—Смейла 351
— Бореля о плотности 421
— Бэра 429
— Гейне—Бореля 431, 437
— Гробмана—Хартмана 299
— Данжуа 387
— Жордана 73
— о конечных приращениях 439
— о гиперболической неподвижной точке 308
— о неявной функции 285
— о промежуточном значении 57–59
— о семействах ϵ -траекторий 313
— о спецификации 311, 321
— об обратной функции 283
— об устойчивом многообразии 297, 305, 349
— Пуанкаре о возвращении 173
— Пуанкаре—Бендиксона 73
— Шарковского 339
— Шеннона—Макмиллана 256
Теория инвариантов перемешивания 326, 334
Топологическая группа 119
— транзитивность 228, 233, 309
— цепь Маркова 232, 245, 271, 312, 316, 318, 329
— энтропия 264–270, 273, 312
Топологически перемешивающее множество 311
— — отображение 222, 243, 257
— транзитивное отображение 111, 120, 221
Топологический инвариант 233
Топологическое перемешивание 226, 233, 318–319
— сопряжение 142
Тор 78, 125, 131, 135, 157, 158, 166, 179, 223, 224, 243
Транзитивная матрица 235

- Трансверсальная периодическая точка 287
Тригонометрический многочлен 119, 164
Убывающая функция 57
Угловой импульс 189
— момент 186
Угол перигелия 187
Удвоение периода 325, 330–332
Узел 93, 100
Универсальное отталкивающее множество 328
Унимодальное отображение 214
Унимодулярная замена координат 422
— решетка 417
Унипотентность 421
Упорядочение 152
Упорядоченное состояние 382
Упорядоченность 143
Уравнение зеркала 206
— Лагранжа 189
— Лотки—Вольтерра 18
— Эйлера—Лагранжа 189
Условие Липшица 44, 62, 75, 289
Усреднение Биркгофа 116, 247, 250
Фазовое пространство 177, 194
Фазовый объем 169, 172, 179
— портрет 64
Фактор 148, 233, 237
Факторизующее отображение 233
Фигуры Лиссажу 127
Филлотаксис 15–16
Фокус 94, 100
Фокусирование 195
Формула Стирлинга 249
Фрактал 262
Фуксова группа 415
Фундаментальная область 131, 134, 157
Функционал действия 189, 374
Функция Ляпунова 366
— ограниченной вариации 151
— распределения 122, 256, 442
Хаос 220, 338, 373
Хаотическое отображение 220, 223–225, 233, 235
Характер 119
Характеристическая функция 116
Центр 100
Центральная сила 186–188
Цепная дробь 112, 414
Цепная дробь 406
Цепочно-рекуррентное множество 309
Цилиндр 78, 182, 193, 194, 231, 234, 235, 271
Частичное топологическое перемешивание 227
Частота 113, 115, 124, 159
Число вращения 139, 142–143, 146, 151–153, 264, 385–390, 396
Шар 75
Ширина 196, 204
Эквивалентность потоков 274
Эквивалентные метрики 75, 430
Экспонента матрицы 101
Экспоненциальная сходимость 48
Эллипсоид 356
Эллиптические острова 393, 394
Эллиптический интеграл 183
Эллиптическое линейное отображение 91
Энтропийная размерность 262–263
Энтропия 256, 265, 325, 342, 347, 356, 357, 392
Эргодическая теорема Биркгофа 252, 320
Эргодичность 252, 320
Явление Лиувилля 409
— Ньюхауса 360
Языки Арнольда 155
Якобиан 171
Японский меч 364

УКАЗАТЕЛЬ ИМЕН

- Алаоглу Л. (Alaoglu L.) 321
Алексеев В. М. 356
Атела П. (Atela P.) 16
Бангерт В. (Bangert V.) 399
Бартон Р. (Barton R.) 338
Бернс К. (Burns K.) 338
Биркгоф Дж. Д. (Birkhoff G. D.) 116, 191,
209, 351, 354, 374, 395, 398
Боуэн Р. Э. (Bowen R. E.) 323
Вайсс Х. (Weiss H.) 356
Ван дер Марк Дж. (van der Mark J.) 19,
20
Ван дер Пол Б. (van der Pol B.) 19, 20,
354
Вейерштрасс К. (Weierstrass K.) 119,
165, 401
Вейль Г. (Weyl H.) 400, 401
Винер Н. (Wiener N.) 338
Винтнер А. (Wintner A.) 338
Галилей Г. (Galilei G.) 8
Гельфрейх В. Г. (Gelfreich V. G.) 358
Герон Александрийский 28
Гиббс Дж. В. (Gibbs J. W.) 321
Голе К. (Golé C.) 16
Грачик Я. (Graczyk J.) 343
Дани Ш. Г. (Dani S. G.) 421
Девени Р. (Devaney R.) 220, 423
Дирихле Л. (Dirichlet L.) 400
Дрейк Ф. (Drake F.) 26
Дрот А. (Dragt A.) 350, 355
Зариский О. (Zariski O.) 421
Иомдин И. 357
Йорк Дж. (Yorke J.) 338
Кампфер Э. (Kämpfer E.) 26
Картрайт М. Л. (Cartwright M. L.) 354
Касселс Дж. В. С. (Cassels J. W. S.) 419
Кеплер И. (Kepler J.) 14, 177
Клингенберг В. (Klingenberg W.) 399
Книпер Г. (Knieper G.) 356
Кокстер Г. С. М. (Coxeter H. S. M.) 15
Колле П. (Collet P.) 345
Кронекер Л. (Kronecker L.) 400
Лагранж Ж.-Л. (Lagrange J.-L.) 189
Лазуткин В. Ф. 208
Ландфорд О. (Landford O.) 335
Лаплас П.-С. (Laplace P.-S.) 8, 9
ле Калвез П. (Le Calvez P.) 396
Левинсон Н. (Levinson N.) 354
Ли Т.-И. (Li T.-Y.) 338
Литлвуд Дж. Э. (Littlewood J. E.) 354
Лоренц Э. (Lorenz E.) 20, 225, 345, 361,
365, 368
Люстерник Л. А. 398
Мазер Дж. (Mather J.) 373, 389
Мане Р. (Mañé R.) 315
Маргулис Г. А. 420–422
Мельников В. К. 358
Мизюрович М. (Misiurewicz M.) 340,
345, 357
Мирберг П. Ю. (Myrberg P. J.) 327
Мэй Р. М. (May R. M.) 21, 338
Нгуен А. (Nguyen A.) 334
Ньютон И. (Newton I.) 8, 177
Ньюхаус Ш. (Newhouse S.) 357
Обри С. (Aubry S.) 373
Пикар Э. (Picard E.) 289
Пуанкаре А. (Poincaré H.) 9, 353, 372,
394
Рагунатан М. (Ragunathan M.) 420, 421
Ратнер М. 421, 422
Роббин Дж. (Robbin J.) 315
Робинсон Р. К. (Robinson R. C.) 315
Ром-Кедар В. (Rom-Kedar V.) 356
Рюэлль Д. (Ruelle D.) 323, 345
Сандерс Ж. (Sanders J.) 358
Свиннертон-Дайер П.
(Swinerton-Dyer P.) 420
Синай Я. Г. 323
Смейл С. (Smale S.) 241, 334, 351, 354,
363

- Такер У. (Tucker W.) 368, 370, 371
 Тёрстон У. (Thurston W.) 397
 Тихонов А. Н. 435
 Улам С. (Ulam S.) 344
 Фейгенбаум М. (Feigenbaum M.) 334, 335, 339, 340
 Фибоначчи Л. (Fibonacci L.) 13, 15, 46, 56, 97, 98, 219
 Финн Дж. М. (Finn J. M.) 355
 фон Нейман Дж. (von Neumann J.) 344
 Франкс Дж. (Franks J.) 395, 396, 399
 Фюрстенберг Х. (Furstenberg H.) 401, 421
 Хаклут Р. (Hakluyt R.) 26
 Хандель М. (Handel M.) 397
 Хинчин А. Я. 410
 Хоттон С. (Hotton S.) 16
 Чебышёв П. Л. 344
 Чуба Ш. (Chuba S.) 325
 Шарковский О. М. (Sharkovsky O. M.) 326
 Швентек Г. (Świątek G.) 329, 343
 Шерер Э. (Scherer A.) 325
 Шнирельман Л. Г. 398
 Экман Ж.-П. (Eckmann J.-P.) 346
 Эрман М. (Herman M.) 387
 Якоби К. (Jacobi C.) 400
 Якобсон М. В. 329, 346

Анатолий Борисович Каток
Борис Хасселблат

ВВЕДЕНИЕ В ТЕОРИЮ ДИНАМИЧЕСКИХ СИСТЕМ С ОБЗОРОМ ПОСЛЕДНИХ ДОСТИЖЕНИЙ

Подписано в печать 25.07.2005 г. Формат 70 × 100 ¹/₁₆. Печать офсетная. Печ. л. 29.
 Тираж 1000 экз. Заказ № 1403.

Издательство Московского центра непрерывного математического образования
 119002, Москва, Большой Власьевский пер., 11. Тел. 241-05-00.

Отпечатано с готовых диапозитивов в ППП «Типография „Наука“».
 119099, Москва, Шубинский пер., 6.

