Пол БАМБЕРГ

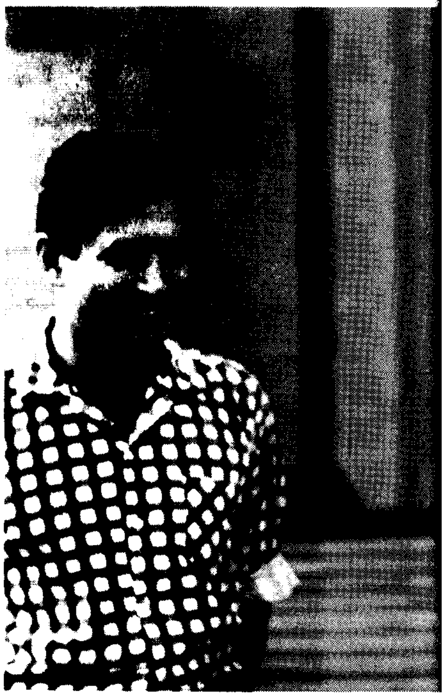
Шломо СТЕРНБЕРГ

КУРС МАТЕМАТИКИ для СТУДЕНТОВ-ФИЗИКОВ

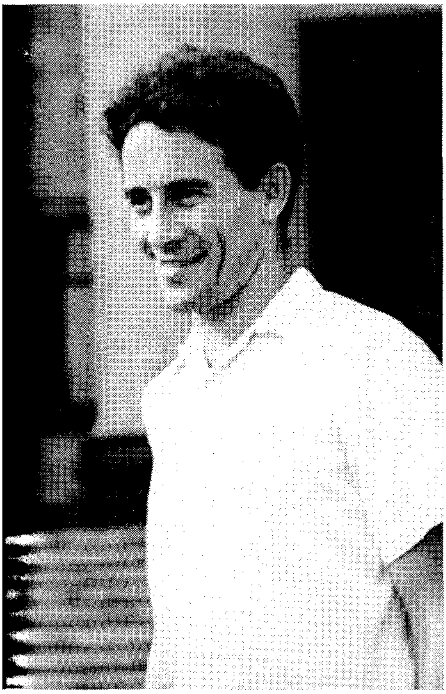
Том I

Предлагаемый курс лекций известных профессоров Гарвардского университета П. Бамберга и Ш. Стернберга пользуется большой популярностью среди физиков и инженеров, считающих для себя необходимым освоить основные элементы современной математики. Помимо традиционных вопросов этот курс содержит изложение основ топологии и геометрии, топологический смысл уравнений Максвелла, комплексный анализ, асимптотический метод Лапласа и начала термодинамики. Объединение всех этих разделов математики и физики в одной монографии, несомненно, полезно для студентов и аспирантов российских университетов и позволит им быстро включиться в проходящие во всём мире исследования в теоретической и математической физике.

Я. Г. Синай
академик РАН



Я. Г. Синай и
на ступенях МГУ
(фото



В. И. Арнольд
сентябре 1963 года
Moser)

История преподавания математического анализа, начиная с его изобретения Ньютоном, изобиловала крутыми поворотами. К середине XX века традиционные изложения стали отличать полное исключение всех геометрических понятий и представлений, с одной стороны, и полный отрыв от физики и всех приложений, с другой.
Предлагаемая книга знаменует возвращение преподавания математики к естественнонаучному подходу, для которого характерны явная геометричность и прямое представление математики как части физики, непосредственно пригодной для многочисленных приложений как в естествознании, так и в инженерном деле.
Это возвращение математики в русло естественных наук происходит сейчас и на уровне новых научных исследований, а его проявление в преподавании давно стало абсолютно необходимым.

В. И. Арнольд
академик РАН

Пол БАМБЕРГ
Шломо СТЕРНБЕРГ

КУРС МАТЕМАТИКИ для СТУДЕНТОВ-ФИЗИКОВ

Том I

Главы 1–11

перевод с английского
к. ф.-м. н., доц. Л. П. Котовой (МФТИ)
под редакцией
к. ф.-м. н., доц. Е. В. Майкова (МГУ)

МОСКВА
2006

УДК 51

Издание осуществлено при поддержке
Центрального аэрогидродинамического
института имени Н. Е. Жуковского,
директор В. Г. Дмитриев
и
Фонда «КНИГА–НАУКА–КУЛЬТУРА»,
директор В. Б. Филиппов

Бамберг П., Стернберг Ш.
Курс математики для студентов-физиков. Том I, главы 1–11.
Перев. с англ. Л. П. Котовой под ред. Е. В. Майкова.
— М.: ФАЗИС, 2006. 592 с. (I–XVIII, 1–574)
ISBN 5-7036-0109-6

Paul Bamberg & Shlomo Sternberg
A Course in Mathematics for Students of Physics

Издательство ФАЗИС
123557 Москва, Пресненский вал, 42-44
e-mail: publisher@phasis.ru *http*://www.phasis.ru

Типография «Наука» Академиздатцентра РАН
121099 Москва Г-99, Шубинский пер., 6
Заказ № 3638

ISBN 5-7036-0109-6

Оглавление тома I

Предисловие

В этой книге излагается курс математики, который мы преподавали в Гарварде в течение восьми лет. Курс предназначен для студентов, интересующихся физикой и имеющих хорошую подготовку по математическому анализу функций одной переменной. Полезно некоторое знакомство с линейной алгеброй, но это не обязательно. Большинство наших студентов одновременно с этим курсом изучают сложные курсы по физике, так что они могут соединить излагаемые здесь сведения со своими физическими знаниями. Такое совмещение полезно, но не обязательно. Основное содержание нашего курса — теория и физические приложения линейной алгебры и математического анализа функций нескольких переменных, включая внешнее исчисление. Наш педагогический метод — изложение по «спирали», т. е. мы рассматриваем одну и ту же тему несколько раз, возвращаясь к ней на все более высоком уровне, одновременно расширяя области ее применения. Такое построение мы предпочитаем «прямолинейному» строгому логическому порядку. Надеемся, что при этом мы избежим логических ошибок типа «порочного круга», но часто будем начинать с простого случая того или иного понятия, а затем возвращаться к нему для обобщений, делая это только после ознакомления с родственными понятиями и, как следствие, приобретения более широкого взгляда на предмет. Такой подход требует от студентов определенного доверия и терпения. Но мы надеемся, что, в конце концов, они будут вознаграждены более глубоким интуитивным пониманием предмета в целом.

Теперь расскажем о содержании книги более детально. Цель первых четырех глав — познакомить читателя с алгеброй и анализом квадратных матриц. Таким образом, изучив эти главы, студент должен научиться думать о матрицах как о самостоятельных объектах, а не просто как о наборе чисел, выстроенных по квадрату. В этих главах мы рассматриваем почти исключительно матрицы 2×2, когда самые сложные вычисления сводятся к решению квадратных уравнений. Но каждый раз мы формулируем результаты, имея в виду матрицы более высокого порядка.

Глава 1 начинается с объяснения связи закона умножения матриц 2×2 и геометрии прямых линий на плоскости. Далее излагается алгебра матриц 2×2, обсуждается детерминант и его отношение к плоскости и ориентации. Мы вводим определение абстрактного векторного пространства в общем случае, объясняем понятие базиса и замены базиса в одномерном и двумерном векторном пространстве.

В главе 2 мы обсуждаем конформную линейную геометрию на плоскости, т.е. геометрию линий и углов и их связь с определенного вида матрицами 2×2. Кроме того, вводятся понятия собственных значений и собственных векторов, столь важные в квантовой механике. Мы пользуемся этими понятиями, чтобы построить алгоритм вычисления степеней матрицы. В качестве примера, где это применяется, обсуждаются основные свойства цепей Маркова.

Основная задача главы 3 — показать, что система однородных линейных дифференциальных уравнений с постоянными коэффициентами может быть записана в виде $d\mathbf{u}/dt = A\mathbf{u}$, где A — матрица, а $\mathbf{u}$ — вектор, и что решение этой системы имеет вид $e^{At}\mathbf{u}_0$, где $\mathbf{u}_0$ задается начальными условиями. При этом, конечно, требуется объяснить, что означает экспонента, в показателе которой стоит матрица. Кроме того, описывается качественное поведение решений в неоднородном случае, в том числе обсуждается явление резонанса.

Глава 4 посвящена изучению скалярных произведений и квадратичных форм. В ней имеется много физических приложений, в том числе обсуждение нормальных форм и детальное изложение специальной теории относительности.

В главах 5 и 6 изложены основные положения дифференциального исчисления.

В главе 5 определяется дифференциал отображения одного векторного пространства на другое, обсуждаются его основные свойства, в частности, правило дифференцирования композиции. В качестве физических приложений рассматриваются задача Кеплера и приближение Борна. Мы даем определения частной производной и производной по направлению, определяем линейные дифференциальные формы.

В главе 6 продолжается изложение дифференциального исчисления. Мы представляем в векторной форме теорему о среднем значении, формулу Тейлора и теорему об обратном отображении. Обсуждаем поведение критических точек и множители Лагранжа.

Главы 7 и 8 знакомят читателя с интегральным исчислением.

Глава 7 посвящена изучению линейных дифференциальных форм и их интегралов. Особое внимание уделено замене переменных. Кроме того, обсуждаются и другие одномерные интегралы, например, длина дуги.

Глава 8 посвящена изучению внешних форм второй степени и соответствующих им двумерных интегралов. Вводится понятие внешней производной и подчеркивается инвариантность относительно замены переменных. Доказывается двумерный вариант теоремы Стокса, т. е. теорема Грина. Рассматриваются поверхностные интегралы в трехмерном пространстве.

В главе 9 результаты первых восьми глав используются при построении физической теории — оптики. Вся глава посвящена приложениям, и ее можно опустить без ущерба для понимания последующего материала.

В главе 10 мы возвращаемся назад и доказываем основные свойства конечномерных векторных пространств и их линейных преобразований. Обычно это непосредственное обобщение результатов, полученных в первых четырех главах в двумерном случае. Но при этом дан новый алгоритм редукции матрицы по строкам. Вводятся два новых понятия — дуальное пространство и факторпространство (в начале курса их было трудно определить). В дальнейшем эти понятия окажутся чрезвычайно важными.

Глава 11 посвящена свойствам детерминантов матриц $n \times n$. Тема рассмотрена аксиоматически, показаны основные вычислительные алгоритмы.

Главы 12–14 знакомят читателя с математикой формы, т. е. с алгебраической топологией.

В главе 12 мы начинаем изучение электрических цепей. При этом рассматриваются два аспекта. Во-первых, изучаются «провода» цепи, т. е. как соединяются между собой различные ветви. На математическом языке это называется топологией одномерных комплексов. Во-вторых, мы изучаем реакцию цепи в целом, если нам известно поведение отдельных ветвей, в частности изменение мощности и энергии. Мы приводим несколько примеров цепей, интересных с физической точки зрения.

В главе 13 мы продолжаем изучение цепей. Исследуем краевые задачи, связанные с емкостными цепями, и применяем изложенные методы для решения ряда классических задач электростатики, включающих проводники.

В главе 14 мы кратко показываем, как результаты, полученные в главах 12 и 13 в одномерном случае, обобщаются на случай большего числа измерений.

Главы 15–18 посвящены разработке внешнего дифференциального исчисления, рассматриваемого как непрерывная версия дискретной теории комплексов.

В главе 15 излагаются основные сведения о внешнем исчислении: внешняя алгебра, k-формы, замена переменных, внешняя производная и теорема Стокса.

Глава 16 посвящена электростатике. Мы рассматриваем диэлектрические свойства вакуума как непрерывный аналог емкости цепи, причем эти диэлектрические свойства определяют евклидову геометрию в трехмерном случае. Излагаем основные положения теории потенциалов.

В главе 17 мы продолжаем изучение внешнего дифференциального исчисления. Основные темы: векторные поля и потоки, внутренние произведения и производные Ли. Полученные результаты применены для решения задач магнитостатики.

Глава 18 заканчивает изучение внешнего исчисления углубленным обсуждением общего случая звездного оператора.

Главу 19 можно считать кульминацией всего курса. В ней результаты, полученные в предыдущих главах, применяются для исследования уравнений Максвелла и связанных с ними волновых уравнений.

Главы 20 и 21 никак не связаны с главами 9–19, и их можно читать независимо. Обычно они не включаются в наш годовой курс. А вот главы 1–9, 20 и 21 могут составить самостоятельный краткий курс математики.

Материал главы 20 представляет собой относительно стандартное изложение теории функций комплексного переменного, которое вполне доступно для студентов, подготовленных к чтению этой книги.

В главе 21 обсуждаются простейшие аспекты асимптотик.

В главе 22 показано, как, следуя идеям Бора и Каратеодори, внешнее исчисление можно использовать в классической термодинамике.

Большинство математиков и все физики, результаты которых представлены в этой книге, работали в первом десятилетии двадцатого века. Это значит, что изложенному материалу уже по крайней мере 90 лет. И, тем не менее, он еще не вошел в наши элементарные курсы математики, хотя бóльшая его часть (за исключением, может быть, теории цепей) должна изучаться современными физиками, что они и делают на определенном этапе своей карьеры. Причины этого, в значительной мере, исторические. У. Гамильтону (1805–1865) было очевидно, что вещественных и комплексных чисел недостаточно для глубого изучения геометрического анализа, что в декартовой геометрии дву- и трехмерного пространства возникает желание рассматривать числа парами и тройками, как самостоятельные объекты с собственными алгебраическими свойствами. Для этого он разработал алгебру кватернионов, которая была очень популярна в Англии в середине девятнадцатого века. Однако, теория имела ряд изъянов: кватернионы более естественны в пространстве четырех, а не трех измерений. Геометрия трех измерений возникала как часть более общей конструкции, а не как самостоятельная теория. Кроме того, кватернионы имели слишком сложную алгебраическую структуру. Например, очень громоздкой была связь между умножением кватернионов и геометрическими построениями в трехмерном пространстве. (Конечно, первый недостаток сегодня считался бы не очень серьезным, но вместо него возникло бы другое возражение: теория *ограничена* четырьмя измерениями.) В конце концов, *векторная алгебра* в трехмерном пространстве с ее скалярным и векторным произведениями была очищена от теории кватернионов. Ее объединили с необходимыми дифференциальными операциями, в результате чего появился *векторный анализ*, окончательно разработанный Дж. Гиббсом (1839–1903) и обнародованный им в своей знаменитой книге «Элементы векторного анализа», оказавшей сильное влияние на дальнейшее развитие науки.

Итак, векторный анализ с его операциями grad, div, curl и т. д. стал стандартным языком изложения геометрических законов физики. Однако, несмотря на то, что векторный анализ прекрасно описывает геометрию трехмерного пространства, он тоже имеет ряд существенных недостатков. Первый — и наименее серьезный — состоит в том, что в нем не просматривается очень важное единство: например, не ясно, что фундаментальная теорема исчисления одна — сейчас ее называют общей теоремой Стокса, а теоремы Грина, Гаусса и классическая теорема Стокса являются ее частными случаями. В рамках векторного анализа мы не можем это увидеть. Более серьезный недостаток состоит в том, что фундаментальные операторы (например, grad, div) предполагают евклидову структуру пространства, т. е. подразумевают трехмерное пространство и ориентацию (например, curl). Это значит, что теория работает только в трехмерном ориентированном евклидовом пространстве. При этом возникает еще одна проблема — при произвольном изменении координат операторы ведут себя неприятно, их выражения в неортогональных координатах очень громоздкие. Уже А. Пуанкаре (1854–1912) в своих фундаментальных научных и философских работах, предшествовавших теории относительности, подчеркивал необходимость делать различие между «топологическими» законами геометрии и физики, т. е. законами, которые зависят только от дифференциальной структуры пространства и поэтому инвариантны относительно гладких преобразований, и законами, зависящими от геометрической структуры, например, от понятия расстояния. Теория относительности оказала огромное влияние на развитие математики главным образом потому, что она поощряла изучение теории многомерных пространств, которая уже существовала в математической литературе, но не считалась важной для изучения геометрии. Кроме того, необходимо было исследовать произвольные изменения координат. Векторный анализ не мог решить эти две задачи и поэтому позднее его дополнили *тензорным анализом*. Но и тензорный анализ с его обилием индексов тоже имеет много серьезных недостатков. Главный изъян состоит в том, что невероятно трудно сказать, какие операции имеют общее геометрическое значение, а какие связаны только с данной системой координат. Таким образом, тензорный анализ очень хорош для вычислений, но трудно сказать, что́ именно вычисляется. Вся работа, начатая еще Гамильтоном и направленная на создание математического аппарата, в котором объекты имеют геометрический смысл, оказалась напрасной. Для того, чтобы заставить теорию работать, необходимо было ввести аффинную связь — довольно изощренную геометрическую конструкцию. Но даже с такой конструкцией геометрическое значение операций не стало прозрачным. По существу интуитивно ясный векторный анализ так и не был вытеснен тензорным анализом из программ начального университетского обучения.

Математики в основном уже согласились, а физики постепенно привыкают к тому, что самым удобным аппаратом для геометрического анализа является внешнее дифференциальное исчисление Грассмана и Картана. Его преимущество состоит в том, что наряду с простыми и четкими правилами вычисления все объекты имеют прозрачный геометрический смысл. Внешнее исчисление работает при любом числе измерений, хорошо себя ведет при отображениях и изменениях системы координат, обладает существенным единством основных теорем и четко различает «топологические» и «метрические» свойства. В рамках внешнего анализа геометрические законы физики принимают простую и элегантную форму. Чтобы это продемонстрировать, полезно привести таблицу, взятую из книги «Курс математической физики», написанную Тиррингом.

Таблица 1. **История записи уравнений Максвелла.** Константы c, μ_0, и ϵ_0 положены равными 1.

Однородное уравнение	**Неоднородное уравнение**
Самая ранняя форма	
$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0$ $\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\dot{B}_x$ $\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -\dot{B}_y$ $\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -\dot{B}_z$	$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \rho$ $\frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} = j_x + \dot{E}_x$ $\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} = j_y + \dot{E}_y$ $\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} = j_z + \dot{E}_z$
Конец XIX века	
$\mathbf{\nabla} \cdot \mathbf{B} = 0$ $\mathbf{\nabla} \times \mathbf{E} = -\dot{\mathbf{B}}$	$\mathbf{\nabla} \cdot \mathbf{E} = \rho$ $\mathbf{\nabla} \times \mathbf{B} = \mathbf{j} + \dot{\mathbf{E}}$
Начало XX века	
$*F^{\beta\alpha}{}_{,\alpha} = 0$	$F^{\beta\alpha}{}_{,\alpha} = J^\beta$
Середина XX века	
$dF = 0$	$\delta F = J$

В 1844 году Г. Грассман (1809–1877) опубликовал книгу «Ausdehnungslehre». Математическое сообщество не оценило эту работу и ведущие немецкие математики того времени проигнорировали ее. Фактически, Грассман так никогда и не получил место в университете; всю свою жизнь он преподавал математику в школе. Однако, похоже, что он прожил счастливую и плодотворную жизнь. У Грассмана была большая семья и он считался признанным экспертом по санскритской литературе. В конце жизни он попробовал еще раз выпустить новое издание «Ausdehnungslehre», но оно также не имело успеха. Только немногие из математиков того времени оценили его работу. Среди них был К. Мебиус (1825–1908). Тем не менее, в книге «Ausdehnungslehre» впервые появились ключевые в современной математике понятия и большинство алгебраических структур, используемых в этой книге. В ней были определены: векторное пространство, внешняя алгебра, внешнее и внутреннее произведение, теорема Стокса в обобщенной форме.

В настоящее время общепризнано, что Э. Картан (1869–1951) является крупнейшим специалистом по геометрии в XX веке. Его ранние работы, относящиеся к группам Ли и дифференциальным уравнениям в частных производных, имеющие огромное значение для современной математики, тоже сначала не были поняты. Но в 1920 году они стали известны широкому кругу ученых. К этому времени теория групп Ли начала играть ключевую роль в математике и физике, и Г. Вейль заново представил их в своих статьях. В настоящее время работы Картана по основным узлам и связям (известные под общим названием «калибровочные теории») являются основой теории элементарных частиц. В 1922 году Картан опубликовал книгу «Лекции по интегральным инвариантам», где показал, что изобретенное им внешнее дифференциальное исчисление является гибким инструментом не только для геометрии, но и для вариационного исчисления с широким кругом физических приложений. Как уже было сказано, признание пришло не сразу, но сегодня и физики, и математики считают, что внешнее исчисление позволяет наилучшим образом формулировать геометрические законы физики. И нам кажется, что пришла пора заменить им векторное исчисление в программах начального университетского обучения.

Необходимо пояснить, почему в книге так много внимания уделено теории электрических цепей, хотя обычно эта тема не входит в учебные программы по математике. Прежде всего, это оправдано с педагогической точки зрения. Данная тема всегда нравится студентам. Она позволяет на очень приземленном уровне ввести понятия дуального пространства и факторпространства. Иначе эти понятия кажутся студентам слишком абстрактными и поэтому трудно воспринимаются ими. Кроме того, в дискретном, алгебраическом подходе к теории цепей теорема

Стокса появляется, по существу, как естественное определение, что позволяет понять роль оператора d и теоремы Стокса при формулировке внешнего исчисления. Конечно, есть более глубокие, философские причины, почему мы приняли решение подробно рассказать о теории цепей. Уже почти сто лет, как люди думают, что силы, удерживающие макроскопические тела, имеют электрическую природу. Таким образом, понятие твердого тела и отсюда евклидова геометрия вытекают из электростатики (конечно, в нерелятивистском приближении, когда понятия твердого тела и геометрия Евклида имеют смысл). Физические исследования как предельно малых объектов (элементарных частиц), так и предельно больших (космология) заново ставят фундаментальные вопросы о геометрии пространства и времени. Мы полагаем, что разумно уже в самом начале изучения математики дать студентам представления о связи геометрии и физики. Без сомнения, сторонники компьютеров и ряда современных теорий в физике зададут вопросы о дискретности, противопоставляемой непрерывному характеру пространства и времени (проблема, поднятая Риманом в его диссертации об основах геометрии). Надеемся, что наше обсуждение может быть полезно тем, кто в будущем станет решать эту проблему.

Наш курс математики читается в течение одного года, и поэтому мы вынуждены были опустить некоторые важные темы. Мы не обсуждаем бесконечномерные векторные пространства, в частности, гильбертовы пространства; не даем определения и не изучаем абстрактные дифференцируемые многообразия и их свойства. Наш опыт показал, что эти темы слишком сложны для студентов, и усилия, которые необходимо затратить на их объяснение, лучше использовать по-другому. Конечно, в книге есть места, где нам приходится компенсировать отсутствие этих понятий. Однако, более серьезно то, что в книге нет подробного обсуждения гармонического анализа Фурье, классической механики и теории вероятностей. Мы только слегка касаемся этих тем, не давая подробного изложения. Нас может извинить только то, что каждая из этих тем потребовала бы семестрового курса, а их подробное изложение на современном уровне легко найти. В конце книги приводится список рекомендуемой литературы.

Нам приятно поблагодарить профессора Даниэла Гороффа, тщательно прочитавшего рукопись, за многочисленные исправления и полезные предложения. Мы искренне благодарим Джин Моррис за отличное качество набора, за то, что она внимательно сопровождала производство книги на протяжении восьми лет — с самого начала проекта до его завершения.

От издателя русского перевода

Еще восемь лет потребовалось для выхода в свет русского варианта этой книги, выполненного типографским способом. На пути русского издания оказалось много препятствий. Невыполнение Российским фондом фундаментальных исследований своих уставных обязанностей по финансированию издания выигравших конкурс произведений (проект №98-01-14152), последствия преступного дефолта августа 1998 года, — эти и другие финансовые трудности были преодолены благодаря энтузиазму тех, кто работал над русским переводом двухтомника.

Однако не только финансовые проблемы снижали скорость подготовки русского издания. Оказалось, что английский оригинал содержит изрядное количество неточностей и ошибок. Поэтому необходимо особо отметить труд переводчика и редакторов, которые провели тщательную работу по адаптации данного курса к практике российского образовательного процесса. Часть ошибок была найдена Л. П. Котовой уже при переводе. Многие неточности и ошибки устранены Е. В. Майковым при его скрупулезном редактировании первого тома. Некоторые исправления в первый том внес и редактор второго тома А. Н. Якивчик; его же трудами в списке рекомендуемой литературы указаны теперь точные библиографические данные последних изданий.

Тем не менее, не исключено, что наш уважаемый Читатель заметит в книге отдельные опечатки, неточности, ошибки. И хотя кое-кто считает, что не следует устранять опечатки, чтобы не лишать читателя удовольствия находить ошибки*, мы все же постоянно ведем работу по уточнению текста книги и будем признательны читателям за соответствующие замечания (наш адрес: `publisher@phasis.ru`).

Москва, май 2006

*См., например: «Задачи Арнольда» (М.: ФАЗИС, 2000, с. X).

Глава 1

Линейные преобразования плоскости

В главе 1 мы рассказываем о связи закона умножения матриц 2×2 с геометрией прямых на плоскости. Излагаем алгебру матриц 2×2, изучаем определитель и его связь с площадью и ориентацией. Вводим понятие абстрактного векторного пространства в общем случае, объясняем понятия базиса и замены базиса в случае одномерного и двумерного векторного пространства.

1.1. Аффинные плоскости и векторные пространства

Евклидова плоскость, знакомая из школьной планиметрии, возникла на раннем этапе развития математики. Физические эксперименты на столе или школьной доске легко демонстрируют ее свойства. Благодаря использованию линейки и транспортира мы привыкаем считать «длину» и «угол» такими же фундаментальными понятиями, как «точка» и «прямая». Занимаясь чистой математикой или ее приложениями к физике и другим дисциплинам, мы часто рассматриваем плоскости, для которых прямые определены, но нет общего определения длины, или когда обычное евклидово определение длины не работает. Такая плоскость может быть изображена на листе бумаги, но физическое расстояние между двумя точками, измеренное линейкой, или угол между двумя прямыми, измеренный транспортиром, может не иметь никакого смысла.

В качестве примера приведем плоскость, где графически изображается движение частицы вдоль оси x (рис. 1.1). Точки P и Q на этой плоскости представляют физическое понятие *события*, определяемого местом и временем. Линия l тоже имеет физический смысл; она соответствует движению частицы, на которую не действует никакая сила. Мы можем сравнивать длины отрезков вдоль оси t (временны́е интервалы) или вдоль оси x (расстояния). Однако измеренное линейкой расстояние между точками P и Q лишено всякого физического смысла. Точно так же не имеет физического смысла начало координат такой плоскости, т. е. точка пересечения осей.

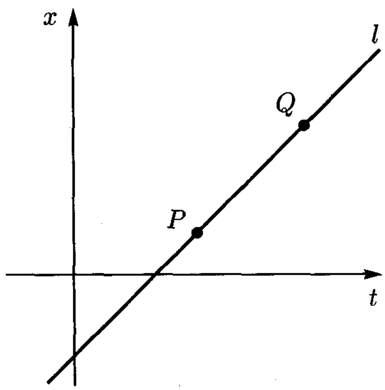


Рис. 1.1

Такой ситуации (и многим другим «двумерным» ситуациям) соответствует математическое понятие *действительной аффинной плоскости*, т. е. такой плоскости, которая представляет много двумерных ситуаций. *Точки* на аффинной плоскости обозначаем заглавными латинскими буквами P, Q и т. д., а прямые линии, в дальнейшем называемые просто *прямыми*, строчными буквами l, m и т. д. Следуя Декарту, считаем, что аффинную плоскость образует множество всех пар действительных чисел. Точка на такой плоскости — упорядоченная пара действительных чисел, обозначаемая $\begin{bmatrix} x \\ y \end{bmatrix}$. Эти числа называют *координатами* точки. Такую плоскость назовем $A\mathbb{R}^2$ (рис. 1.2). Здесь A говорит нам, что плоскость аффинная, и мы не выделяем на ней начала координат. Буква $\mathbb{R}$ соответствует множеству действительных чисел, а верхний индекс два указывает, что мы рассматриваем пары действительных чисел. (Как обычно, изображая плоскость на бумаге, мы откладываем x по горизонтальной оси, а y определенного вида — вертикально вверх вдоль перпендикулярной оси. Однако, в настоящий момент понятия «перпендикулярности» или величины угла не определены. С одинаковым успехом можно откладывать x и y вдоль любых[1] осей.) Прямая — это определенного вида множество точек. Мы предполагаем, что в результате изучения геометрии читатель знаком с прямыми и, в частности, знает, как описываются прямые в аналитической геометрии.

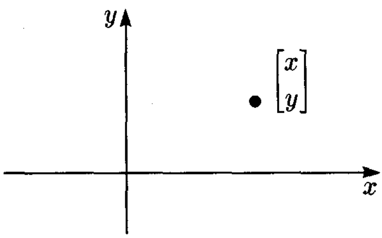


Рис. 1.2

Прямые на аффинной плоскости $A\mathbb{R}^2$ могут описываться поразному. Например, уравнением, которому удовлетворяют все точки, лежащие на прямой,

$$l = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| \, ax + by = c \right\}.$$

Формулу следует читать так: «l — это множество точек $\begin{bmatrix} x \\ y \end{bmatrix}$, координаты которых удовлетворяют уравнению $ax + by = c$». Предполагается при этом, что a и b одновременно не равняются нулю.

Такой способ описания прямой не очень удобен, потому что параметры прямой a, b, c не определяются прямой однозначно.

[1]Не параллельных друг другу. — *Прим. ред.*

Например, формулы

$$\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| ax + by = c \right\} \quad \text{и} \quad \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| 3ax + 3by = 3c \right\}$$

определяют одну и ту же прямую. Вообще параметры ra, rb, rc при $r \neq 0$ описывают ту же прямую, что и a, b, c.

Второй способ задания прямой состоит в том, что на плоскости $A\mathbb{R}^2$ выбираются две различные точки $P_0 = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$, $P_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$, и через них проводят прямую. Тогда точки, лежащие на такой прямой, определяются формулой

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_0 + t(x_1 - x_0) \\ y_0 + t(y_1 - y_0) \end{bmatrix},$$

где параметр t — любое действительное число. Это описание прямой еще более громоздко, чем предыдущее: мы можем заменить выбранные точки P_0 и P_1 любой другой парой точек, принадлежащих этой прямой.

Есть более удобный способ описания прямой («динамический» в отличие от «статического» описания). Мы задаем точку на прямой и «направляющий вектор прямой». Таким образом, набор точек, определяемых по формуле $\left\{ \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + t \begin{pmatrix} u \\ v \end{pmatrix} \middle| t \in \mathbb{R} \right\}$, где $\begin{pmatrix} u \\ v \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ — заданный вектор, и есть прямая. (Представим себе частицу, движущуюся вдоль этой траектории с вектором скорости $\begin{pmatrix} u \\ v \end{pmatrix}$, находившуюся в точке $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ в нулевой момент времени). Здесь для описания прямой мы используем четыре параметра. При этом можно умножить $\begin{pmatrix} u \\ v \end{pmatrix}$ на не равный нулю скаляр и получить ту же прямую (это будет движение с другой скоростью) или можем сдвинуть начальную точку $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ вдоль прямой. В этом случае мы имеем два избыточных параметра.

Конечно, эта формула переходит в формулу, полученную вторым способом,

$$
\begin{aligned}
u &= x_1 - x_0,\\
v &= y_1 - y_1.
\end{aligned}
$$

Ну, а теперь опишем четвертый, привычный способ задания прямой. В нем нет избыточных параметров, его неудобство состоит в отсутствии единообразия записи для различных прямых. Если a и b — любые действительные числа, то набор точек

$$l = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| \, y = ax + b \right\}$$

задает прямую, пересекающую ось y в точке $\begin{bmatrix} 0 \\ b \end{bmatrix}$ и имеющую «угловой» коэффициент a, т. е. для всех точек прямой увеличение x на единицу влечет за собой увеличение y на величину a. Прямая задается простым уравнением, параметры a и b определяют геометрические свойства прямой. Но не все прямые могут быть заданы таким образом. Необходимо добавить прямые, параллельные оси y:

$$l = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| \, x = d \right\}.$$

Следуя строгой логике, мы должны принять один из этих четырех способов описания за *определение* прямой. Например, следует сказать: по определению, прямая l — это набор точек на плоскости $A\mathbb{R}^2$,

$$l = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| \, ax + by = c \right\},$$

где три параметра a, b и c — действительные числа, причем a и b одновременно не равны нулю. Теперь нам следовало бы доказать, что этот же набор точек может быть получен и в трех других способах описания. Но здесь мы не станем углубляться в логические тонкости, которые можно реализовать в рамках элементарной аналитической геометрии.

Важно помнить, что в аффинной плоскости не выделено начало координат, и бессмысленно складывать точки на такой плоскости. Точка $\begin{bmatrix}0\\0\end{bmatrix}$ не имеет особого значения, и мы устоим перед соблазном складывать точки $\begin{bmatrix}2\\1\end{bmatrix}$ и $\begin{bmatrix}3\\6\end{bmatrix}$, суммируя их координаты. Однако, существует близкое понятие двумерного *векторного пространства*, где операция сложения определена. Мы построим векторное пространство из аффинной плоскости, вводя для любой пары ее точек вектор смещения $\overrightarrow{PQ}$, такой, что «хвост» вектора находится в точке P, а его «голова» — в точке Q (рис. 1.3). Такие векторы будем обозначать жирными строчными латинскими буквами $\mathbf{v}$, $\mathbf{w}$ и т. д. Вектор $\mathbf{v}$ определяется парой действительных чисел, например, $\mathbf{v} = \begin{pmatrix}5\\2\end{pmatrix}$. (Заметим, что для векторов мы используем круглые скобки ( ), а для точек — квадратные [ ]).

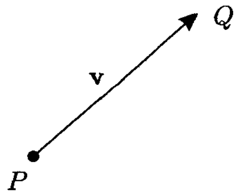


Рис. 1.3

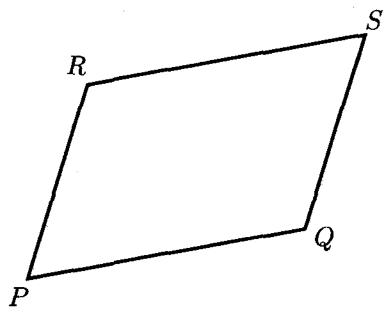


Рис. 1.4

Например, вектор $\mathbf{v} = \begin{pmatrix}5\\2\end{pmatrix}$ следует понимать как смещение, которое переводит точку $\begin{bmatrix}1\\3\end{bmatrix}$ в точку $\begin{bmatrix}6\\5\end{bmatrix}$, точку $\begin{bmatrix}-3\\2\end{bmatrix}$ в точку $\begin{bmatrix}2\\4\end{bmatrix}$, и вообще, произвольную точку $P = \begin{bmatrix}x\\y\end{bmatrix}$ в точку $Q = \begin{bmatrix}x+5\\y+2\end{bmatrix}$. Таким образом, вектор $\mathbf{v}$ определяет некое преобразование аффинной плоскости, а именно, жесткую трансляцию всей плоскости. Если P — любая точка аффинной плоскости, то смещенную точку Q обозначаем P "+" $\mathbf{v}$: знак "+" обозначает действие вектора на точки. Таким образом, вектор $\mathbf{v}$ переводит точку P в точку $Q = P$ "+" $\mathbf{v}$. В координатной форме это выглядит так: если $P = \begin{bmatrix} x \\ y \end{bmatrix}$ и $\mathbf{v} = \begin{pmatrix} a \\ b \end{pmatrix}$, то P "+" $\mathbf{v} = \begin{bmatrix} x+a \\ y+b \end{bmatrix}$.

Мы ставим кавычки вокруг знака "+", потому что в этом случае операция производится над объектами различной природы, над точками и векторами, и поэтому отличается от обычной процедуры сложения. Далее, для любой пары точек P и Q определим единственный вектор $\mathbf{v} = Q$ "−" P такой, что

$$P \text{ “+” } \mathbf{v} = Q.$$

Вокруг знака минус мы тоже ставим кавычки, потому что здесь тоже участвуют объекты разного рода — из пары точек получается вектор. Нарисовав несколько примеров на бумаге в клеточку, вы можете самостоятельно убедиться, что две пары точек P, Q и R, S определяют один и тот же вектор (т. е. Q "−" $P = S$ "−" R) тогда и только тогда, когда векторы $\overrightarrow{PQ}$ и $\overrightarrow{RS}$ являются противоположными сторонами параллелограмма (рис. 1.4). Поэтому часто говорят, что вектор определяется величиной и направлением. Но здесь мы не будем вводить понятия величины и направления вектора, потому что они не являются инвариантными.

По определению, суммой двух векторов $\mathbf{u} = \begin{pmatrix} a \\ b \end{pmatrix}$ и $\mathbf{v} = \begin{pmatrix} c \\ d \end{pmatrix}$ является вектор

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix} a+c \\ b+d \end{pmatrix}.$$

Заметим, что

$$P \text{ “+” } (\mathbf{u} + \mathbf{v}) = (P \text{ “+” } \mathbf{u}) \text{ “+” } \mathbf{v}, \tag{1.1}$$

так как, если $\mathbf{u} = \begin{pmatrix} a \\ b \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} c \\ d \end{pmatrix}$ и $P = \begin{bmatrix} x \\ y \end{bmatrix}$, то и левая, и правая части равенства (1.1) определяют точку $\begin{bmatrix} a+c+x \\ b+d+y \end{bmatrix}$. Формула (1.1) означает, что смещение, соответствующее $\mathbf{u} + \mathbf{v}$, может быть получено последовательным применением операций смещения $\mathbf{u}$ и $\mathbf{v}$. При этом заметим, что $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$. Мы можем в обозначениях между $\mathbb{R}^2$ и $A\mathbb{R}^2$. Дело в том, что в пространстве $\mathbb{R}^2$ точка $\begin{pmatrix}0\\0\end{pmatrix}$ имеет особое значение (нейтральный элемент), и имеет смысл сложение двух векторов. В пространстве $A\mathbb{R}^2$ ничего этого нет.

1.2. Векторные пространства и их аффинные пространства

Легко проверить, что операция сложения векторов в пространстве $\mathbb{R}^2$ и операция умножения векторов на действительные числа удовлетворяют следующим аксиомам.

Сложение векторов

Ассоциативность сложения: $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$.

Коммутативность сложения: $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.

Существование нулевого элемента: существует такой вектор $\mathbf{0}$, что $\mathbf{0}+\mathbf{v}=\mathbf{v}$ для всех векторов $\mathbf{v}$.

Существование противоположного элемента: для любого вектора $\mathbf{v}$ существует $-\mathbf{v}$ такой, что $\mathbf{v}+(-\mathbf{v})=\mathbf{0}$.

Умножение векторов на действительные числа

Свойство числа 1: $1\mathbf{v}=\mathbf{v}$ для любого $\mathbf{v}$.

Ассоциативность и дистрибутивность умножения на число: для любых действительных чисел r и s и любых векторов $\mathbf{u}$ и $\mathbf{v}$

$$(rs)\mathbf{v}=r(s\mathbf{v}),$$
$$(r+s)\mathbf{v}=r\mathbf{v}+s\mathbf{v},$$
$$r(\mathbf{u}+\mathbf{v})=r\mathbf{u}+r\mathbf{v}.$$

Эти аксиомы известны как аксиомы *векторного пространства*. По определению, векторное пространство — это такое множество V объектов, называемых векторами $\mathbf{u}$, $\mathbf{v}$ и т. д., в котором введены операция сложения, означающая, что каждой паре векторов $\mathbf{u}$ и $\mathbf{v}$ соответствует третий вектор $\mathbf{u}+\mathbf{v}$, и операция умножения такая, что каждому действительному числу t и каждому вектору $\mathbf{v}$ соответствует вектор $t\mathbf{v}$. При этом должны выполняться вышеперечисленные аксиомы.

Итак, мы показали, что $\mathbb{R}^2$ является векторным пространством. В качестве второго примера рассмотрим пространство $\mathbb{R}^3$, в котором вектор является триплетом, состоящим из действительных чисел:

$$\mathbf{v} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

Сложение векторов производится сложением компонент, как и в пространстве $\mathbb{R}^2$:

$$\text{если} \quad \mathbf{v}_1 = \begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix} \quad \text{и} \quad \mathbf{v}_2 = \begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix}, \quad \text{то} \quad \mathbf{v}_1 + \mathbf{v}_2 = \begin{pmatrix} a_1 + a_2 \\ b_1 + b_2 \\ c_1 + c_2 \end{pmatrix}.$$

Пространство $\mathbb{R}^3$ — это знакомое нам трехмерное пространство векторов. Понятие размерности пространства мы обсудим позднее.

Все действительные числа тоже можно рассматривать как векторное пространство $\mathbb{R} = \mathbb{R}^1$, в котором процедурами сложения и умножения являются обычные операции сложения и умножения чисел. Когда мы определим понятие размерности пространства, станет ясно, что $\mathbb{R}^1$ — это пример одномерного векторного пространства.

В качестве другого примера векторного пространства рассмотрим множество всех полиномов. Мы умеем складывать полиномы, просто складывая коэффициенты при одинаковых степенях x, например:

$$(1 + 3x + 7x^2) + (2 - x^2 + x^4 - x^6) = 3 + 3x + 6x^2 + x^4 - x^6,$$

Мы умеем умножать полиномы на действительное число:

$$7(1 + 3x + 3x^2) = 7 + 21x + 21x^2.$$

Легко проверить, что в этом случае все аксиомы векторного пространства тоже выполняются. Далее, можно рассматривать пространство полиномов не выше заданной степени. Например, наиболее общий вид полинома не выше второй степени таков:

$$P = ax^2 + bx + c.$$

Сумма двух полиномов

$$P_1 = a_1x^2 + b_1x + c_1 \quad \text{и} \quad P_2 = a_2x^2 + b_2x + c_2$$

равна

$$P_1 + P_2 = (a_1 + a_2)x^2 + (b_1 + b_2)x + (c_1 + c_2).$$

Например, если

$$P_1 = 3x^2 + 2x + 1, \quad P_2 = 7x^2 - 10x + 2,$$

то

$$P_1 + P_2 = 10x^2 - 8x + 3.$$

Множество всех полиномов степени не выше двух также является векторным пространством. Заметим, что оно похоже на $\mathbb{R}^3$, в том смысле, что предыдущие равенства можно записать в виде

$$\begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 7 \\ -10 \\ 2 \end{pmatrix} = \begin{pmatrix} 10 \\ -8 \\ 3 \end{pmatrix},$$

В дальнейшем мы еще вернемся к этому вопросу.

Предположим, что *задано* векторное пространство V; например, это может быть $\mathbb{R}^1$, $\mathbb{R}^2$ или $\mathbb{R}^3$. *Аффинным пространством*, ассоциированным с V, мы называем множество A, состоящее из точек P, Q, и т. д., при этом операция "+" означает, что каждой точке $P \in A$ и каждому вектору $\mathbf{v} \in V$ в аффинном пространстве A соответствует другая точка, обозначаемая P "+" $\mathbf{v}$. При этом должны выполняться следующие аксиомы.

Ассоциативность: $(P \text{ "+" } \mathbf{u}) \text{ "+" } \mathbf{v} = P \text{ "+" } (\mathbf{u} + \mathbf{v})$ для $P \in A$ и $\mathbf{u}, \mathbf{v} \in V$.

«Нуль» действует как тождественное отображение:
$P \text{ "+" } \mathbf{0} = P$ для любой точки $P \in A$.

Транзитивность: для заданных P и $Q \in A$ существует $\mathbf{v} \in V$ такой, что $P \text{ "+" } \mathbf{v} = Q$.

Точность: если для любой точки P выполняется $P \text{ "+" } \mathbf{u} = P \text{ "+" } \mathbf{v}$, то $\mathbf{u} = \mathbf{v}$.

Комбинируя две последних аксиомы, можно сказать, что для любых заданных точек P и Q существует единственный вектор

$\mathbf{v}$ такой, что P "+" $\mathbf{v} = Q$. Иногда это удобно записывать как $\mathbf{v} = Q$ "−" P.

Концепция векторного пространства и ассоциированного с ним аффинного пространства привлекает внимание физиков на протяжении трех столетий. Началось это еще с ньютоновской механики, затем продолжалось в специальной теории относительности и в квантовой механике. Цель этой главы — исследуя на интуитивном уровне простой случай двумерного[2] векторного пространства $\mathbb{R}^2$, продемонстрировать ключевые идеи, необходимые для изучения таких структур. Начнем, однако, с одномерного случая, концепция которого настолько очевидна, что подробное обсуждение может показаться слишком педантичным. Однако стоит постараться.

Векторное пространство V называется *одномерным*, если оно удовлетворяет следующим двум условиям: (i) в нем существует вектор $\mathbf{v} \neq \mathbf{0}$; и (ii) если $\mathbf{v} \neq \mathbf{0}$, то любой вектор $\mathbf{u} \in V$ может быть записан как $\mathbf{u} = r\mathbf{v}$, где r — некоторое действительное число. Заметим, что число r единственное: если $r_1\mathbf{v} = r_2\mathbf{v}$, то обязательно $r_1 = r_2$. Действительно, равенство $r_1\mathbf{v} = r_2\mathbf{v}$ можно переписать в виде

$$(r_1 - r_2)\mathbf{v} = \mathbf{0}.$$

Если $r_1 - r_2 \neq 0$, тогда, вводя обозначение $s = (r_1 - r_2)^{-1}$, имеем

$$\mathbf{0} = s[(r_1 - r_2)\mathbf{v}] = (s(r_1 - r_2))\mathbf{v} = 1\mathbf{v} = \mathbf{v},$$

т. е. $\mathbf{v} = \mathbf{0}$, что противоречит начальному предположению $\mathbf{v} \neq \mathbf{0}$. (Проверьте самостоятельно, какую из аксиом векторного пространства мы использовали на каждом этапе рассуждений.) Итак, если выбрать в одномерном пространстве вектор $\mathbf{v} \neq \mathbf{0}$, то для каждого вектора $\mathbf{u}$ можно указать вещественное число r,

$$\mathbf{u} \to r \quad \text{такое, что} \quad \mathbf{u} = r\mathbf{v}.$$

Если $\mathbf{u}_1 = r_1\mathbf{v}$ и $\mathbf{u}_2 = r_2\mathbf{v}$, то $\mathbf{u}_1 + \mathbf{u}_2 = (r_1 + r_2)\mathbf{v}$. Таким образом, вектору $\mathbf{u}_1 + \mathbf{u}_2$ соответствует число $r_1 + r_2$. Аналогично,

---

[2]Мы дадим строгое определение термина «двумерный» в разделе 1.12, термина «одномерный» на следующей странице, а общее понятие размерности векторного пространства — в главе 10. — *Прим. ред.*

если $\mathbf{u} = r\mathbf{v}$ и t — любое действительное число, то $t\mathbf{u} = (tr)\mathbf{v}$, т. е. вектору $t\mathbf{u}$ соответствует число tr. Итак, каждый вектор соответствует вещественному числу, и векторные операции соответствуют операциям в пространстве $\mathbb{R}^1$. В этом случае мы говорим, что имеется *изоморфизм* одномерного векторного пространства V с пространством $\mathbb{R}^1$. Такое сопоставление пространства V пространству $\mathbb{R}^1$ зависит от выбора вектора $\mathbf{v}$. Выбор вектора $\mathbf{v}$ называется выбором *базиса* V, а число r, связанное с $\mathbf{u}$ соотношением $\mathbf{u} = r\mathbf{v}$, называется координатой вектора $\mathbf{u}$ относительно базиса $\mathbf{v}$. Предположим, что мы выбрали другой базис $\mathbf{v}'$. Здесь $\mathbf{v}' = a\mathbf{v}$, где a — действительное число, не равное нулю. Если $\mathbf{u} = r\mathbf{v}$, то

$$\mathbf{u} = (ra^{-1})a\mathbf{v},$$

поэтому

$$\mathbf{u} = r'\mathbf{v}', \qquad r' = a^{-1}r.$$

Таким образом, изменение базиса посредством замены $\mathbf{v}$ на $a\mathbf{v}$ вызывает замену координаты любого вектора r на $a^{-1}r$. Выбор базиса в одномерном пространстве очень похож на выбор системы единиц для физической величины. Если один килограмм (единицу массы) мы заменим на новую единицу массы — один грамм, то вес тела, равный 1.3 кг, станет равным 1300 г. Отличие состоит в том, что для многих обычных физических величин измерение объектов производится с помощью положительных (или нулевых) значений. Бессмысленно говорить, что тело имеет отрицательный объем, массу и т. д. Есть исключение только в теории электричества, где электрический заряд может быть положительным и отрицательным. Например, мы можем представить себе ситуацию, когда в качестве единицы измерения выбран заряд электрона. В таком базисе электрон будет иметь заряд $+1$, а не $-1.602191 \times 10^{-19}$ Кл, где кулон (Кл) — стандартная единица, т. е. базис, принятый международным соглашением.

Пусть A — аффинное пространство, ассоциированное с одномерным векторным пространством V. Если мы выберем точку O в A, то каждая точка P определяет вектор $\mathbf{u} = P$ "$-$" O. Если в пространстве V выберем базис $\mathbf{v}$, то каждой точке P будет соответствовать число $x(P)$,

$$P = O \text{ “+” } x(P)\mathbf{v}.$$

Значение $x(P)$ назовем координатой точки P. Теперь у нас есть две возможности: можно взять точку O в качестве начала координат, и тогда считать другие точки векторами; либо выбрать базис в пространстве V, и тогда отождествить векторы с числами. Если мы заменим базис $\mathbf{v}$ на $a\mathbf{v}$, то x перейдет в x' согласно формуле

$$x'(P) = a^{-1}x(P).$$

Если, кроме того, мы заменим начало O на O', где $O' = O$ "+" $\mathbf{w}$, то получим

$$P \text{ “}-\text{” } O' = (P \text{ “}-\text{” } O) - \mathbf{w}.$$

Если $\mathbf{w} = b\mathbf{v}$, то это равносильно замене x' на x'' по формуле

$$x''(P) = a^{-1}x(P) - b.$$

Сравним вышеприведенные рассуждения с концепцией абсолютного времени, введенной Ньютоном. Ньютон писал:

> Абсолютное, истинное, математическое время или длительность течет равномерно в соответствии со своей внутренней природой, независимо от каких-либо внешних условий. Относительное, кажущееся, обычное время есть некоторая мера длительности, измеренная посредством движения (например, движения стрелки часов). Оно обычно используется вместо истинного времени.

В нашей терминологии мысль Ньютона можно изложить так: есть понятие абсолютного времени, и множество всех абсолютных времен обладает структурой одномерного аффинного пространства. Идея гладкого и равномерного течения времени делает математически точным утверждение, что для множества абсолютных времен в одномерном векторном пространстве V существует операция "+". Именно эта постулированная операция позволяет нам сравнивать различные интервалы времени. Ньютон различает истинное и обычное время. В нашей дискуссии это соответствует степени произвольности при определении координат на аффинной прямой.

Давайте сделаем небольшую паузу и еще раз обдумаем этот абстрактный постулат Ньютона, который является краеугольным камнем в физике двух последних столетий. Каждый из нас имеет собственное психологическое восприятие времени. Наше психологическое время во многих отношениях отличается от абсолютного времени Ньютона. Основное отличие в том, что для *нас время имеет определенное направление*. Будущее в какой-то степени неизвестно и подвержено нашим желаниям и нашему вмешательству. Во многих европейских языках, например, будущее время указывается словами, выражающими желание или необходимость (по-английски I will go — я хочу идти или по-французски j'irai — я должен идти). А вот прошлое в какой-то степени мы знаем и помним. Однако законы Ньютона не зависят от изменения направления времени. Если бы мы прокрутили кинофильм ньютоновской планетарной системы (во всех отношениях реальной) в обратном направлении, то не обнаружили бы противоречий с законами Ньютона. Второе существенное отличие в том, что психологическое время не течет гладко и равномерно, по крайней мере так, как это происходит с ньютоновским абсолютным временем. У нас есть определенные физические функции, повторяющиеся периодически, которые побуждают нас ввести понятие временнóго интервала: через определенный промежуток времени после еды мы опять хотим есть. Но величина такого промежутка сильно меняется, она определяется уровнем сахара в крови, который, в свою очередь, зависит от того, что мы съели, что делали, от нашего психологического состояния в целом и т. д. Кроме того, сильно меняется психологическое восприятие временных интервалов. Время летит быстро, когда нам интересно, и мы увлечены тем, что делаем, и наоборот, оно тянется очень медленно, если нам скучно. И тем не менее, оказывается, что наш внутренний ритм как-то коррелирует с периодичностью мира вокруг нас. Из самых древних источников мы знаем, что измерение внешнего времени для ведения календаря или для научных целей всегда основывалось на вращении небесных тел. В качестве стандарта временнóго интервала суток обычно использовали видимое вращение Солнца, т. е. интервалы между последовательными прохождениями Солнцем зенита. Египтяне разделили сутки на 24 равных промежутка (24 часа), а греки делили периоды от восхода до заката солнца на двенадцать одинаковых часов, аналогично делили ночные периоды. Эти временны́е интервалы отмечались с помощью разных приборов, например, с помощью водяных часов, а днем с помощью солнечных часов. Те, кто принял греческую систему, должны были оборудовать водяные часы специальным компенсирующим приспособлением, чтобы модифицировать временны́е интервалы в соответствии со временем года. (Длительность календарного, или солнечного, дня в течение года непостоянна из-за неравномерности движения Солнца на небесной сфере. Простейшая, относительно точная мера времени — это звездный день. Он определяется по вращению Земли вокруг своей оси и измеряется наблюдением за какой-то фиксированной звездой: период между двумя последовательными прохождениями фиксированной звезды определенного меридиана называют звездным днем. Календарный день в среднем на четыре минуты длиннее звездного дня.)

Самые первые часы, по-видимому, появились в Европе в тринадцатом веке, но они были очень неточными. Первый крупный шаг в деле усовершенствования часов был сделан в XVII веке, когда Галилей обнаружил, что период колебаний маятника (измеренный, например, по нормальному сердечному пульсу) постоянен. Но Галилей, похоже, мало использовал это открытие на практике и только изобрел небольшой прибор для врачей, чтобы измерять пульс у пациентов. Однако говорят, что его сын использовал маятник для создания часов. Во всяком случае, с этого времени усовершенствование механических часов происходило стремительно. Таким образом, во времена Ньютона уже существовал метод, согласно которому, в принципе, можно было делить время на произвольно малые равные интервалы. Следует отметить, что направление движения стрелок механических часов совершенно условно, и оно легко обратимо. Простой заменой зубчатой передачи мы можем заставить стрелки вращаться в противоположном направлении. Интересно пофантазировать на тему: как развитие механических часов связано с ньютоновской концепцией времени?

1.3. Функции и аффинные функции

В нескольких следующих параграфах мы будем рассматривать те отображения $A\mathbb{R}^2$ на себя, при которых прямые линии преобразуются в прямые линии. Начнем с общего обсуждения понятий «отображения» или «функции».

Пусть W и X — некоторые множества. Правило $f : W \to X$, согласно которому элементу w множества W соответствует элемент $f(w) \in X$, называется *функцией* (или *отображением*, *картой*, *оператором*), переводящей множество W в множество X. Множество W называется *областью определения* f. Если A — подмножество W, то $f(A)$ обозначает подмножество X, состоящее из элементов $f(w)$, где $w \in A$:

$$f(A) = \{f(w) \mid w \in A\}.$$

Множество $f(W)$ называется *образом* f, оно является подмножеством X.

Например, допустим, что f — это отображение $\mathbb{R}^2$ на себя, определяемое равенством $f(P) = P$ "+" $\mathbf{v}$, где $\mathbf{v}$ — фиксированный вектор. Тогда $f(A)$ получается из A переносом A на вектор $\mathbf{v}$. Если $A = l = \{P + t\mathbf{u}\}$ — прямая, то $f(l) = \{P + \mathbf{v} + t\mathbf{u}\}$ — другая прямая (рис. 1.7).

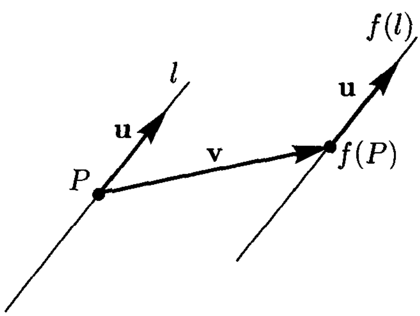


Рис. 1.7

Это очень общее и мощное определение функции. Уточним, что считать результатом действия функции. Например, невозможно иметь функцию $f : \mathbb{R} \to \mathbb{R}$, обладающую свойствами:

$f(1) = 2$ и $f(1) = 3$, но все будет в порядке, если для функции $f : \mathbb{R} \to \mathbb{R}^2$ мы будем иметь $f(1) = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$.

Следует знать стандартную терминологию, касающуюся области определения и образа функции f.

1. Если два различных элемента $w_1, w_2 \in W$ всегда отображаются в различные точки $x_1, x_2 \in X$, то f называют *инъективным* отображением. Другими словами, f инъективно, если из равенства $f(w_1) = f(w_2)$ следует, что $w_1 = w_2$.

2. Если образ $f(W)$ составляет все множество X, то f называется *сюръективным* (иногда говорят, что f является отображением на X). Другими словами, если уравнение $f(w) = x$ имеет *по крайней мере одно* решение для каждого $x \in X$, то f сюръективно.

3. Если отображение f и инъективно, и сюръективно, то оно называется *биективным.* Другими словами, f называется биективным, если уравнение $f(w) = x$ имеет *единственное* решение w для каждого элемента $x \in X$. В этом случае существует *обратная* функция $f^{-1} : X \to W$, которая отображает каждый элемент $x \in X$ в единственный w, для которого $f(w) = x$.

Эти свойства отображений проиллюстрированы на рис. 1.8.

Во многих случаях мы можем описать функции с помощью формул. Существует два способа записи соответствующих формул. Например, операцию возведения в квадрат $F : \mathbb{R} \to \mathbb{R}$ мы можем записывать как $F(x) = x^2$ или

$$F : x \mapsto x^2.$$

При любой форме записи x — немой символ. Та же функция может записываться как

$$F(t) = t^2$$

или

$$F : \alpha \mapsto \alpha^2.$$

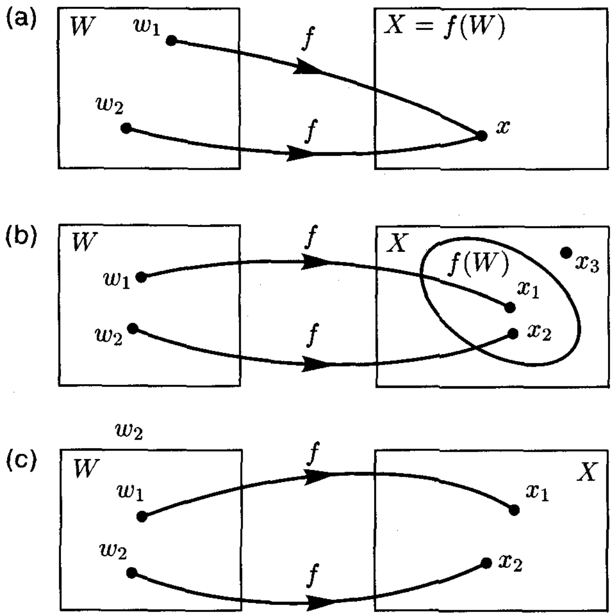


Рис. 1.8. (a) Отображение сюръективно, но не инъективно. Невозможно определить $f^{-1}(x)$. (b) Отображение инъективно, но не сюръективно. Нет определения для $f^{-1}(x_3)$. (c) Отображение биективно. $f^{-1}(x_1) = w_1$ и $f^{-1}(x_2) = w_2$.

Формула для описания функции может содержать несколько численных аргументов, например,

$$G(x, y) = 2x + 3y \quad \text{или} \quad G : (x, y) \mapsto 2x + 3y.$$

Такая функция G упорядоченную пару чисел (x, y) переводит в $2x + 3y$.

Определим еще одно понятие, относящееся к функциям — *композицию*. Пусть W, X, Y, Z — некоторые множества, и допустим, что у нас есть функции

$$\begin{aligned} f &: W \to X, \\ g &: X \to Y, \\ h &: Y \to Z. \end{aligned}$$

Рассмотрим функцию, которая берет $w \in W$, действует на этот элемент посредством f, в результате чего получается элемент $f(w)$ множества X; затем на этот элемент $f(w)$ действует g, производя элемент множества Y. Такую функцию обозначим $g \circ f$ и назовем ее *композицией* g с f. Короче: $(g \circ f)(w) = g(f(w))$. Заметим, что композиция

$$h \circ g \circ f : W \to Z$$

(сначала действует f, потом g, потом h) совпадает с отображениями $h \circ (g \circ f)$ и $(h \circ g) \circ f$. Таким образом, для композиций выполняется такой же «ассоциативный закон», как и для умножения действительных чисел.

Теперь рассмотрим функции, определенные на аффинных прямых, плоскостях и векторных пространствах. Начнем с одномерных примеров, которые, конечно, очень важны, но при этом достаточно просты для изучения.

Пусть $\mathbb{A}$ — аффинная прямая, изображенная на рис. 1.9. Мы можем выбрать некоторое начало и направление и приписать координату каждой точке на этой прямой.

Рис. 1.9

На математическом языке это значит, что мы выбрали начало координат O на прямой $\mathbb{A}$ и базис $\mathbf{v}$ в пространстве V, как это сделано в предыдущем параграфе. Таким образом, мы построили аффинную координатную функцию

$$x : \mathbb{A} \to \mathbb{R}.$$

Конечно, на прямой существует много возможных аффинных координатных функций, которые можно построить в зависимости от выбора начала координат и единицы измерения. Мы называем x *координатной* функцией, потому что она обратима: зная $x(P)$, можно восстановить P. Заметим, что x сохраняет «интерполяционное свойство» действительной аффинной прямой: если

$$R = (1 - t)P + tQ,$$

то
$$x(R) = (1-t)x(P) + tx(Q).$$
В частности, если R — средняя точка отрезка PQ, то
$$x(R) = \frac{1}{2}x(P) + \frac{1}{2}x(Q).$$

Вы, вероятно, никогда раньше не думали об x как о *функции*. Вы не можете написать для нее формулу. Однако без этой функции невозможно изучать физику, потому что именно она позволяет вам записывать другие функции на прямой в виде формул. Например, сила, действующая на частицу, которая находится на прямой, является функцией положения частицы
$$f : \mathbb{A} \to \mathbb{R};$$
причем вы не можете написать формулу для f, но можете ввести аффинную координату
$$x : \mathbb{A} \to \mathbb{R}$$
и функцию $F : \mathbb{R} \to \mathbb{R}$, после чего написать
$$f(P) = F(x(P)) = (F \circ x)(P).$$
Именно такая композиция, например, формула «сила $= \sin x$», и применяется для представления функции на прямой.

Время — это тоже аффинная прямая, точками которой являются «мгновения». Аффинная координатная функция $t : \mathbb{A} \to \mathbb{R}$ приписывает каждому мгновению число. Чтобы узнать t, мы смотрим на часы. Часы, отсчитывающие время с разной скоростью, дают различные функции t, но любые «хорошие» часы производят аффинную функцию. В то же время плохие часы, например, часы, у которых длина маятника меняется с температурой, произведут неаффинную функцию.

Движение частицы по прямой линии определяет зависимость одной действительной аффинной прямой $\mathbb{A}_t$ (*время*) от другой действительной аффинной прямой $\mathbb{A}_x$ (*пространство*). Такая функция $f : \mathbb{A}_t \to \mathbb{A}_x$ каждому моменту времени E ставит в соответствие точку P на прямой, так что $P = f(E)$. Мы не можем записать формулу для f, потому что E и P не являются числами. Поэтому, если мы хотим иметь формулу для описания движения частицы, то должны ввести аффинные координатные функции t и x. После этого можно написать

$$x(P) = F(t(E)),$$

где $F : R \to R$ может быть представлена формулой типа, например, $F(\alpha) = x_0 + v_0\alpha + \frac{1}{2}a\alpha^2$.

1.4. Евклидовы и аффинные преобразования

Отображение $f : \mathbb{R}^2 \to \mathbb{R}^2$ называется *евклидовым* преобразованием, если оно сохраняет расстояние. Это значит, что расстояние между двумя любыми точками $P_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ и $P_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$ равняется расстоянию между $f(P_1)$ и $f(P_2)$. Если мы выразим f через две функции $\phi : \mathbb{R}^2 \to \mathbb{R}$ и $\psi : \mathbb{R}^2 \to \mathbb{R}$ так, что

$$f\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \phi(x,y) \\ \psi(x,y) \end{bmatrix},$$

то из условия сохранения расстояния получим равенство

$$[\phi(x_2,y_2)-\phi(x_1,y_1)]^2+[\psi(x_2,y_2)-\psi(x_1,y_1)]^2 = (x_2-x_1)^2+(y_2-y_1)^2,$$

которое должно выполняться для всех значений x_1, y_1, x_2, y_2.

Можно считать, что евклидова геометрия изучает свойства тех подмножеств плоскости, которые инвариантны при любом евклидовом преобразовании. Например, если A — окружность и f — евклидово преобразование, то $f(A)$ — тоже окружность. Если l — прямая, то $f(l)$ — тоже прямая. Из определения ясно, что если f и g — евклидовы преобразования, то композиция $g \circ f$ тоже будет евклидовым преобразованием.

Отображение $f : \mathbb{R}^2 \to \mathbb{R}^2$ называется *аффинным* преобразованием, если оно переводит прямые в прямые. Таким образом, для любой прямой l образ $f(l)$ — тоже прямая. Например, пусть преобразование f задается как

$$f\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2x + y + 1 \\ y - x + 5 \end{bmatrix}.$$

Самое общее выражение для прямой на плоскости записывается уравнением

$$ax + by + c = 0.$$

Это значит, что

$$l = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| \, ax + by + c = 0 \right\}.$$

Поэтому

$$f(l) = \left\{ \begin{bmatrix} w \\ z \end{bmatrix} \middle| \, w = 2x + y + 1, \quad z = y - x + 5 \ \text{ и } \ ax + by + c = 0 \right\}.$$

Мы можем решить уравнения

$$\begin{aligned} w &= 2x + y + 1, \\ z &= y - x + 5 \end{aligned}$$

для x и y, выразив их через w и z:

$$\begin{aligned} x &= \frac{1}{3}(w - 1) - \frac{1}{3}(z - 5), \\ y &= \frac{1}{3}(w - 1) + \frac{2}{3}(z - 5). \end{aligned}$$

Тогда условие

$$ax + by + c = 0$$

можно переписать в виде

$$a\left[\frac{1}{3}(w - 1) - \frac{1}{3}(z - 5)\right] + b\left[\frac{1}{3}(w - 1) + \frac{2}{3}(z - 5)\right] + c = 0,$$

или иначе:

$$\frac{1}{3}(a + b)w + \left(\frac{2}{3}b - \frac{1}{3}a\right)z + c + \frac{4}{3}a - \frac{11}{3}b = 0.$$

Другими словами,

$$f(A) = \left\{ \begin{bmatrix} w \\ z \end{bmatrix} \middle| \, ew + gz + h = 0 \right\},$$

где

$$e=\frac{1}{3}(a+b),\quad g=\frac{2}{3}b-\frac{1}{3}a,\quad \text{и}\quad h=c+\frac{4}{3}a-\frac{11}{3}b.$$

Мы видим, что в результате отображения f получилась снова прямая.

Заметим, что f из этого примера не является евклидовым преобразованием. Аффинная геометрия изучает те свойства подмножеств плоскости, которые инвариантны при всех взаимно-однозначных аффинных преобразованиях. Таким образом, если A — окружность, то $f(A)$ не обязана быть окружностью (позднее мы увидим, что это всегда эллипс).

Предположим, что f и есть такое аффинное преобразование. Тогда, по определению, f преобразует прямые в прямые и поскольку различные точки переходят в различные точки, то параллельные прямые переходят в параллельные прямые. Поэтому преобразование f переводит параллелограммы в параллелограммы. Таким образом, в аффинной геометрии существует понятие параллелограмма — рис. 1.10, но нет понятий прямоугольника и квадрата — рис. 1.11.

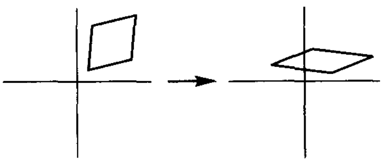


Рис. 1.10

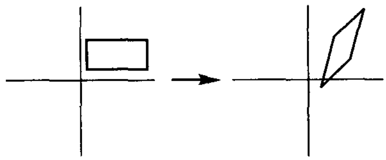


Рис. 1.11

1.5. Линейные преобразования

Простейшим случаем аффинных (и евклидовых) преобразований являются *трансляции*

$$\begin{bmatrix} x \\ y \end{bmatrix} \to \begin{bmatrix} x+a \\ y+b \end{bmatrix}.$$

С помощью трансляций любую точку пространства можно переместить в любую другую. Прежде чем переходить к дальнейшему, давайте ограничимся аффинными преобразованиями, при которых одна точка, скажем $\begin{bmatrix}0\\0\end{bmatrix}$, неподвижна, т. е. $\begin{bmatrix}0\\0\end{bmatrix} \mapsto \begin{bmatrix}0\\0\end{bmatrix}$. Производя трансляцию, мы сможем из этой точки перейти в любую другую.

Пусть f — взаимно-однозначное аффинное преобразование, при котором начальная точка $O = \begin{bmatrix}0\\0\end{bmatrix}$ неподвижна. Теперь любую точку $P = \begin{bmatrix}x\\y\end{bmatrix}$ можно задавать с помощью *вектора положения* $\mathbf{v} = \begin{pmatrix}x\\y\end{pmatrix}$, так что $P = O$ "+" $\mathbf{v}$. В дальнейшем мы будем опускать обозначение [ ], тем самым не будем различать $A\mathbb{R}^2$ и $\mathbb{R}^2$. Из того, что f преобразует параллелограмм в параллелограмм, немедленно следует, что если вектор положения точки Q есть $\mathbf{v} + \mathbf{w}$, то вектор положения $f(Q)$ будет $f(\mathbf{v}) + f(\mathbf{w})$. Поэтому

$$f(\mathbf{v} + \mathbf{w}) = f(\mathbf{v}) + f(\mathbf{w}) \quad \text{при} \quad \mathbf{v} \neq \mathbf{w}. \tag{1.2}$$

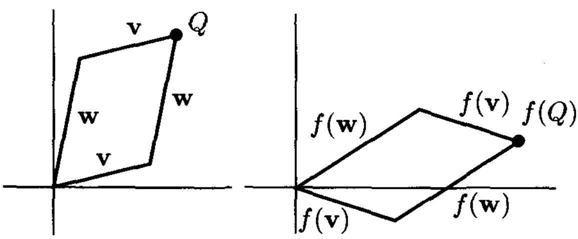


Рис. 1.12

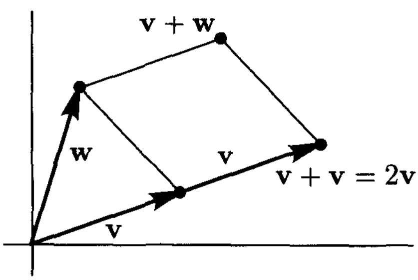


Рис. 1.13

Теперь можно показать, что преобразование f сохраняет отношение длин отрезков на любой прямой. Если параллелограмм натянут на векторы $\mathbf{w}$, $\mathbf{v}$, $\mathbf{v}+\mathbf{w}$ и $2\mathbf{v}$, то из рис. 1.12, 1.13 видно, что

$$f(2\mathbf{v}) = 2f(\mathbf{v}),$$

и уравнение (1.2) выполняется при $\mathbf{v} = \mathbf{w}$. Повторяя те же аргументы, получаем

$$f(n\mathbf{v}) = nf(\mathbf{v})$$

для любого целого числа $n \geq 0$. Проводя такое же рассуждение для вектора $(1/m)\mathbf{v}$, в результате получим

$$f(a\mathbf{v}) = af(\mathbf{v})$$

для любого рационального числа $a \geq 0$.

Из рассмотрения параллелограмма с вершинами O, $\mathbf{w}$, $-\mathbf{v}+\mathbf{w}$, $-\mathbf{v}$ видим, что

$$f(-\mathbf{v}) = -f(\mathbf{v}),$$

поэтому

$$f(a\mathbf{v}) = af(\mathbf{v})$$

для любого рационального числа a, положительного и отрицательного, и для всех $\mathbf{v}$.

Если мы *предположим*, что f — непрерывное преобразование, то получим, что

$$f(a\mathbf{v}) = af(\mathbf{v})$$

для всех вещественных чисел a. Но оказывается, такое предположение излишне. Действительно, из свойств системы вещественных чисел и из того, что f преобразует прямую на плоскости в другую прямую, следует, что f — непрерывное преобразование. Это значит, что равенство $f(a\mathbf{v}) = af(\mathbf{v})$ справедливо для всех вещественных чисел a и всех векторов $\mathbf{v}$. Доказательство этого утверждения несколько искусственно, мы приведем его в добавлении в конце этой главы. А сейчас ограничимся только теми аффинными преобразованиями, для которых условие $f(a\mathbf{v}) = af(\mathbf{v})$ выполняется при всех вещественных числах a, хотя, как уже было сказано, здесь нет никакого ограничения. Для таких преобразований f имеем тождество

$$f(a\mathbf{v} + b\mathbf{w}) = af(\mathbf{v}) + bf(\mathbf{w}), \tag{1.3}$$

справедливое для любых вещественных чисел a, b и для любых векторов $\mathbf{v}$, $\mathbf{w}$ в пространстве $\mathbb{R}^2$.

Отображение $f : \mathbb{R}^2 \to \mathbb{R}^2$, удовлетворяющее условию (1.3), называется *линейным* преобразованием плоскости. Таким образом, изучение *аффинных* преобразований $\mathbb{R}^2$, при которых фиксировано начало координат, мы свели к изучению линейных преобразований векторного пространства $\mathbb{R}^2$.

По определению, любое отображение $\mathbb{R}^2 \to \mathbb{R}^2$, удовлетворяющее условию (1.3), линейно. Но не каждое линейное преобразование является взаимно-однозначным. Например, преобразование, которое отображает каждый вектор пространства $\mathbb{R}^2$ в нулевой вектор:

$$f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{для всех} \quad \begin{pmatrix} x \\ y \end{pmatrix},$$

линейно, но не взаимно-однозначно.

Если f — линейное преобразование, то

$$f(\mathbf{v} + t\mathbf{w}) = f(\mathbf{v}) + tf(\mathbf{w}).$$

Если f, кроме того, взаимно-однозначно, то из условия $\mathbf{w} \neq 0$ следует, что $f(\mathbf{w}) \neq 0$. Таким образом, f преобразует прямую $\{\mathbf{v} + t\mathbf{w} \mid t \in \mathbb{R}\}$ в прямую $\{f(\mathbf{v}) + tf(\mathbf{w}) \mid t \in \mathbb{R}\}$, т. е. f всегда прямые преобразует в прямые. Следовательно, любое взаимно-однозначное линейное преобразование является аффинным преобразованием. *Взаимно-однозначное* линейное преобразование называется *регулярным* или *несингулярным*. Линейное преобразование, *не* являющееся взаимно-однозначным, называется *сингулярным*. Мы уже видели, что любое регулярное линейное преобразование является аффинным. Покажем, что сингулярные преобразования переводят всю плоскость в прямую или в начальную точку.

Очевидно, что если f и g — линейные преобразования (регулярные или нет), то $g \circ f$ — тоже линейное преобразование.

Действительно, поскольку f линейно, то $(g \circ f)(a\mathbf{v} + b\mathbf{w}) = g(af(\mathbf{v}) + bf(\mathbf{w}))$, и поскольку g линейно, это выражение равно $ag \circ f(\mathbf{v}) + bg \circ f(\mathbf{w})$, из чего следует линейность $g \circ f$.

Итак, по определению, преобразование f линейно, если условие (1.3) выполняется для всех пар векторов $\mathbf{v}$ и $\mathbf{w}$ и для всех вещественных чисел a и b. Аффинное преобразование — это взаимно-однозначное отображение пространства $\mathbb{R}^2$ в себя, переводящее прямые в прямые. Любое аффинное преобразование может быть записано как регулярное линейное преобразование с последующей трансляцией, т. е. для любого аффинного преобразования f выполняется

$$f(\mathbf{w}) = l(\mathbf{w}) + \mathbf{v},$$

где l — регулярное линейное преобразование. Верно и обратное утверждение: если преобразование f имеет такой вид, то оно аффинно.

1.6. Матрица линейного преобразования

Пусть $f : \mathbb{R}^2 \to \mathbb{R}^2$ *— линейное преобразование.* На плоскости каждый вектор $\begin{pmatrix} x \\ y \end{pmatrix}$ можно записать в виде

$$\begin{pmatrix} x \\ y \end{pmatrix} = x\begin{pmatrix} 1 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

так что

$$f\begin{pmatrix} x \\ y \end{pmatrix} = xf\begin{pmatrix} 1 \\ 0 \end{pmatrix} + yf\begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Последняя формула показывает, что преобразование f полностью определяется тем, во что оно переводит базисные векторы $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ и $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Предположим, что $f\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix}$ и $f\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix}$. Тогда f полностью определяется четырьмя числами a, b, c и d

$$f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}.$$

Запишем эти числа в форме квадратной *матрицы*

$$\mathrm{Mat}(f) = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

где первый столбец есть образ вектора $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, а второй — образ вектора $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Тогда образ любой точки $\begin{pmatrix} x \\ y \end{pmatrix}$ задается формулой

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}, \tag{1.4}$$

причем соотношение (1.4) рассматривается как определение умножения матриц. Оно показывает, как вектор $\begin{pmatrix} x \\ y \end{pmatrix}$ умножать на матрицу $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, чтобы получить другой вектор. Согласно этому правилу для получения каждой компоненты вектора надо «умножить строку на столбец». Подробнее, верхняя компонента $ax+by$ вектора справа в (1.4) получена умножением верхней строки матрицы (a, b) на столбец $\begin{pmatrix} x \\ y \end{pmatrix}$, аналогично получена нижняя компонента.

Например, пусть R_θ есть поворот на плоскости против часовой стрелки на угол θ. Тогда (рис. 1.14)

$$R_\theta \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$$

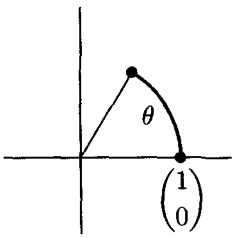


Рис. 1.14

и

$$R_\theta \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix},$$

а поворот R_θ имеет матрицу

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Образ любой точки $\begin{pmatrix} x \\ y \end{pmatrix}$ имеет вид

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{pmatrix}.$$

Формула (1.4) показывает соответствие между линейными преобразованиями и матрицами, т. е. матрицы 2×2 можно отождествить с линейными преобразованиями пространства $\mathbb{R}^2$.

1.7. Умножение матриц

Допустим, что F — линейное преобразование, которому соответствует матрица $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, а линейному преобразованию G соответствует матрица $\begin{pmatrix} e & f \\ g & h \end{pmatrix}$. Мы знаем, что преобразование $F \circ G$ тоже линейно и ему соответствует матрица, в которой первый столбец равен

$$F \circ G \begin{pmatrix} 1 \\ 0 \end{pmatrix} = F \begin{pmatrix} e \\ g \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e \\ g \end{pmatrix} = \begin{pmatrix} ae + bg \\ ce + dg \end{pmatrix},$$

а второй столбец равен

$$F \circ G \begin{pmatrix} 0 \\ 1 \end{pmatrix} = F \begin{pmatrix} f \\ h \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} f \\ h \end{pmatrix} = \begin{pmatrix} af + bh \\ cf + dh \end{pmatrix}.$$

Мы определим умножение матриц так, чтобы их произведение соответствовало композиции линейных преобразований: $(\mathrm{Mat} F) \times (\mathrm{Mat}\, G) = \mathrm{Mat}\, F \circ G$. Итак, правило умножения матриц таково:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix}.$$

Каждый элемент матрицы-произведения получается умножением той или иной строки первой матрицы на некоторый столбец второй.

Например, если R_θ есть поворот против часовой стрелки на угол θ, а R_ϕ есть аналогичный поворот на угол ϕ, то $R_\theta \cdot R_\phi = R_{\theta+\phi}$ и

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \times \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$
$$= \begin{pmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi & -\cos\theta\sin\phi - \sin\theta\cos\phi \\ \sin\theta\cos\phi + \cos\theta\sin\phi & \cos\theta\cos\phi - \sin\theta\sin\phi \end{pmatrix}.$$

Сравнение этого выражения с матрицей $R_{\theta+\phi}$

$$\begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix}$$

дает стандартные тригонометрические формулы для $\cos(\theta+\phi)$ и $\sin(\theta+\phi)$. Таким образом, уже нет необходимости помнить формулы для синуса и косинуса суммы углов. Вы можете получить их из более общего правила умножения матриц.

Заметим, что, вообще говоря, умножение матриц *не* коммутативно. Например,

$$\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 4 & 10 \\ 0 & 15 \end{pmatrix},$$

а произведение этих же матриц в обратном порядке дает другой результат:

$$\begin{pmatrix} 4 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 4 & 8 \\ 0 & 15 \end{pmatrix}.$$

(Два поворота в пространстве $\mathbb{R}^2$ коммутируют между собой, потому что не важно, на какой угол мы поворачиваем сначала. Но в *общем* случае коммутативности нет.)

В качестве другого примера умножения матриц докажем утверждение о разложении матрицы в произведение трех матриц. Позже это утверждение нам пригодится. Мы докажем, что любая матрица $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ при $a \neq 0$ может быть записана как тройное произведение в виде

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix} \begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}. \tag{1.5}$$

Чтобы это доказать, мы просто предложим способ вычисления констант y, r, s, x. Сначала умножим две матрицы, стоящие в (1.5) справа. Получаем:

$$\begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} r & rx \\ 0 & s \end{pmatrix}.$$

Мы хотим, чтобы выполнялось равенство

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix} \begin{pmatrix} r & rx \\ 0 & s \end{pmatrix} = \begin{pmatrix} r & rx \\ ry & rxy + s \end{pmatrix}.$$

Приравняем соответствующие элементы матриц справа и слева. Сначала получаем $a = r$, и по нашему предположению $a \neq 0$, $r \neq 0$. Далее, имеем $b = rx$ и, следовательно, $x = b/r = b/a$ (напомним, что $a \neq 0$). Аналогично, $c = ry$, поэтому $y = c/r = c/a$. И наконец, $d = rxy + s$, и, следовательно,

$$s = d - rxy = d - r(b/r)(c/r) = d - (bc)/a.$$

Аналогичное разложение матрицы, важное при изучении систем линз, имеет вид

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & y \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & f \\ e & 0 \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}.$$

Это разложение, возможное для любой матрицы при $c \neq 0$, доказывается столь же просто: достаточно выполнить умножение в правой части формулы и приравнять соответствующие элементы в обеих частях равенства.

1.8. Алгебра матриц

Пусть F и G — линейные преобразования в пространстве $\mathbb{R}^2$. Определим их сумму равенством:

$$(F+G)(\mathbf{v}) = F(\mathbf{v}) + G(\mathbf{v}).$$

Заметим, что

$$\begin{aligned}(F+G)(a\mathbf{v}+b\mathbf{w}) &= F(a\mathbf{v}+b\mathbf{w}) + G(a\mathbf{v}+b\mathbf{w})\\ &= aF(\mathbf{v}) + bF(\mathbf{w}) + aG(\mathbf{v}) + bG(\mathbf{w})\\ &= a(F(\mathbf{v}) + G(\mathbf{v})) + b(F(\mathbf{v}) + G(\mathbf{v}))\\ &= a(F+G)(\mathbf{v}) + b(F+G)(\mathbf{w}).\end{aligned}$$

Таким образом, сумма преобразований $F+G$ есть тоже линейное преобразование. Очевидно, что такая операция сложения обладает свойствами ассоциативности, коммутативности, нулевое преобразование имеет вид $O(\mathbf{v}) = \mathbf{0}$ для всех $\mathbf{v}$, условие $(-F)(\mathbf{v}) = -F(\mathbf{v})$ определяет противоположное к F преобразование $(-F)$, т. е. $(-F) + F = O$, где O обозначает нулевое линейное преобразование.

Если H — третье линейное преобразование, то композиция, представляющая произведение матриц, обладает свойством:

$$\begin{aligned}H\circ(F+G)(v) &= H[(F+G)(v)]\\ &= H[F(v)+G(v)]\\ &= H[F(v)] + H[G(v)]\\ &= (H\circ F)(v) + (H\circ G)(v),\end{aligned}$$

справедливым для всех $\mathbf{v}$. Это можно записать кратко

$$H\circ(F+G) = H\circ F + H\circ G.$$

Аналогично,

$$(F+G)\circ H = F\circ H + G\circ H.$$

Это значит, что умножение матриц дистрибутивно по отношению к введенной операции сложения.

Из определения суммы линейных преобразований непосредственно следует, что если имеются матрицы F и G:

$$\mathrm{Mat}(F) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{и} \quad \mathrm{Mat}(G) = \begin{pmatrix} e & f \\ g & h \end{pmatrix},$$

то для матрицы преобразования $F + G$ имеем

$$\mathrm{Mat}(F+G) = \begin{pmatrix} a+e & b+f \\ c+g & d+h \end{pmatrix} = \mathrm{Mat}(F) + \mathrm{Mat}(G).$$

Другими словами, при сложении матриц мы складываем их соответствующие элементы. Матрицу можно умножать на число: например, $2\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 2a & 2b \\ 2c & 2d \end{pmatrix}$. Заметим, что

$$2\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 2a & 2b \\ 2c & 2d \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Итак, мы определили сложение и умножение матриц 2×2; правила сложения и умножения матриц точно такие же, как большинство известных правил сложения и умножения чисел. Таким образом,

- сложение ассоциативно, коммутативно, существуют нулевое и противоположное преобразования;
- умножение матриц ассоциативно, дистрибутивно по отношению к сложению, существует единичная матрица $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.[3]

Однако, у матриц есть два существенных свойства, отличающие их от чисел.

1. Умножение матриц *не* коммутативно.
2. Произведение двух ненулевых матриц может равняться нулю, так что закон сокращения при умножении матриц работает не всегда:

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \times \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

[3]Имеется в виду, что умножение любой матрицы M на матрицу $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ (и слева, и справа) дает M. — *Прим. ред.*

И тем не менее, во многих отношениях с линейными преобразованиями можно работать так же, как и с числами.

Теперь рассмотрим линейные преобразования не в $\mathbb{R}^2$, а в векторном пространстве $\mathbb{R}^3$. Векторы пространства $\mathbb{R}^3$ описываются столбцами $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$. Существуют три базисных вектора $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ и $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Наиболее общее линейное преобразование в пространстве $\mathbb{R}^3$ описывается матрицей 3×3:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

где первый столбец является образом первого базисного вектора $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ и т. д. Формула умножения таких матриц имеет вид

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix},$$

где для любых i, j, принимающих значения 1, 2, 3, элемент результирующей матрицы равен

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j}.$$

Таким образом, как и в $\mathbb{R}^2$, для получения любого элемента произведения надо строку первой матрицы умножать на столбец второй. Например, пусть $i = 2$ и $j = 3$. Тогда написанная выше формула условно изображена на рис. 1.15.

При сложении матриц 3×3 матричный элемент результирующей матрицы равен сумме соответствующих матричных элементов. При этом сохраняются все свойства, которыми обладают матрицы 2×2: ассоциативность и коммутативность для сложения; ассоциативность и дистрибутивность по отношению к сложению — для умножения.

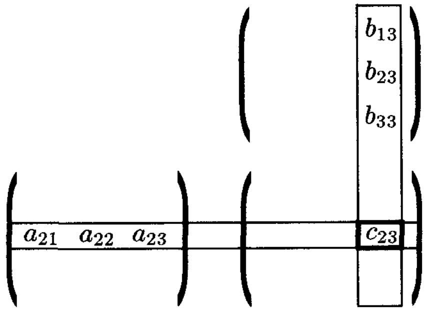


Рис. 1.15

Аналогично, мы можем рассматривать матрицы 4×4, 5×5 и вообще матрицы $n \times n$.

Можно умножить вектор в пространстве $\mathbb{R}^3$ на матрицу 3×3:

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} ax + by + cz \\ dx + ey + fz \\ gx + hy + iz \end{pmatrix},$$

в общем случае можно умножить вектор в пространстве $\mathbb{R}^n$ на матрицу $n \times n$.

1.9. Площади и определители

Опять вернемся к плоскости. Пусть f — несингулярное линейное преобразование плоскости. Поскольку $f(\mathbf{v} + t\mathbf{w}) = f(\mathbf{v}) + tf(\mathbf{w})$, преобразование f прямые переводит в прямые, следовательно, f является аффинным преобразованием и квадраты переводит в параллелограммы. Далее, пусть $\square_{\mathbf{v}}$ — единичный квадрат, левый нижний угол которого находится в $\mathbf{v}$ (рис. 1.16):

$$\begin{aligned} \square_{\mathbf{v}} &= \left\{ \mathbf{v} + s\begin{pmatrix} 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} 0 \\ 1 \end{pmatrix} \middle| \, 0 \le s < 1, \quad 0 \le t < 1 \right\} \\ &= \{\mathbf{v} + \mathbf{w} \mid \mathbf{w} \in \square_{\mathbf{0}}\}. \end{aligned}$$

В результате преобразования f получается образ квадрата $\square_{\mathbf{v}}$, обозначаемый $f(\square_{\mathbf{v}})$, который может быть получен

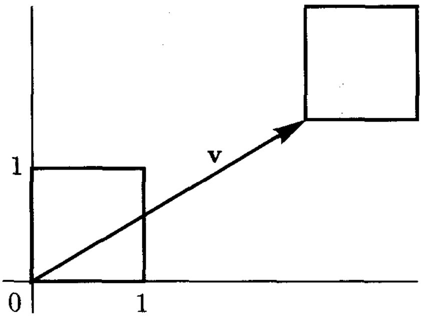


Рис. 1.16

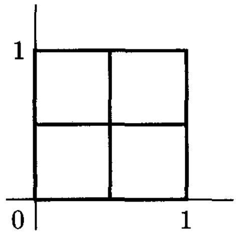


Рис. 1.17

трансляцией образа $f(\square_\mathbf{0})$ квадрата $\square_\mathbf{0}$:

$$\begin{aligned} f(\square_\mathbf{v}) &= \{f(\mathbf{v}) + f(\mathbf{w}) \mid \mathbf{w} \in \square_\mathbf{0}\} \\ &= \{f(\mathbf{v}) + \mathbf{u} \mid \mathbf{u} \in f(\square_\mathbf{0})\}. \end{aligned}$$

Это значит, что $f(\square_\mathbf{v})$ имеет ту же самую площадь, что и $f(\square_\mathbf{0})$. Очевидно, что то же самое справедливо не только для единичного квадрата, но и для квадрата любого размера. С другой стороны, можно разделить единичный квадрат на четыре конгруэнтных квадрата (рис. 1.17). Соответствующие образы тоже будут конгруэнтны и сложенные вместе образуют образ $\square_\mathbf{0}$, т. е. образ каждого из них имеет площадь, равную $\frac{1}{4} \times (f(\square_\mathbf{0}))$ (рис. 1.18).

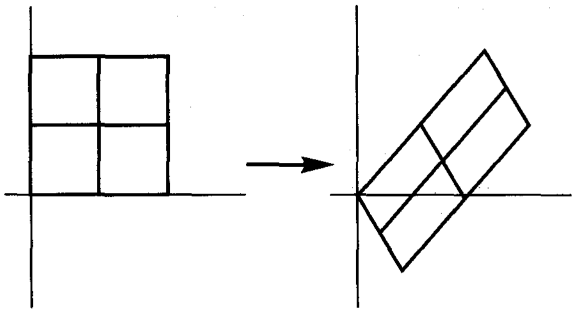


Рис. 1.18

Продолжив деление квадратов дальше, можно сделать вывод, что для квадрата $\square$, длина стороны которого $1/2^k$, отношение

$$\frac{\text{площадь } f(\square)}{\text{площадь } \square}$$

есть число, не зависящее от $\square$ (или от величины 2^k). Обозначим это число $\mathrm{Ar}(f)$, т. е.

$$\mathrm{Ar}(f) = \frac{\text{площадь}\, f(\square)}{\text{площадь}\,\square}.$$

Пусть D — некоторое «хорошее» подмножество на плоскости (рис. 1.19), т. е. подмножество, которое можно с любой степенью точности аппроксимировать объединением описанных выше квадратиков (рис. 1.20). (Строго говоря, следовало бы аппроксимировать D изнутри и снаружи. Если мы можем покрыть границу множества D конечным объединением квадратиков сколь угодно малой суммарной площади, то общую площадь их образов тоже можно сделать сколь угодно малой. Именно такие аппроксимации нам необходимы, и именно их возможность подразумевается нашим термином «хорошее» множество[4]). Для таких множеств также будет выполняться[5] равенство

$$\frac{\text{площадь}\, f(D)}{\text{площадь}\, D} = \mathrm{Ar}(f).$$

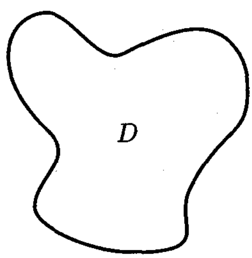


Рис. 1.19

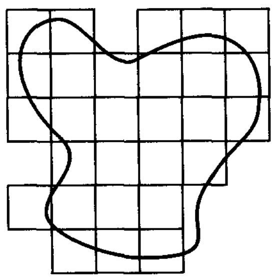


Рис. 1.20

Итак, $\mathrm{Ar}(f)$ есть коэффициент, показывающий, как изменяется площадь в результате преобразования f. Если f и g — два

[4]Здесь авторы фактически предполагают знакомство читателя с понятием площади для довольно широкого класса подмножеств плоскости, называемых «измеримыми по Жордану», или «квадрируемыми». Именно такие множества авторы называют «хорошими». Читатель должен также знать понятие «границы множества» и ее простейшие свойства. — *Прим. ред.*

[5]При очевидном условии, что (площадь D) $\neq 0$. — *Прим. ред.*

несингулярных линейных преобразования, то

$$\frac{\text{площадь } (f\circ g)(\Box)}{\text{площадь } \Box} = \frac{\text{площадь } f\circ g(\Box)}{\text{площадь } g(\Box)} \times \frac{\text{площадь } g(\Box)}{\text{площадь } \Box}$$

и, следовательно,

$$\mathrm{Ar}(f\circ g) = (\mathrm{Ar}(f)) \times (\mathrm{Ar}(g)).$$

Теперь вычислим коэффициенты $\mathrm{Ar}(f)$ для нескольких линейных преобразований специального вида, а затем используем полученные результаты, чтобы получить этот коэффициент в общем случае. Заметим, что $\mathrm{Ar}(f)$ = площадь $[f(\Box_0)]$ = площади образа единичного квадрата в результате действия преобразования f.[6]

Случай 1a. Преобразование f задается матрицей

$$F = (f) = \begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix}, \qquad r > 0, \quad s > 0.$$

Тогда, как видно из рисунка 1.21(a), $\mathrm{Ar}(f) = rs$.

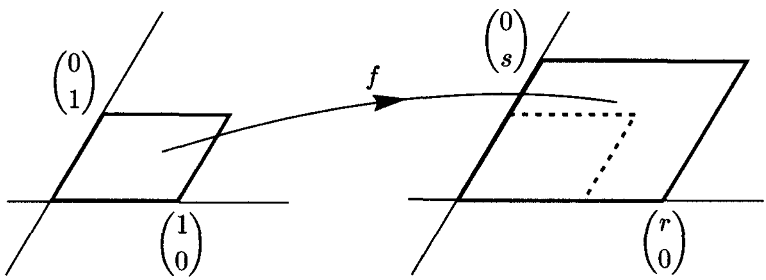


Рис. 1.21(a)

Случай 1b. Преобразование f задается матрицей

$$F = \begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix}, \qquad r < 0, \quad s > 0.$$

Тогда (рис. 1.21(b)) $\mathrm{Ar}(f) = |rs|$.

[6]Если принять, что площадь единичного квадрата $\Box_0$ равна единице. — *Прим. ред.*

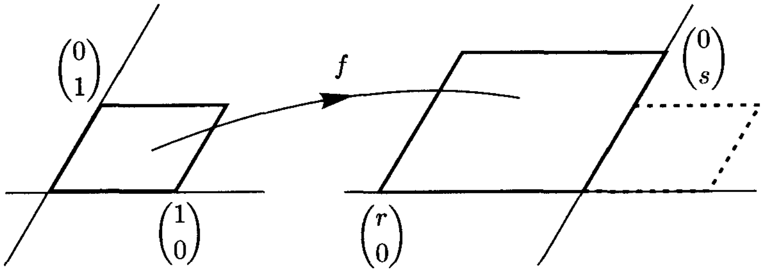


Рис. 1.21(b)

Из рассмотрения этих двух случаев видно, что $\mathrm{Ar}(f) = |rs|$ для любой диагональной матрицы F.

Случай 2a. Преобразование f задается матрицей $F = \begin{pmatrix}1 & x\\0 & 1\end{pmatrix}$.

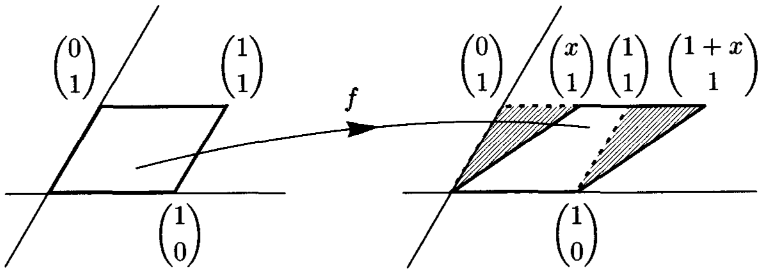


Рис. 1.22(a)

Заштрихованный треугольник (см. рис. 1.22(a)) с вершинами $\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1+x\\1\end{pmatrix}$ можно получить из заштрихованного треугольника с вершинами $\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}x\\1\end{pmatrix}$ трансляцией, прибавляя к каждой вершине вектор $\begin{pmatrix}1\\0\end{pmatrix}$. Поэтому образ единичного параллелограмма имеет ту же площадь, что и сам единичный параллелограмм. Это значит, что при таком *преобразовании сдвига* площадь не изменяется и коэффициент $\mathrm{Ar}(f) = 1$.

Случай 2b. Преобразование f задается матрицей $F = \begin{pmatrix}1 & 0\\y & 1\end{pmatrix}$.

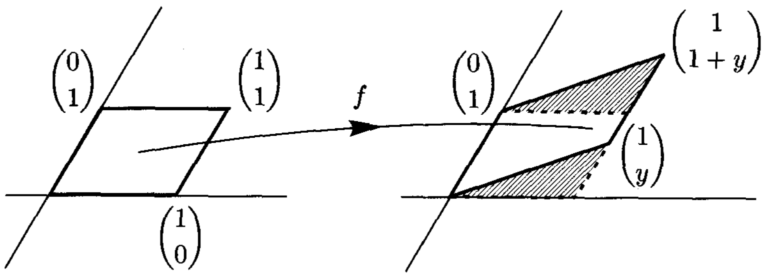


Рис. 1.22(b)

И в этом случае два заштрихованных треугольника имеют одинаковую площадь, т. е. коэффициент $\mathrm{Ar}(f) = 1$.

Теперь поговорим о связи между свойствами линейного преобразования и определителем его матрицы.

Определителем матрицы $F = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ называется число

$$\mathrm{Det}\, F = ad - bc. \tag{1.6}$$

Докажем важную формулу:

$$\mathrm{Ar}(f) = |\mathrm{Det}\, F|, \tag{1.7}$$

где F обозначает матрицу преобразования f. Чтобы доказать эту формулу, воспользуемся важным свойством определителей:

$$\mathrm{Det}(F \circ G) = (\mathrm{Det}\, F) \times (\mathrm{Det}\, G).$$

Оно проверяется простым умножением: если $G = \begin{pmatrix} e & f \\ g & h \end{pmatrix}$, то

$$\begin{aligned} \mathrm{Det}(F \circ G) &= \mathrm{Det} \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix} \\ &= (ae + bg)(cf + dh) - (af + bh)(ce + dg) \\ &= (ad - bc)(eh - fg) = (\mathrm{Det}\, F) \times (\mathrm{Det}\, G). \end{aligned}$$

Принимая во внимание два правила,

$$\mathrm{Ar}(f \circ g) = (\mathrm{Ar}(f) \times \mathrm{Ar}(g)) \quad \text{и} \quad \mathrm{Det}(F \circ G) = (\mathrm{Det}\ \mathrm{F}) \times (\mathrm{Det}\, G),$$

получаем, что формула

$$\boxed{\mathrm{Ar}(f) = |\mathrm{Det}\, F|}$$

верна для любого преобразования, которое можно записать как композицию преобразований, для которых уже проверена ее справедливость. В разделе 1.7 мы доказали, что если $a \neq 0$, можно записать (см. (1.5)):

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix} \begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix} = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}.$$

Таким образом, формула доказана для всех $a \neq 0$. Если $a = 0$, мы можем действовать двумя способами:

1. Непосредственной проверкой: $\mathrm{Ar}\begin{pmatrix} 0 & b \\ c & d \end{pmatrix} = |bc| = \left|\mathrm{Det}\begin{pmatrix} 0 & b \\ c & d \end{pmatrix}\right|$. Доказательства предоставляем читателю[7].

2. Воспользоваться непрерывностью[8]: заметим, что и $\mathrm{Ar}(f)$ и $\mathrm{Det}\, F$ — непрерывные функции коэффициентов матрицы F (т. е. если мы мало изменяем эти коэффициенты, то значения $\mathrm{Ar}(f)$ и $\mathrm{Det}\, F$ тоже изменяются мало.) Если матрица $\begin{pmatrix} 0 & b \\ c & d \end{pmatrix}$ несингулярна, то для достаточно малых e матрица $\begin{pmatrix} e & b \\ c & d \end{pmatrix}$ тоже несингулярна. Далее, поскольку мы знаем, что равенство

$$\mathrm{Ar}\begin{pmatrix} e & b \\ c & d \end{pmatrix} = \left|\mathrm{Det}\begin{pmatrix} e & b \\ c & d \end{pmatrix}\right|$$

выполняется для всех e, близких к нулю, то оно верно при $e = 0$.

[7]Легкое доказательство: если $c = 0$, то преобразование сингулярно, если же $c \neq 0$, то $\begin{pmatrix} 0 & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix} \times \begin{pmatrix} 1 & d/c \\ 0 & 1 \end{pmatrix}$. — *Прим. ред.*

[8]Следующее рассуждение использует не определенные ранее понятия непрерывной зависимости от параметра нескольких объектов: преобразования, его матрицы и функций от них. — *Прим. ред.*

Перейдем к выяснению геометрического смысла знака $\operatorname{Det} F$, если $\operatorname{Det} F \neq 0$. (Смысл модуля этого определителя мы уже обсудили.) Смысл знака $\operatorname{Det} F$ лучше всего показать на примере (рис. 1.23). Преобразование $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ — это отражение плоскости относительно оси x.

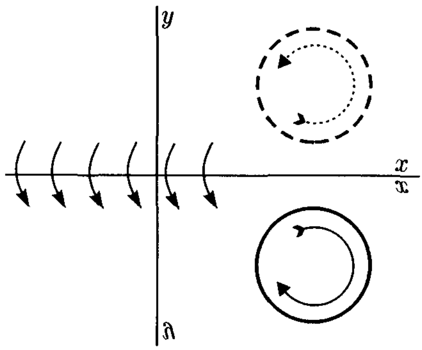


Рис. 1.23

Результатом отражения является изменение направления вращения против часовой стрелки на вращение по часовой стрелке. Таким образом, тот факт, что определитель отрицателен, означает изменение ориентации плоскости на противоположное, см. рис. 1.23.

В качестве иллюстрации давайте рассмотрим евклидово линейное преобразование плоскости, обозначаемое F. Поскольку вектор $F\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ имеет единичную длину, то для некоторого угла θ можно написать

$$F\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}.$$

Так как $F\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ тоже имеет единичную длину, и он перпендикулярен вектору $\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$, у нас есть две возможности:

$$F\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} \quad \text{или} \quad F\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \sin\theta \\ -\cos\theta \end{pmatrix}.$$

В первом случае мы имеем дело с поворотом на угол θ:

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \text{и} \quad \operatorname{Det} F = 1.$$

Во втором случае преобразование $F = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$, для которого $\operatorname{Det} F = -1$, является *отражением* относительно вектора $\begin{pmatrix} \cos\theta/2 \\ \sin\theta/2 \end{pmatrix}$. Действительно, из тригонометрических формул следует, что

$$\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta/2 \\ \sin\theta/2 \end{pmatrix} = \begin{pmatrix} \cos\theta/2 \\ \sin\theta/2 \end{pmatrix}.$$

Поэтому вектор $\begin{pmatrix} \cos\theta/2 \\ \sin\theta/2 \end{pmatrix}$ после преобразования F остается на месте и прямая, проведенная из начала вдоль этого вектора, отображается сама в себя. Далее $F \circ F = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, т. е. преобразование F действительно является отражением.

1.10. Обратные матрицы

Пусть $F = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ — любая матрица порядка 2×2.

Матрица G называется обратной к F, если $GF = FG = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Если $\operatorname{Det} F = 0$, то матрица F не имеет обратной, поскольку $(\operatorname{Det} F) \cdot (\operatorname{Det} G) = \operatorname{Det} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1 \neq 0$. Если же $\operatorname{Det} F \neq 0$, то мы найдем обратную к F матрицу G с помощью следующего построения.

Определим матрицу F^a согласно формуле

$$F^a = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Непосредственное умножение показывает, что

$$F^aF = FF^a = \begin{pmatrix} \operatorname{Det} F & 0 \\ 0 & \operatorname{Det} F \end{pmatrix} = \operatorname{Det} F \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Поэтому ясно, что в качестве матрицы, обратной к F, можно взять

$$G = (\operatorname{Det} F)^{-1}F^a.$$

Таким образом, мы доказали теорему: матрица F имеет обратную тогда и только тогда, когда $\operatorname{Det} F \neq 0$. В этом случае для обратной матрицы F^{-1} справедлива формула

$$F^{-1} = \begin{pmatrix} d(\operatorname{Det} F)^{-1} & -b(\operatorname{Det} F)^{-1} \\ -c(\operatorname{Det} F)^{-1} & a(\operatorname{Det} F)^{-1} \end{pmatrix} \tag{1.8}$$

Нам следует понять геометрический смысл F^{-1}; эта матрица «уничтожает» действие матрицы F. Если сначала мы произведем операцию F, а потом F^{-1}, то вернемся в исходное состояние[9], поскольку $FF^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ — соответствует тождественному отображению.

Заметим, что если F сингулярна, то она не имеет обратной матрицы, потому что для любой матрицы F, имеющей обратную, из $F\mathbf{v} = F\mathbf{w}$ следует $F^{-1}F\mathbf{v} = F^{-1}F\mathbf{w}$, т. е. $\mathbf{v} = \mathbf{w}$.

Тот факт, что условие условия $\operatorname{Det} F = 0$ соответствует сингулярности F, можно интерпретировать в терминах площадей. Действительно, пусть параллелограмм имеет одну вершину в начале и стороны, параллельные векторам $\begin{pmatrix} a \\ c \end{pmatrix}$ и $\begin{pmatrix} b \\ d \end{pmatrix}$. Параллелограмм имеет площадь, не равную нулю, если эти вектора не лежат на одной прямой, проходящей через начало. Это значит, что $\operatorname{Det} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq 0$, потому что $\left|\operatorname{Det} \begin{pmatrix} a & b \\ c & d \end{pmatrix}\right|$ равен площади этого параллелограмма. В этом случае существует обратная матрица, и мы можем написать

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix}, \quad \text{где} \quad F^{-1} = \begin{pmatrix} e & f \\ g & h \end{pmatrix}.$$

[9]Здесь и иногда далее авторы не делают различия между линейным преобразованием и его матрицей. — *Прим. ред.*

То же самое можно написать по-другому: $\begin{pmatrix}1\\0\end{pmatrix} = e\begin{pmatrix}a\\c\end{pmatrix} + g\begin{pmatrix}b\\d\end{pmatrix}$ и $\begin{pmatrix}0\\1\end{pmatrix} = f\begin{pmatrix}a\\c\end{pmatrix} + h\begin{pmatrix}b\\d\end{pmatrix}$. А это значит, что любой вектор на плоскости может быть представлен как линейная комбинация векторов $\begin{pmatrix}a\\c\end{pmatrix}$ и $\begin{pmatrix}b\\d\end{pmatrix}$, где e, f, g и h — просто числа.

Вспомним (см. параграф 1.1), что вектора $\mathbf{u}$ и $\mathbf{v}$ называются *линейно зависимыми*, если существуют вещественные числа r и s, не равные нулю, такие, что

$$r\mathbf{u} + s\mathbf{v} = 0.$$

Если векторы $\mathbf{u}$, $\mathbf{v}$ не являются линейно зависимыми, то их называют *линейно независимыми*.

Пусть $\mathbf{u} = \begin{pmatrix}x\\y\end{pmatrix}$ и $\mathbf{v} = \begin{pmatrix}z\\t\end{pmatrix}$. Вектора $\mathbf{u}$ и $\mathbf{v}$ линейно независимы тогда и только тогда, когда матрица

$$M = \begin{pmatrix}x & z\\ y & t\end{pmatrix}$$

несингулярна. Действительно, матрица M переводит единичные вектора $\begin{pmatrix}1\\0\end{pmatrix}$ и $\begin{pmatrix}0\\1\end{pmatrix}$ в $\mathbf{u}$ и $\mathbf{v}$ соответственно. Таким образом, $\mathbf{u}$ и $\mathbf{v}$ будут лежать на одной прямой (а это и означает их линейную зависимость) тогда и только тогда, когда матрица M преобразует единичный квадрат в вырожденный параллелограмм нулевой площади, т. е. при условии, что $\operatorname{Det} M = 0$.

Предположим, что $\mathbf{u}$ и $\mathbf{v}$ линейно независимы, и мы имеем матрицу M, определенную выше. Пусть $\mathbf{w}$ есть любой вектор на плоскости $\mathbb{R}^2$. Поскольку M несингулярна, то мы можем построить обратную матрицу M^{-1} и взять вектор $M^{-1}\mathbf{w}$. Этот вектор определен на плоскости $\mathbb{R}^2$ и, следовательно, можно написать

$$M^{-1}\mathbf{w} = \begin{pmatrix}a\\b\end{pmatrix} = a\begin{pmatrix}1\\0\end{pmatrix} + b\begin{pmatrix}0\\1\end{pmatrix}.$$

Умножаем обе части этого уравнения на M и получаем

$$MM^{-1}\mathbf{w} = \mathbf{w} = aM\begin{pmatrix}1\\0\end{pmatrix} + bM\begin{pmatrix}0\\1\end{pmatrix} = a\mathbf{u} + b\mathbf{v}.$$

Таким образом, если $\mathbf{u}$ и $\mathbf{v}$ линейно независимы, то любой вектор на плоскости может быть записан как линейная комбинация этих векторов. Обратно, допустим, что любой вектор на плоскости можно записать в виде линейной комбинации векторов $\mathbf{u}$ и $\mathbf{v}$. Очевидно, что $\mathbf{u}$ и $\mathbf{v}$ не могут лежать на одной прямой, проходящей через начало, так как это означало бы, что любой вектор на плоскости лежит на той же прямой. Таким образом,

> векторы $\mathbf{u}$ и $\mathbf{v}$ линейно независимы тогда и только тогда, когда любой вектор на плоскости может быть представлен как линейная комбинация $\mathbf{u}$ и $\mathbf{v}$.

Допустим, что F — матрица, а $\mathbf{u}_1$ и $\mathbf{u}_2$ — два линейно независимых вектора. Параллелограмм, построенный на векторах $\mathbf{u}_1$ и $\mathbf{u}_2$, имеет ненулевую площадь, следовательно, $F(\mathbf{u}_1)$ и $F(\mathbf{u}_2)$ будут линейно независимы тогда и только тогда, когда $\operatorname{Det} F \neq 0$.

Подытожим сказанное.

Все утверждения, приведенные ниже, эквивалентны.

1. Матрица $F = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ имеет обратную.
2. $\operatorname{Det} F \neq 0$.
3. Векторы $\begin{pmatrix} a \\ c \end{pmatrix}$ и $\begin{pmatrix} b \\ d \end{pmatrix}$ не лежат на одной прямой, проходящей через начало координат.
4. Любой вектор на плоскости можно записать в виде линейной комбинации $\begin{pmatrix} a \\ c \end{pmatrix}$ и $\begin{pmatrix} b \\ d \end{pmatrix}$.
5. Для некоторой пары векторов $\mathbf{u}_1$, $\mathbf{u}_2$ соответствующие вектора $F(\mathbf{u}_1)$ и $F(\mathbf{u}_2)$ линейно независимы.
6. Для *любой* пары линейно независимых векторов $\mathbf{v}_1$, $\mathbf{v}_2$ вектора $F(\mathbf{v}_1)$ и $F(\mathbf{v}_2)$ линейно независимы.
7. F — несингулярная матрица.

Давайте используем эти результаты для иллюстрации ряда положений из *аффинной геометрии*. Для начала отметим, что в рамках аффинной геометрии не имеет смысла не только длина отрезка, но и относительные длины двух отрезков, не лежащих на одной прямой. Действительно, если $\mathbf{u}$ и $\mathbf{v}$ — два линейно независимых вектора, то существует единственное линейное преобразование f, которое выполняет операции: $\mathbf{u} \mapsto r\mathbf{u}$ и $\mathbf{v} \mapsto s\mathbf{v}$, где r и s — любые числа, не равные нулю. Таким образом, подбирая s/r, мы можем сделать любым отношение длин $f(\mathbf{u})$ и $f(\mathbf{v})$.

С другой стороны, отношение длин отрезков, лежащих на одной прямой, в аффинной геометрии имеет смысл. Действительно, раз трансляции сохраняют длину, мы можем предположить, что прямая l и ее образ $f(l)$ проходят через начало. Так как вращения тоже сохраняют длину, то мы можем сделать поворот и считать, что $f(l) = l$. Но тогда, если $\mathbf{0} \neq \mathbf{u} \in l$, образ $f(\mathbf{u})$ тоже лежит на прямой l, поэтому $f(\mathbf{u}) = c\mathbf{u}$, где c — постоянный коэффициент. Следовательно, $f(\mathbf{v}) = c\mathbf{v}$ для любого вектора $\mathbf{v} \subset l$. Это значит, что f изменяет длины всех отрезков на прямой l одинаково, умножая их на $|c|$.

Следует также отметить, что для заданных треугольников Δ_1 и Δ_2 существует аффинное преобразование f, для которого $f(\Delta_1) = \Delta_2$. На самом деле, с помощью трансляций одну из вершин треугольника Δ_1 можно поместить в начало. Пусть концы векторов $\mathbf{u}_1$ и $\mathbf{v}_1$ находятся в двух других вершинах. Вектора $\mathbf{u}_1$ и $\mathbf{v}_1$ линейно независимы, поэтому вершины треугольника $\mathbf{0}$, $\mathbf{u}_1$, $\mathbf{v}_1$ не лежат на одной прямой. Аналогично можно считать, что вершинами треугольника Δ_2 являются точки $\mathbf{0}$, $\mathbf{u}_2$, $\mathbf{v}_2$. Но тогда существует единственное линейное преобразование f, для которого $f(\mathbf{u}_1) = \mathbf{u}_2$ и $f(\mathbf{v}_1) = \mathbf{v}_2$. Теперь мы можем сформулировать теорему: *для любого треугольника три прямые, соединяющие вершины со средними точками противоположных сторон, пересекаются в одной точке.*

Понятие средней точки в рамках аффинной геометрии существует, поэтому утверждение, что эти три прямые пересекаются в одной (общей) точке, имеет смысл. Чтобы *доказать* эту теорему, достаточно рассмотреть один треугольник, потому что мы можем найти аффинное преобразование, переводящее один треугольник в другой. Поэтому, если теорема справедлива для одного треугольника, то она будет справедливой и для любого другого. Теорема очевидна для равносторонних треугольников. Поэтому можно считать, что она доказана и в общем случае.

1.11. Сингулярные матрицы

Давайте разберемся, что происходит, когда $\operatorname{Det} F = 0$. В этом случае возможны два варианта:

$$F = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{или} \quad F \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Если F — нулевая матрица, то каждый вектор она отображает в $\mathbf{0}$, вся плоскость стягивается в начальную точку. В другом случае, когда F — не нулевая матрица, но $\operatorname{Det} F = 0$, справедливы следующие утверждения.

1. Существует такая прямая l, что $F(\mathbf{u}) \in l$ для каждого вектора $\mathbf{u}$ на плоскости $\mathbb{R}^2$. Более того, каждый вектор $\mathbf{v} \in l$ является образом некоторого $\mathbf{u} \in \mathbb{R}^2$, т. е. $\mathbf{v} = F(\mathbf{u})$. Другими словами, матрица F стягивает плоскость в прямую l.

2. Существует такая прямая k, что $F(\mathbf{w}) = \mathbf{0}$ тогда и только тогда, когда $\mathbf{w} \in k$. Другими словами, матрица F стягивает прямую k в начальную точку и не переводит в $\mathbf{0}$ ни одного вектора, не принадлежащего прямой k.

Докажем утверждение 1. Пусть

$$\mathbf{c}_1 = \begin{pmatrix} a \\ c \end{pmatrix} = F\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) \quad \text{и} \quad \mathbf{c}_2 = \begin{pmatrix} b \\ d \end{pmatrix} = F\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$$

образуют два столбца матрицы $F = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Поскольку F — ненулевая матрица, $\mathbf{c}_1$ и $\mathbf{c}_2$ не могут одновременно быть равными $\mathbf{0}$. С другой стороны, если $\mathbf{c}_1$ и $\mathbf{c}_2$ не лежат на одной прямой, то из утверждений 2 и 3 предыдущего параграфа следует, что $\operatorname{Det} F \neq 0$, что противоречит исходному предположению. Таким образом, $\mathbf{c}_1$ и $\mathbf{c}_2$ лежат на одной прямой.

Обозначим ее l. Любой вектор $\mathbf{u} \in \mathbb{R}^2$ можно представить в виде

$$\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

поэтому $F(\mathbf{u}) = x\mathbf{c}_1 + y\mathbf{c}_2$ лежит на той же прямой l.

Если $\mathbf{c}_1 \neq 0$, то любой вектор $\mathbf{v} \in l$ равен $\mathbf{v} = x\mathbf{c}_1$, где x — некоторое число. Следовательно, $\mathbf{v} = F\left(\begin{pmatrix} x \\ 0 \end{pmatrix}\right)$. Если же $\mathbf{c}_2 \neq \mathbf{0}$, то $\mathbf{v} = y\mathbf{c}_2$ и вектор $\mathbf{v} = F\left(\begin{pmatrix} 0 \\ y \end{pmatrix}\right)$. Таким образом, мы доказали утверждение 1.

Теперь докажем утверждение 2. Пусть $\mathbf{b}_1$ и $\mathbf{b}_2$ — первый и второй столбцы матрицы F^α, т. е.

$$\mathbf{b}_1 = \begin{pmatrix} d \\ -c \end{pmatrix} \quad \text{и} \quad \mathbf{b}_2 = \begin{pmatrix} -b \\ a \end{pmatrix}.$$

Поскольку $\operatorname{Det} F^\alpha = \operatorname{Det} F = 0$ и из (7) мы знаем, что $\mathbf{b}_1$ и $\mathbf{b}_2$ линейно зависимы. Кроме того, они не могут быть одновременно равны $\mathbf{0}$, т. к. мы предположили, что исходная матрица $F \neq 0$. Значит, эти вектора порождают прямую, которую мы обозначим k. Простое вычисление показывает, что $F\mathbf{b}_1 = F\mathbf{b}_2 = \mathbf{0}$. Следовательно, каждый вектор $\mathbf{w} \in k$ удовлетворяет соотношению $F\mathbf{w} = \mathbf{0}$.

Обратно, пусть вектор $\mathbf{w} = \begin{pmatrix} x \\ y \end{pmatrix}$ удовлетворяет условию $F\mathbf{w} = \mathbf{0}$, и пусть хотя бы одно из чисел a, b не равно нулю. Тогда координаты вектора $\mathbf{w}$ удовлетворяют уравнению $ax + by = 0$, которое, очевидно, задает прямую k. Если же $a = b = 0$, то хотя бы одно из чисел c, d не равно нулю, и выполняется равенство $cx + dy = 0$, что снова является уравнением прямой k. В обоих случаях мы получили $\mathbf{w} \in k$, что и доказывает утверждение 2.

Обозначим $\operatorname{Im}(F)$ подмножество $\mathbb{R}^2$, состоящее из всех элементов вида $F(\mathbf{u})$. Запишем это символически:

$$\operatorname{Im}(F) = \{\mathbf{v} \mid \mathbf{v} = F(\mathbf{u}) \quad \text{для некоторого } \mathbf{u}\}.$$

Читается обозначение $\operatorname{Im}(F)$ так: «образ отображения F».

Аналогично определяем «ядро отображения F», обозначая его символом $\mathrm{Ker}(F)$. Так мы будем обозначать множество векторов, которые преобразование F переводит в $\mathbf{0}$. Запишем это формулой:

$$\mathrm{Ker}(F) = \{\mathbf{w} \mid f(\mathbf{w}) = \mathbf{0}\}.$$

Такм образом, в зависимости от коэффициентов матрицы F существует три возможности.

1. $\mathrm{Det}\, F \neq 0$. Тогда $\mathrm{Im}(F) = \mathbb{R}^2$ и $\mathrm{Ker}(F) = \{\mathbf{0}\}$.

2. $\mathrm{Det}\, F = 0$, но F — ненулевая матрица. Тогда $\mathrm{Im}(F) = l$ есть прямая и $\mathrm{Ker}(F) = k$ тоже прямая. Другими словами, $\mathrm{Im}(F)$ и $\mathrm{Ker}(F)$ образуют одномерные векторные пространства (рис. 1.24).

3. F — нулевая матрица. Тогда $\mathrm{Im}(F) = \{\mathbf{0}\}$ и $\mathrm{Ker}(F) = \mathbb{R}^2$.

Если рассматривать $\{\mathbf{0}\}$ как векторное пространство нулевой размерности, то можно написать общее выражение

$$\text{размерность } \mathrm{Im}(F) + \text{размерность } \mathrm{Ker}(F) = 2.$$

Специальные случаи сингулярных преобразований. Сейчас мы изучим несколько частных случаев сингулярных линейных преобразований.

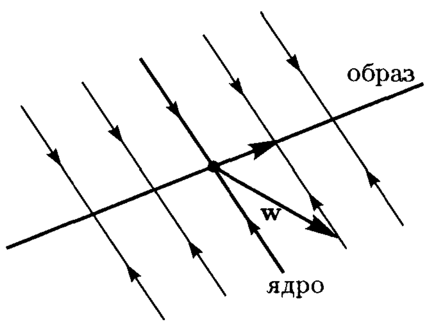


Рис. 1.24

1. Проекции. Будем называть преобразование p *проекцией*, если оно обладает свойством:

$$\text{если } \mathbf{v} \text{ есть вектор из образа } p, \text{ то } p(\mathbf{v}) = \mathbf{v}.$$

Для произвольного сингулярного преобразования $f : \mathbb{R}^2 \to \mathbb{R}^2$ можно было бы лишь сказать, что при $\mathbf{v} \in \operatorname{Im}(f)$ вектор $f(\mathbf{v})$ лежит на той же прямой, что и $\mathbf{v}$, т. е. $f(\mathbf{v}) = \alpha\mathbf{v}$ для некоторого числа α. Если $\mathbf{w}$ — произвольный вектор, то $\mathbf{v} = p(\mathbf{w})$ принадлежит образу p и

$$(p \circ p)(\mathbf{w}) = p(\mathbf{v}) = \mathbf{v} = p(\mathbf{w}).$$

Отсюда следует, что $p \circ p = p$, и поэтому матрица P, представляющая проекцию, удовлетворяет условию $P^2 = P$. Таким образом,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2 + bc & ab + bd \\ ac + cd & bc + d^2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Поэтому $ab + bd = b$, и для $b \neq 0$ получаем $a + d = 1$. Далее, $ac + cd = c$, для $c \neq 0$ снова получаем $a + d = 1$. В случае, когда $b = 0$ и $c = 0$, мы имеем $a^2 = a$, $d^2 = d$ и $\operatorname{Det} P = ad = 0$, т. е. либо $a = 1$, $d = 0$ либо $a = 0$, $d = 1$, либо $P = 0$. Поэтому, кроме случая $P = 0$, *след* P, обозначаемый $\operatorname{Tr} P = a + d$, должен быть равен 1.

Итак, ненулевая (сингулярная) проекция p удовлетворяет условию $p \circ p = p$; матрица P удовлетворяет условию $P^2 = P$ и имеет нулевой определитель ($ad - bc = 0$) и единичный след ($a + d = 1$).

Обратно, предположим, что $ad - bc = 0$ и $a + d = 1$. Тогда $ab+bd = (a+d)b = b$ и $ac+cd = c$, кроме того, $a^2+bc = a^2+ad = a$ и $bc+d^2 = d$. Таким образом, $P^2 = P$ и p есть проекция на прямую.

В общем случае будем называть оператор p *проекцией*, если $p \circ p = p$. Тогда у нас есть три возможности.

1. P несингулярна. В этом случае обе части уравнения $P^2 = P$ можно умножить на P^{-1} и получить

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

   В этом случае $\operatorname{Tr} P = 2 = \dim(\operatorname{Im}(P))$, где $\dim$ обозначает размерность.

2. P сингулярна, но не равна нулю. Мы уже рассмотрели этот случай раньше. Здесь P отображает $\mathbb{R}^2$ на прямую и $\operatorname{Tr} P = 1 = \dim(\operatorname{Im}(P))$.

3. $P = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, поэтому P стягивает всю плоскость в начальную точку и $\operatorname{Tr} P = 0 = \dim(\operatorname{Im}(P))$.

Во всех трех случаях выполняется равенство $\operatorname{Tr} P = \dim(\operatorname{Im}(P))$.

2. Нильпотенты. Для произвольной сингулярной матрицы определены две прямые $\operatorname{Im}(F)$ и $\operatorname{Ker}(F)$. Рассмотрим частный случай, а именно, преобразование n, для которого прямые его образа и его ядра совпадают. Действуя преобразованием n на произвольный вектор $\mathbf{w}$, получаем вектор $\mathbf{v} = n(\mathbf{w})$, который принадлежит одновременно образу и ядру преобразования n. Отсюда следует, что $n(\mathbf{v}) = n \circ n(\mathbf{w}) = \mathbf{0}$. Таким образом, преобразование $n \circ n$ стягивает всю плоскость в начальную точку и матрица N, определяющая n, удовлетворяет условию $N^2 = 0$. Поэтому

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2+bc & ab+bd \\ ac+cd & bc+d^2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

В частности $ab + bd = 0$ и $ac + cd = 0$, поэтому если $b \neq 0$ или $c \neq 0$, то $a + d = 0$. Если $b = c = 0$, то $a^2 = d^2 = 0$ и $N = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. В любом случае матрица N имеет нулевой след. И обратно, если $\operatorname{Tr} N = a + d = 0$ и $\operatorname{Det} N = ad - bc = 0$, то $N^2 = 0$.

Будем называть матрицу *нильпотентом*, если $N^k = 0$ для некоторого k. У такой матрицы необходимо $\operatorname{Det} N = 0$, т. к. $\operatorname{Det}(N)^k = (\operatorname{Det} N)^k$. Итак, $ad - bc = 0$ и, следовательно,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 = \begin{pmatrix} (a+d)a & (a+d)b \\ (a+d)c & (a+d)d \end{pmatrix}.$$

Тогда

$$N^k = \begin{pmatrix} (a+d)^{k-1}a & (a+d)^{k-1}b \\ (a+d)^{k-1}c & (a+d)^{k-1}d \end{pmatrix}.$$

Эта матрица становится нулевой, если $a + d = 0$. Но в этом случае уже $N^2 = 0$. Таким образом, на плоскости матрица N является нильпотентом тогда и только тогда, когда выполняется условие $N^2 = 0$, а оно выполняется тогда и только тогда, когда

$$\operatorname{Det} N = 0 \quad \text{и} \quad \operatorname{Tr} N = 0.$$

1.12. Двумерные векторные пространства

Векторное пространство V называется *двумерным*, если в V мы можем найти такие два вектора $\mathbf{u}_1$ и $\mathbf{u}_2$, что каждый $\mathbf{v} \in V$ можно единственным способом представить в виде

$$\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2.$$

«Единственность» означает, что если

$$\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 \qquad \text{и} \qquad \mathbf{v} = b_1\mathbf{u}_1 + b_2\mathbf{u}_2,$$

то необходимо выполняются равенства

$$a_1 = b_1 \qquad \text{и} \qquad a_2 = b_2.$$

Упорядоченная пара векторов $\mathbf{u}_1$, $\mathbf{u}_2$, обладающая описанным свойством, называется *базисом* векторного пространства V. Выбор базиса определяет отображение $L = L_{\mathbf{u}_1,\mathbf{u}_2}$ векторного пространства V на плоскость $\mathbb{R}^2$

$$V \xrightarrow{L} \mathbb{R}^2$$

такое, что если

$$\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2, \quad \text{то} \quad L(\mathbf{v}) = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}.$$

Предполагаемая единственность гарантирует, что отображение L однозначно определено, компоненты a_1 и a_2 полностью определяются вектором $\mathbf{v}$. Отображение L сюръективно: при любых заданных $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ для $\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2$ равенство $L(\mathbf{v}) = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ удовлетворяется очевидным образом[10].

Отметим некоторые свойства отображения L.

Если $\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2$ и $\mathbf{w} = b_1\mathbf{u}_1 + b_2\mathbf{u}_2$, то

$$\mathbf{v} + \mathbf{u} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + b_1\mathbf{u}_1 + b_2\mathbf{u}_2 = (a_1 + b_1)\mathbf{u}_1 + (a_2 + b_2)\mathbf{u}_2,$$

[10] Следовало бы еще отметить инъективность отображения L. — *Прим. ред.*

и, следовательно,

$$L(\mathbf{v}+\mathbf{w}) = L(\mathbf{v}) + L(\mathbf{w}).$$

Аналогично,

$$L(r\mathbf{v}) = rL(\mathbf{v})$$

для любого вещественного числа r и любого вектора $\mathbf{v}$ в пространстве V.

Имея в виду все эти свойства, мы говорим, что L — *изоморфное*[11] отображение (или *изоморфизм*) пространства V на $\mathbb{R}^2$. Это позволяет нам все свойства и операции векторного пространства V отождествить с соответствующими свойствами и операциями в $\mathbb{R}^2$ — так же, как в одномерном случае свойства векторного пространства мы перенесли на $\mathbb{R}^1$. Конечно, как и в одномерном случае, изоморфизм L зависит от выбора базиса. Таким образом, выбор базиса $\{\mathbf{u}_1, \mathbf{u}_2\}$ есть двумерный аналог «выбора единиц измерения». Нас будут интересовать только те свойства, которые не зависят от выбора базиса и имеют истинно геометрическую природу.

Давайте коротко изучим, как изменяется L при замене базиса. Сначала рассмотрим, как, зная L, можно восстановить базис $\{\mathbf{u}_1, \mathbf{u}_2\}$. Действительно,

$$\mathbf{u}_1 = L^{-1}\begin{pmatrix}1\\0\end{pmatrix} \quad \text{и} \quad \mathbf{u}_2 = L^{-1}\begin{pmatrix}0\\1\end{pmatrix}.$$

Таким образом, *задание базиса равнозначно заданию изоморфизма*, $L : V \to \mathbb{R}^2$. Выбрав L, мы можем из этих уравнений получить базисные вектора $\mathbf{u}_1$ и $\mathbf{u}_2$. Поскольку любой вектор в $\mathbb{R}^2$ может быть однозначно записан в виде линейной комбинации векторов $\begin{pmatrix}1\\0\end{pmatrix}$ и $\begin{pmatrix}0\\1\end{pmatrix}$ и поскольку L — изоморфное отображение, то каждый вектор в V может быть однозначно записан в виде линейной

[11] В математике термин «изоморфизм» означает взаимно однозначное (биективное) отображение, сохраняющее все относящиеся к делу структуры. Для векторных пространств V и W мы говорим, что отображение L пространства V на W является изоморфизмом, если оно линейно, инъективно и сюръективно (т. е. биективно).

комбинации векторов $\mathbf{u}_1$, $\mathbf{u}_2$. Таким образом, изоморфизм, связанный с $\{\mathbf{u}_1, \mathbf{u}_2\}$, очевидно, есть L.

Линейным преобразованием $F : V \to V$ называется отображение, удовлетворяющее знакомому тождеству

$$F(a\mathbf{u} + b\mathbf{v}) = aF(\mathbf{u}) + bF(\mathbf{v}).$$

Выбрав базис в пространстве V, определяющий отображение $L : V \to \mathbb{R}^2$, мы можем определить линейное преобразование пространства $\mathbb{R}^2$ следующим образом:

$$LFL^{-1}.$$

Здесь $L^{-1} : \mathbb{R}^2 \to V$, $F : V \to V$ и $L : V \to \mathbb{R}^2$. Эту ситуацию лучше всего видеть на диаграмме:

$$\begin{tikzcd}
V \arrow[r, "F"] & V \arrow[d, "L"] \\
\mathbb{R}^2 \arrow[u, "L^{-1}"] \arrow[r] & \mathbb{R}^2
\end{tikzcd}$$

Преобразование $LFL^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$, соответствующее нижней линии, получается после трех последовательных преобразований: вверх по левой линии диаграммы, вправо по верхней линии и вниз по правой. Линейное преобразование $\mathbb{R}^2 \to \mathbb{R}^2$ задается матрицей. Таким образом, выбрав базис L, мы определяем соответствующую матрицу преобразования F

$$\mathrm{Mat}_L(F) = \mathrm{Mat}(LFL^{-1})$$

для любого линейного оператора $F : V \to V$. Если $G : V \to V$ — другое линейное преобразование, то

$$LGFL^{-1} = LGL^{-1}LFL^{-1},$$

поэтому

$$\mathrm{Mat}_L(G \circ F) = \mathrm{Mat}_L(G)\mathrm{Mat}_L(F).$$

Другими словами, композиции линейных преобразований соответствует умножение их матриц. Аналогичный вывод можно сделать для сложения линейных преобразований. Таким образом, алгебра линейных преобразований двумерного векторного пространства преобразуется в алгебру матриц 2×2.

Пространство $\mathbb{R}^2$ является векторным пространством. Оно имеет «естественный» базис, состоящий из векторов $\begin{pmatrix}1\\0\end{pmatrix}$ и $\begin{pmatrix}0\\1\end{pmatrix}$. Если $f : \mathbb{R}^2 \to \mathbb{R}^2$ — линейное преобразование, то, пользуясь языком предыдущих параграфов, соответствующая естественному базису матрица — это F. Отображение $L : \mathbb{R}^2 \to \mathbb{R}^2$, соответствующее этому базису — просто тождественное отображение I. Таким образом, связь между f и F следует записать в виде

$$F = \mathrm{Mat}_I(f).$$

Следуя строгой логике, нам следовало бы с самого начала использовать обозначение $\mathrm{Mat}_I(f)$ вместо F, но это было бы слишком громоздко. С настоящего момента, владея понятием линейного преобразования произвольного векторного пространства, мы не будем делать различие между большими и малыми буквами в скобках.

Привязка $\mathrm{Mat}_L(F)$ к F зависит от выбора базиса. Давайте посмотрим, что произойдет, если базис изменить. Предположим, что у нас есть два базиса. Это значит, что нам заданы два изоморфизма: $L : V \to \mathbb{R}^2$ и $M : V \to \mathbb{R}^2$. Рассмотрим матрицу $B = ML^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$, обладающую свойством

$$M = BL.$$

Эту ситуацию легко видеть на диаграмме:

$$\begin{tikzcd} & V \arrow[dl, "L"'] \arrow[dr, "M"] & \\ \mathbb{R}^2 \arrow[rr, "B"'] & & \mathbb{R}^2 \end{tikzcd}$$

Матрица B называется «матрицей замены базиса» (устанавливает связь между базисами L и M). Это двумерный аналог множителя 1000, на который мы должны умножать численные значения масс при переходе от килограммов к граммам в новой системе единиц. Итак, повторяем: $L(\mathbf{v})$ и $M(\mathbf{v})$ — это две точки в $\mathbb{R}^2$, соответствующие одной и той же точке $\mathbf{v}$ в пространстве V для двух различных базисов L и M. Эти две точки в $\mathbb{R}^2$ связаны между собой матрицей замены базиса:

$$M(\mathbf{v}) = BL(\mathbf{v}).$$

Матрица B задает изоморфизм $\mathbb{R}^2 \to \mathbb{R}^2$ (это композиция двух изморфизмов), т. е. она не сингулярна. (Ясно, что если мы имеем изоморфизм L и дана любая несингулярная матрица B, то можно определить $M = BL$, при этом матрица M — это изоморфизм $V \to \mathbb{R}^2$. Итак, коль скоро мы выбрали базис, т. е. имеем отображение $L : V \to \mathbb{R}^2$, то множество других базисов в пространстве V параметризуется множеством всех обратимых матриц 2×2, обозначенных у нас B.)

Пусть $F : V \to V$ — некоторое линейное преобразование. Тогда $\mathrm{Mat}_L(F) = LFL^{-1}$ и $\mathrm{Mat}_M(F) = MFM^{-1}$.

Следующая выкладка приводит к важной формуле:

$$MFM^{-1} = (BL)F(BL)^{-1} = BLFL^{-1}B^{-1} = B(LFL^{-1})B^{-1},$$

поэтому

$$\mathrm{Mat}_M(F) = B\mathrm{Mat}_L(F)B^{-1}.$$

Эта формула показывает, как две матрицы одного и того же преобразования связаны в разных базисах между собой, если мы знаем матрицу замены базиса B.

Для заданного линейного преобразования $F : V \to V$ можно *выбрать* такой базис L, что $\mathrm{Mat}_L(F)$ имеет наиболее удобную форму. Например, допустим, что $F : V \to V$ стягивает все пространство V в прямую и эту прямую преобразует в точку $\mathbf{0}$. Другими словами, пусть $\mathrm{Im}\, F = \mathrm{Ker}\, F$. Возьмем вектор $\mathbf{u}_2$, *не* принадлежащий $\mathrm{Ker}\, F$; обозначим $\mathbf{u}_1 = F(\mathbf{u}_2)$, так что $\mathbf{u}_1 \neq \mathbf{0}$, а $F(\mathbf{u}_1) = \mathbf{0}$. Вектора $\mathbf{u}_1$ и $\mathbf{u}_2$ выберем в качестве базиса. Тогда $LFL^{-1}\begin{pmatrix}1\\0\end{pmatrix} = LF(\mathbf{u}_1) = L(\mathbf{0}) = \begin{pmatrix}0\\0\end{pmatrix}$ и $LFL^{-1}\begin{pmatrix}0\\1\end{pmatrix} = LF(\mathbf{u}_2) = L(\mathbf{u}_1) = \begin{pmatrix}1\\0\end{pmatrix}$. При таком выборе базиса матрица преобразования

F имеет следующий вид:

$$\operatorname{Mat}_L(F) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \tag{1.9}$$

В ходе этих рассуждений ничто не мешает нам рассмотреть случай, когда векторное пространство V есть $\mathbb{R}^2$. Используя стандартный базис, мы задали линейное преобразование в виде матрицы. Другими словами, когда в параграфах 1.5 и 1.6 мы писали F, на самом деле следовало писать $\operatorname{Mat}_I(F)$. Например, пусть N — ненулевая нильпотентная матрица. Тогда $N = \operatorname{Mat}_I(F)$, где F — линейное преобразование $\mathbb{R}^2$ с $\operatorname{Ker} F = \operatorname{Im} F$. (Это значит, что N есть матрица линейного преобразования F в стандартном базисе.) Предыдущие рассуждения позволяют нам найти другой базис L, для которого выполняется (1.9). Формула замены базиса (переходим от базиса L к базису I) дает нам

$$N = B \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} B^{-1}. \tag{1.10}$$

Итак, мы доказали следующее: задав любую ненулевую нильпотентную матрицу N, можно найти такую обратимую матрицу B, что выполняется (1.10). В следующей главе мы еще к этому вернемся (в частности, рассмотрим, как получить B).

Важные физические приложения результатов этой главы изложены в главе 9. В ней мы покажем, как гауссова оптика сводится к изучению матриц 2×2. Однако, большую часть главы 9 можно читать, только зная материал главы 1.

Добавление: фундаментальная теорема аффинной геометрии

Мы хотим доказать следующее:

Пусть f — аффинное преобразование $\mathbb{R}^2$, удовлетворяющее условию $f\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. *Тогда f линейно.*

Для доказательства этой теоремы сделаем ряд упрощений. Заметим, что если g — обратимое линейное преобразование, то $g \circ f$

линейно тогда и только тогда, когда f линейно. Пусть вектора $f\begin{pmatrix}1\\0\end{pmatrix}$ и $f\begin{pmatrix}0\\1\end{pmatrix}$ не лежат на одной прямой, проходящей через начало координат, т. е. они линейно независимы, и, следовательно, мы можем построить линейное преобразование g, обеспечивающее равенства $g \circ f\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix}$ и $g \circ f\left(\begin{pmatrix}0\\1\end{pmatrix}\right) = \begin{pmatrix}0\\1\end{pmatrix}$. Таким образом, с учетом замены f на $g \circ f$ нам достаточно доказать следующее:

> *Пусть f — аффинное преобразование, удовлетворяющее условиям $f\left(\begin{pmatrix}0\\0\end{pmatrix}\right) = \begin{pmatrix}0\\0\end{pmatrix}$, $f\left(\begin{pmatrix}1\\0\end{pmatrix}\right) = \begin{pmatrix}1\\0\end{pmatrix}$ и $f\left(\begin{pmatrix}0\\1\end{pmatrix}\right) = \begin{pmatrix}0\\1\end{pmatrix}$. Тогда f — тождественное преобразование.*

Доказательство. Из параграфа 1.5 мы знаем, что $f\left(\begin{pmatrix}1\\1\end{pmatrix}\right) = f\left(\begin{pmatrix}1\\0\end{pmatrix}\right)+f\left(\begin{pmatrix}0\\1\end{pmatrix}\right) = \begin{pmatrix}1\\0\end{pmatrix}+\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}$. (По существу, мы доказали, что $f\left(\begin{pmatrix}r\\s\end{pmatrix}\right) = \begin{pmatrix}r\\s\end{pmatrix}$ для любых рациональных чисел r и s.)

Таким образом, f переводит оси x, y и прямую $x = y$, проходящую через точки $\begin{pmatrix}0\\0\end{pmatrix}$ и $\begin{pmatrix}1\\1\end{pmatrix}$, в самих себя.

Итак, для любого вещественного числа a

$$f\left(\begin{pmatrix}a\\0\end{pmatrix}\right) = \begin{pmatrix}\phi(a)\\0\end{pmatrix},$$

где ϕ — некоторая функция. (Мы хотим доказать, что $\phi(a) = a$ для всех a.) Аналогично $f\left(\begin{pmatrix}0\\b\end{pmatrix}\right) = \begin{pmatrix}0\\\psi(b)\end{pmatrix}$ для некоторой функции ψ. Поскольку

$$f\left(\begin{pmatrix}a\\b\end{pmatrix}\right) = f\left(\begin{pmatrix}a\\0\end{pmatrix}\right) + f\left(\begin{pmatrix}0\\b\end{pmatrix}\right),$$

мы имеем

$$f\left(\begin{pmatrix}a\\b\end{pmatrix}\right) = \begin{pmatrix}\phi(a)\\\psi(b)\end{pmatrix}.$$

Мы утверждаем, что ψ и ϕ — одна и та же функция. Действительно, возьмем прямую $x = a$ (рис. 1.25). Она параллельна оси y и, следовательно, ее образ при преобразовании f должен быть параллелен оси y, т. е. образ должен быть $x = \phi(a)$. Прямая $x = a$ пересекает прямую $x = y$ в точке $\begin{pmatrix} a \\ a \end{pmatrix}$, а прямая $x = \phi(a)$ пересекает прямую $x = y$ в точке $\begin{pmatrix} \phi(a) \\ \phi(a) \end{pmatrix}$. Следовательно,

$$f\left(\begin{pmatrix} a \\ a \end{pmatrix}\right) = \begin{pmatrix} \phi(a) \\ \phi(a) \end{pmatrix},$$

т. е. $\phi(a) = \psi(a)$ для всех a.

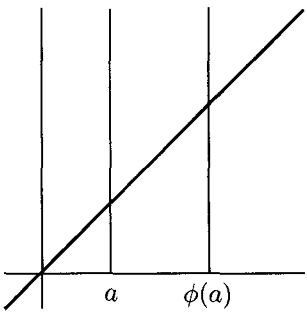


Рис. 1.25

Мы знаем, что $\left(\begin{pmatrix} a+b \\ 0 \end{pmatrix}\right) = \left(\begin{pmatrix} a \\ 0 \end{pmatrix}\right) + \left(\begin{pmatrix} b \\ 0 \end{pmatrix}\right)$, т. е.

$$\phi(a+b) = \phi(a) + \phi(b).$$

Воспользуемся аргументами из параграфа 1.2 и докажем удивительное равенство

$$\phi(ab) = \phi(a)\phi(b).$$

Действительно, давайте посмотрим на рис. 1.26. Прямая, соединяющая точки $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ и $\begin{pmatrix} b \\ 0 \end{pmatrix}$, параллельна прямой, соединяющей точки $\begin{pmatrix} a \\ a \end{pmatrix}$ и $\begin{pmatrix} ab \\ 0 \end{pmatrix}$. Таким образом, значение ab может быть

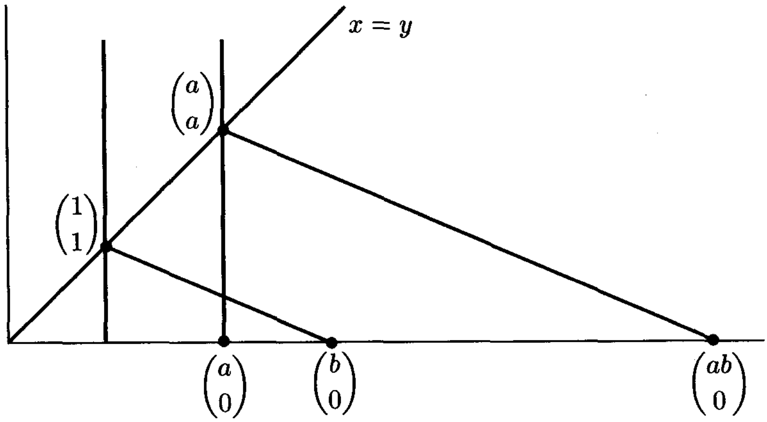


Рис. 1.26

получено графически с помощью параллельных и пересекающихся прямых. Поэтому, рисуя аналогичную диаграмму для $\begin{pmatrix} \phi(a) \\ 0 \end{pmatrix}$ и $\begin{pmatrix} \phi(b) \\ 0 \end{pmatrix}$, мы получим

$$\begin{pmatrix} \phi(ab) \\ 0 \end{pmatrix} = f\begin{pmatrix} ab \\ 0 \end{pmatrix} = \begin{pmatrix} \phi(a)\phi(b) \\ 0 \end{pmatrix},$$

откуда следует, что

$$\phi(ab) = \phi(a)\phi(b).$$

Вещественное число x положительно тогда и только тогда, когда $x = y^2$ при некотором положительном y. Поэтому $\phi(x) = \phi(y^2) = \phi(y)^2$, и из неравенства $x > 0$ следует, что $\phi(x) > 0$. Итак, из $a - b > 0$ вытекает, что $\phi(a) - \phi(b) > 0$. И если $r < a < s$, то $\phi(r) < \phi(a) < \phi(s)$. Для любого вещественного числа a мы можем найти рациональные числа r и s, удовлетворяющие условию $r < a < s$, сделав величину $s - r$ сколь угодно малой. Однако, для рациональных чисел $\phi(r) = r$ и $\phi(s) = s$. Отсюда

$$r < \phi(a) < s.$$

Следовательно, $|a - \phi(a)| < s - r$. Поскольку число $s - r$ может быть выбрано сколь угодно малым, то для всех вещественных чисел выполняется $\phi(a) = a$. Этим заканчивается наше доказательство.

Резюме

A. Преобразования плоскости

Прочитав эту главу, вы должны усвоить определения аффинного преобразования, линейного преобразования и евклидова преобразования.

Вы должны уметь определять геометрические свойства, которые сохраняются при аффинных преобразованиях, и свойства, которые сохраняются при евклидовых преобразованиях.

B. Алгебра матриц

Вы должны уметь складывать и умножать квадратные матрицы одинакового ранга, вычислять определители матриц 2×2 и писать обратную матрицу для обратимой матрицы 2×2.

C. Матрицы и линейные преобразования

Имея достаточную информацию о линейном преобразовании плоскости, вы должны уметь писать матрицу 2×2, определяющую это преобразование.

Вы должны понимать связь между умножением матриц и композицией линейных преобразований и уметь пользоваться этой связью.

Вы должны уметь определять матрицы 2×2, которые представляют частные случаи преобразований (преобразования вращения, отражения, проекции, нильпотентные преобразования).

Задачи

1.1. Ниже приводятся несколько теорем евклидовой геометрии плоскости. Разберитесь, какие из них справедливы для *аффинной* геометрии плоскости.

(a) Все медианы треугольника пересекаются в одной точке и делятся этой точкой в отношении $2 : 1$ (считая от вершины угла).

(b) Длины биссектрис равнобедренного треугольника равны.

(c) Диагонали ромба взаимно перпендикулярны.

(d) Диагонали параллелограмма в точке пересечения делятся пополам.

(e) Пусть PQR и $P'Q'R'$ — два таких треугольника, что параллельны следующие три пары прямых: PQ и $P'Q'$, QR и $Q'R'$, PR и $P'R'$. Тогда три прямые PP', QQ' и RR' либо параллельны друг другу, либо пересекаются в одной точке.

1.2. (a) Пусть $\mathbb{A}_1$ и $\mathbb{A}_2$ — аффинные прямые, x — аффинная координата на $\mathbb{A}_1$, а y — аффинная координата на $\mathbb{A}_2$. Далее, пусть $f : \mathbb{A}_1 \to \mathbb{A}_2$ — *аффинное* отображение, а соответствующая ему функция $F : \mathbb{R} \to \mathbb{R}$ такова, что если $Q = f(P)$, то $y(Q) = F \circ x(P)$. Покажите, что наиболее общий вид функции F есть $F(\alpha) = r\alpha + s$.

(b) Пусть $x' = ax + b$, $y' = cy + d$, где x' и y' — новые аффинные координаты на $\mathbb{A}_1$ и $\mathbb{A}_2$ соответственно, и пусть $y(Q) = F \circ x(P)$, где $F(\alpha) = r\alpha + s$. Получите формулу для такой функции $F'(\beta)$, чтобы $y'(Q) = F' \circ x'(P)$.

1.3. Функцию $u : \mathbb{R}^2 \to \mathbb{R}$ назовем *аффинной*, если она аффинна для каждой прямой в плоскости и для любого параллелограмма, изображенного на рис. 1.27, выполняется равенство $u(P)+u(R) = u(Q) + u(S)$. Пусть $u : \mathbb{R}^2 \to \mathbb{R}$ — аффинная функция и известны следующие ее значения: $u\begin{bmatrix}1\\2\end{bmatrix} = 3$, $u\begin{bmatrix}3\\8\end{bmatrix} = 8$, $u\begin{bmatrix}2\\-1\end{bmatrix} = 9$.

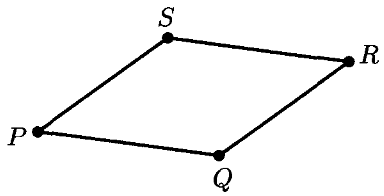


Рис. 1.27

(a) Получить формулу для $u\begin{bmatrix}x\\y\end{bmatrix}$.

(b) Нарисовать $\mathbb{R}^2$ и показать прямые $u = \text{const}$.

1.4. Для линейного преобразования, представленного каждой из матриц, следующих ниже, получить образ прямоугольника $ABDE$, изображенного на рис. 1.28. Вычислить определители преобразований и проверить, что они правильно предсказывают площади и ориентации образов прямоугольника.

(a) Вращение $R_{\pi/2} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

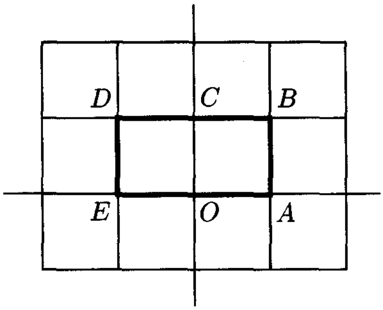


Рис. 1.28

(b) Вращение $R_{\pi/4} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$.

(c) «Искажение» $D_2 = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}$.

(d) Преобразование Лоренца $L_2 = \begin{pmatrix} 5/4 & 3/4 \\ 3/4 & 5/4 \end{pmatrix}$.

(e) Сдвиг $S_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

(f) Сдвиг $S_1' = \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 3/2 \end{pmatrix}$.

(g) Отражение $M_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

(h) Отражение $M_{\pi/4} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

(i) Проекция $P_{\pi/4} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$.

(j) Нильпотентное преобразование $N_0 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$.

(k) Нильпотентное преобразование $N_{\pi/4} = \begin{pmatrix} -1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix}$.

1.5. Алгебраически вычислите произведения матриц, определенных в предыдущем упражнении, и дайте геометрическую интерпретацию полученных преобразований:

(a) $R_{\pi/4}R_{\pi/2}$,

(b) S_1^2,

(c) $R_{\pi/2}M_0$,

(d) $P^2_{\pi/4}$,

(e) $N^2_{\pi/4}$.

1.6. Для каждой из следующих матриц вычислите обратную матрицу и дайте геометрическую интерпретацию полученного результата:

(a) $R_{\pi/4} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$,

(b) $L_2 = \begin{pmatrix} 5/4 & 3/4 \\ 3/4 & 5/4 \end{pmatrix}$,

(c) $S_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$,

(d) $M_{\pi/4} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

1.7. Для каждой из следующих матриц, имеющих нулевой определитель, найдите образ (Im) и ядро (Ker):

(a) $P_{\pi/4} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$,

(b) $N_{\pi/4} = \begin{pmatrix} -1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix}$,

(c) $A = \begin{pmatrix} 1 & 2 \\ -1 & -2 \end{pmatrix}$.

1.8. В параграфе 1.7 доказано разложение матрицы в произведение трех множителей. Представьте матрицу

$$\begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix}$$

в виде

$$\begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix}\begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix}\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}.$$

1.9. Найдите процедуру, представляющую любую матрицу $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ при $c \neq 0$ в виде произведения

$$\begin{pmatrix} 1 & y \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & f \\ e & 0 \end{pmatrix}\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}.$$

С помощью этой процедуры разложите матрицу $\begin{pmatrix} -12 & 26 \\ -4 & -8 \end{pmatrix}$ в произведение трех множителей.

1.10. Докажите, что $\mathrm{Ar}\begin{pmatrix} 0 & b \\ c & d \end{pmatrix} = \left|\mathrm{Det}\begin{pmatrix} 0 & b \\ c & d \end{pmatrix}\right|$:

(a) путем прямой проверки (найдите образ квадрата);

(b) используя разложение из предыдущего упражнения, которое справедливо даже при $a = 0$.

1.11. Постройте матрицы 2×2, которые представляют следующие преобразования плоскости:

(a) преобразование P, удовлетворяющее условию $P^2 = P$, которое всю плоскость отображает на прямую $y = 2x$, а прямую $y = -2x$ отображает в начало координат;

(b) преобразование сдвига S, которое каждую точку прямой $y = 2x$ отображает в себя, ось y отображает в прямую $y = -x$ и удовлетворяет условию $(S - I)^2 = 0$;

(c) преобразование, переводящее $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ в $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$, а $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ в $\begin{pmatrix} 1/2 \\ -1 \end{pmatrix}$;

(d) нильпотентное преобразование N, удовлетворяющее условию $N^2 = 0$, для которого прямая $y = 3x$ является одновременно образом и ядром.

1.12. Упражнение на умножение матриц 3×3. Возьмите две матрицы

$$U = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad L = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

и вычислите UL, LU, U^3.

1.13. Известно, что определитель матрицы 3×3 равен:

$$\begin{aligned}
&\mathrm{Det}\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \\
&= a_{11}\mathrm{Det}\begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{12}\mathrm{Det}\begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} + a_{13}\mathrm{Det}\begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \\
&= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} \\
&\quad + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.
\end{aligned}$$

Докажите, что

$$\operatorname{Det}(F \circ G) = \operatorname{Det} F \times \operatorname{Det} G.$$

1.14. Покажите, что если матрица

$$F = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

удовлетворяет условиям $a_{11} \neq 0$ и $a_{11}a_{22} - a_{12}a_{21} \neq 0$, то ее можно записать в виде произведения трех матриц

$$F = \begin{pmatrix} 1 & 0 & 0 \\ y_{21} & 1 & 0 \\ y_{31} & y_{32} & 1 \end{pmatrix} \begin{pmatrix} e & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & g \end{pmatrix} \begin{pmatrix} 1 & x_{12} & x_{13} \\ 0 & 1 & x_{23} \\ 0 & 0 & 1 \end{pmatrix}.$$

1.15. В пространстве $\mathbb{R}^3$ мы можем определить $\operatorname{Vol} F$ для любого несингулярного линейного преобразования F так же, как мы определяли Ar на плоскости. Таким образом,

$$\operatorname{Vol} F = \frac{\text{объем } F(D)}{\text{объем } D}$$

для любой области D. Кроме того,

$$\operatorname{Vol}(F \circ G) = \operatorname{Vol} F \times \operatorname{Vol} G$$

и $\operatorname{Vol} F = \text{объем } F(\square)$, где $\square$ — единичный куб. Докажите, что

$$\operatorname{Vol} F = |\operatorname{Det} F|.$$

1.16. Рассмотрите аффинное преобразование, которое *не* оставляет начало координат на месте:

$$F\begin{pmatrix} x \\ y \end{pmatrix} = A\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix},$$

где A — аффинное преобразование, оставляющее начало координат на месте. Покажите, что такое преобразование может быть представлено матрицей 3×3:

$$\left(\begin{array}{cc|c} \multicolumn{2}{c|}{\multirow{2}{*}{A}} & a \\ & & b \\ \hline 0 & 0 & 1 \end{array}\right),$$

где A — матрица 2×2, при условии, что вектор на плоскости $\begin{pmatrix} x \\ y \end{pmatrix}$ представляется трехкомпонентным вектором $\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$. Докажите следующие утверждения.

(a) Такая матрица 3×3 имеет определитель, равный $\operatorname{Det} A$.

(b) Когда такая матрица 3×3 действует на вектор $\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$, третья компонента результирующего вектора равна 1.

(c) Матрицы $T(a, b) \equiv \begin{pmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}$, представляющие чистые трансляции, удовлетворяют закону

$$T(a, b)T(c, d) = T(a + c, b + d).$$

Это упражнение может быть интерпретировано геометрически. Рассмотрим аффинную плоскость как плоскость $z = 1$ в пространстве $\mathbb{R}^3$. Группу аффинных преобразований этой плоскости отождествим с группой линейных преобразований в пространстве $\mathbb{R}^3$. Точку $\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$ на плоскости $z = 1$ отождествим с прямой, соединяющей эту точку с началом координат. Любая точка на плоскости определяет единственную такую прямую. Если прямая, проведенная из начальной точки, пересекает плоскость $z = 1$, то она пересекает ее в единственной точке (рис. 1.29). Таким образом, мы можем говорить, что каждой точке на нашей аффинной плоскости соответствует прямая в пространстве $\mathbb{R}^3$, выходящая из начальной точки и пересекающая плоскость $z = 1$. Такая геометрическая интерпретация позволяет нам подойти к определению понятия перспективы: две плоские фигуры в пространстве $\mathbb{R}^3$ (не содержащие начало) перспективно совпадают, если они определяют одно и то же семейство прямых, выходящих из начала координат (рис. 1.30).

Это позволяет нам рассмотреть новую геометрию, в которой «точки» — это прямые в пространстве $\mathbb{R}^3$, проходящие через начало; мы не требуем, чтобы прямая пересекала плоскость $z = 1$.

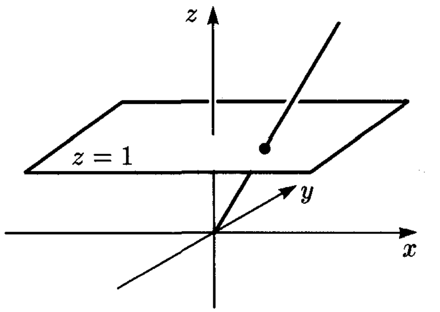


Рис. 1.29

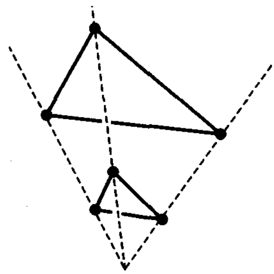


Рис. 1.30

Новые «точки», которые мы добавили, соответствуют тем прямым в пространстве $\mathbb{R}^3$, которые лежат в плоскости $z = 0$ и проходят через начальную точку. Пусть $\begin{pmatrix} a \\ b \\ 0 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ — точка пространства $\mathbb{R}^3$, лежащая в плоскости $z = 0$. Буквой P обозначим прямую, проходящую через начальную точку и точку $\begin{pmatrix} a \\ b \\ 0 \end{pmatrix}$, т. е. P — это одна из наших «новых» точек. Таким образом, $P = \left\{ \begin{pmatrix} at \\ bt \\ 0 \end{pmatrix}_{t \in \mathbb{R}} \right\}$. С точки зрения пространства $\mathbb{R}^3$, где P является прямой, мы можем аппроксимировать P семейством прямых P_ϵ, проходящих через начальную точку (рис. 1.31) и задаваемых формулой

$$P_\epsilon = \left\{ \begin{pmatrix} at \\ bt \\ \epsilon t \end{pmatrix}_{t \in \mathbb{R}} \right\}.$$

В некотором смысле $P_\epsilon \to P$ при $\epsilon \to 0$. Но P_ϵ пересекает плоскость $z = 1$ при $t = 1/\epsilon$ в точке $\begin{pmatrix} a/\epsilon \\ b/\epsilon \\ 1 \end{pmatrix}$.

Точки прямых P_ϵ, принадлежащие аффинной плоскости, стремятся к бесконечности при $\epsilon \to 0$ в определенном направлении,

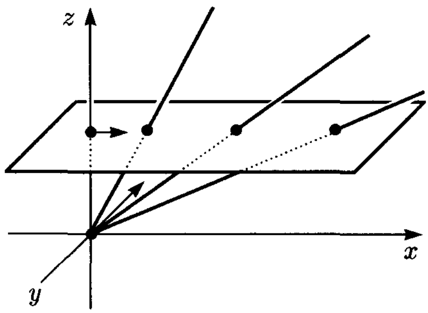


Рис. 1.31

задаваемым вектором $\begin{pmatrix} a \\ b \end{pmatrix}$. Поэтому будем считать, что новая «точка» P — это точка, лежащая на бесконечности в аффинной плоскости. Понятие о таких бесконечно удаленных точках было впервые введено в пятнадцатом и шестнадцатом веках художниками и геометрами, изучавшими перспективу.

Итак, мы определили новое пространство $\mathbb{P}^2$, называемое *проективной плоскостью.* «Точка» в пространстве $\mathbb{P}^2$ — это прямая, проведенная из начальной точки в пространстве $\mathbb{R}^3$. Давайте теперь посмотрим, как определить «прямую» в $\mathbb{P}^2$. С точки зрения пространства $\mathbb{R}^3$ «прямая» на аффинной плоскости $z = 1$ состоит из прямых, проходящих через начало координат и пересекающих плоскость $z = 1$ во всевозможных точках некоторой прямой (рис. 1.32).

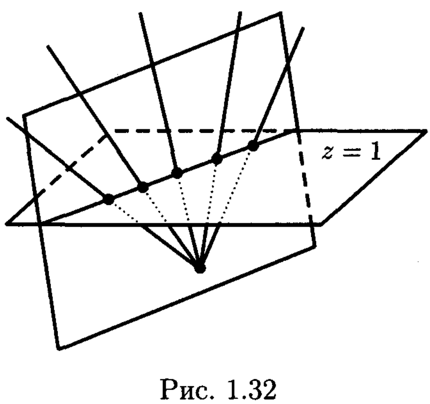


Рис. 1.32

Такое семейство прямых образует плоскость в пространстве $\mathbb{R}^3$. Другими словами, с точки зрения пространства $\mathbb{R}^3$, прямая на аффинной плоскости есть линия пересечения плоскости в пространстве $\mathbb{R}^3$, проходящей через начальную точку, и плоскости $z = 1$. Теперь ясно, что делать дальше. Мы опускаем условие пересечения плоскостей и определяем «прямую» на $\mathbb{P}^2$ как плоскость в пространстве $\mathbb{R}^3$, проходящую через начальную точку. Существует только одна плоскость в пространстве $\mathbb{R}^3$, проходящая через начало и не пересекающая плоскость $z = 1$. Это плоскость $z = 0$. Нам осталось добавить к множеству прямых на аффинной плоскости «бесконечно удаленную прямую». Точка P принадлежит «прямой» l, если прямая, проведенная из начала, лежит в плоскости, проходящей через это начало. Две разные «точки» P и Q (т. е. две разные прямые, проведенные через начало) определяют единственную плоскость, проходящую через начало, т. е. две различные «точки» определяют единственную «прямую». Две различные плоскости, проходящие через начало в пространстве $\mathbb{R}^3$, пересекаются по прямой, проходящей через начало. Таким образом, две разные «прямые» на плоскости $\mathbb{P}^2$ всегда пересекаются в «точке». (Заметим, что здесь проявляется различие с аффинной геометрией, где две прямые могут быть параллельны. Две параллельные прямые, принадлежащие аффинной плоскости, пересекаются «на бесконечности» в проективной плоскости.)

Выводы: любая «точка» на плоскости $\mathbb{P}^2$ — это прямая, проходящая через начало в пространстве $\mathbb{R}^3$; любая «прямая» на плоскости $\mathbb{P}^2$ — это плоскость, проходящая через начало в пространстве $\mathbb{R}^3$; две различные «прямые» пересекаются в единственной «точке».

Любая обратимая матрица 3×3, действуя в пространстве $\mathbb{R}^3$, переводит прямые, проходящие через начало, в прямые, тоже проходящие через начало; переводит плоскости, содержащие начало, в плоскости, тоже содержащие начало.

1.17. (а) Покажите, что любая обратимая матрица 3×3 определяет взаимно-однозначное преобразование проективной плоскости $\mathbb{P}^2$, переводящее «прямые» в «прямые».

(b) Покажите, что две обратимые матрицы A и B третьего порядка определяют одно и то же преобразование плоскости $\mathbb{P}^2$ тогда и только тогда, когда $A = cB$ для вещественного числа c, не равного нулю.

1.18. Три вектора $\mathbf{u}$, $\mathbf{v}$ и $\mathbf{w}$ в пространстве $\mathbb{R}^3$ называются линейно независимыми, если уравнение

$$a\mathbf{u} + b\mathbf{v} + c\mathbf{w} = \mathbf{0}$$

не может выполняться, если только не все числа a, b, c равны нулю. Покажите, что если $\mathbf{u}$, $\mathbf{v}$ и $\mathbf{w}$ линейно независимы, то существует единственная обратимая матрица A третьего порядка, для которой

$$A\mathbf{u} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad A\mathbf{v} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad A\mathbf{w} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

(Обобщение этой теоремы для любого конечномерного векторного пространства будет доказано позже.)

1.19. (a) Пусть P_1, P_2, P_3 — различные «точки» на плоскости $\mathbb{P}^2$, определяемые прямыми, проходящими через начальную точку, и векторами $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ и $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ соответственно. Пусть Q_1, Q_2, Q_3 — три любые другие различные «точки» $\mathbb{P}^2$. Покажите, что существует обратимая матрица 3×3, которая преобразует точку Q_1 в P_1, Q_2 в P_2 и Q_3 в P_3.

(b) Пусть Q_1, Q_2, Q_3, Q_4 — «точки» на плоскости $\mathbb{P}^2$, из которых никакие три не лежат на одной прямой. Пусть R_1, R_2, R_3, R_4 — другой набор четырех «точек», из которых никакие три не лежат на одной прямой. Покажите, что существует такая матрица 3×3, которая преобразует Q_1 в P_1, Q_2 в P_2, Q_3 в P_3 и Q_4 в P_4.

(c) Докажите фундаментальную теорему проективной геометрии: любое взаимно-однозначное преобразование плоскости $\mathbb{P}^2$, переводящее «прямые» в «прямые», осуществляется матрицей 3×3. (Подсказка: сведите все к фундаментальной теореме аффинной геометрии, доказанной в приложении к этой главе.)

1.20. В качестве иллюстрации применения упражнения 1.19(b) докажите теорему Фано (рис. 1.33). Пусть A, B, C, D — четыре точки, из которых никакие три не лежат на одной прямой, и пусть P — точка пересечения прямых AB и CD, Q — точка пересечения AC и BD, R — точка пересечения AD и BC. Тогда P, Q и R *не* лежат на одной прямой. (Подсказка: рассмотрите частный случай, например, пусть A, B, C — три вершины равностороннего треугольника, а D — его центр.)

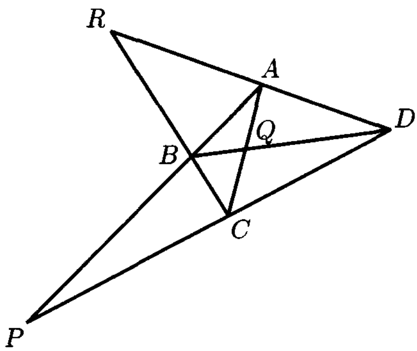


Рис. 1.33

Глава 2

Собственные векторы и собственные значения

Во второй главе мы обсудим конформную линейную геометрию на плоскости, т. е. геометрию прямых, углов и их связь с матрицами 2×2 определенного вида. Обсудим очень важные для квантовой механики понятия собственных векторов и собственных значений[1]. С помощью этих понятий получим алгоритм вычисления степеней матрицы. В качестве приложения изучим основные свойства цепей Маркова.

2.1. Конформные линейные преобразования

Мы рассмотрим такие линейные преобразования f пространства $\mathbb{R}^2$, которые

(1) сохраняют угол,

(2) сохраняют ориентацию, т. е. $\operatorname{Det} F > 0$,

где F — матрица, представляющая f. Заметим, что если f и g — два таких преобразования, то их композиция $g \circ f$ обладает теми же свойствами.

---

[1]Конечно, эти понятия не менее важны в самых различных разделах математики и ее приложениях. — *Прим. ред.*

Допустим, что f сохраняет угол и ориентацию (рис. 2.1). Мы можем найти такой поворот $r_{-\theta}$, что $r_{-\theta} \circ f$ переводит $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ в точку на положительной оси x. Но раз преобразование $r_{-\theta} \circ f$ сохраняет углы и $\operatorname{Det}(r_{-\theta} \circ f) > 0$, то $r_{-\theta} \circ f\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$ находится на положительной оси y.

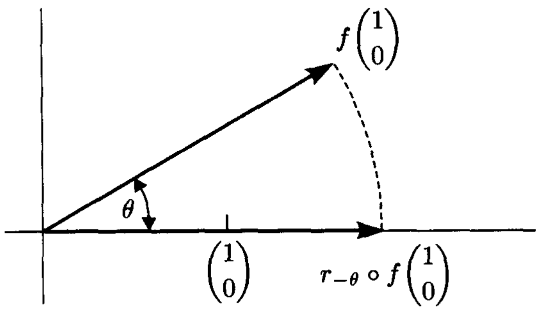


Рис. 2.1

Таким образом, матрица, представляющая $r_{-\theta} \circ f$, имеет вид:

$$r_{-\theta} \circ f = \begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix},$$

где $r > 0$, $s > 0$. Так как $r_{-\theta} \circ f$ сохраняет углы, прямая, идущая вдоль вектора $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, переходит в себя. Если мы говорим, что $\begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ лежит на прямой, заданной вектором $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, это значит, что $r = s$. Следовательно,

$$r_{-\theta} \circ f = \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix}.$$

Поэтому матрица, представляющая f, имеет вид

$$F = \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} r\cos\theta & -r\sin\theta \\ r\sin\theta & r\cos\theta \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

где $a = r\cos\theta$, $b = r\sin\theta$. Ясно, что любая такая матрица удовлетворяет условию $\operatorname{Det} F = a^2 + b^2 = r^2 > 0$.

Обратное утверждение: любая ненулевая матрица $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ сохраняет угол и ориентацию, потому что взяв

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \quad a^2 + b^2 \neq 0,$$

можно положить $r^2 = a^2 + b^2$ и потом найти такой угол θ, при котором

$$\cos\theta = ar^{-1}, \qquad \sin\theta = br^{-1},$$

поскольку $a \leq r$ и $\sin^2\theta + \cos^2\theta = (a^2+b^2)/r^2 = 1$. Следовательно, матрица принимает вид

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} r\cos\theta & -r\sin\theta \\ r\sin\theta & r\cos\theta \end{pmatrix}.$$

Итак, в общем случае матрица

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \quad \text{где} \quad a^2 + b^2 \neq 0,$$

сохраняет угол и ориентацию, причем

$$\operatorname{Det}\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = a^2 + b^2.$$

Очевидно, что произведение двух таких матриц дает матрицу такого же вида. Кроме того, заметим, что

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}\begin{pmatrix} a' & -b' \\ b' & a' \end{pmatrix} = \begin{pmatrix} aa' - bb' & -(ba' + ab') \\ ba' + ab' & aa' - bb' \end{pmatrix} = \begin{pmatrix} a' & -b' \\ b' & a' \end{pmatrix}\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

т. е. в этом случае произведение матриц коммутативно. Более того, при ненулевых значениях a и b существует обратная матрица (т. к. определитель матрицы равен a^2+b^2). Наконец, если мы сложим две матрицы

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} + \begin{pmatrix} a' & -b' \\ b' & a' \end{pmatrix} = \begin{pmatrix} a + a' & -(b + b') \\ b + b' & a + a' \end{pmatrix},$$

то *сумма* будет опять матрицей типа $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. Это замечательное свойство не следует прямо из определения. Матрицу типа $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ будем называть *конформной.* (При этом мы допускаем возможность, что $a = b = 0$. Таким образом, ненулевыми конформными матрицами называются такие матрицы, которые сохраняют угол и ориентацию.)

Итак, мы доказали, что множество всех конформных матриц замкнуто относительно операций сложения и умножения. Для таких матриц умножение является коммутативной операцией. Каждая ненулевая конформная матрица имеет обратную. Это значит, что поведение конформных матриц во многом похоже на поведение чисел.

Любую конформную матрицу можно записать в виде

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = a\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Заметим, что $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ есть поворот на девяносто градусов и что

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Введем обозначения:

$$\mathbb{I} \quad \text{для} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

и

$$\mathrm{i} \quad \text{для} \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Тогда

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = a\mathbb{I} + b\mathrm{i},$$

где

$$\mathrm{i}^2 = -\mathbb{I}.$$

Другими словами, множество конформных матриц можно отождествлять с множеством комплексных чисел.

Обычное представление комплексного числа в виде точки на плоскости есть просто отождествление комплексного числа с образом точки $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ при преобразовании, задаваемом матрицей $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. Для конформных матриц точка $\begin{pmatrix} a \\ b \end{pmatrix}$ определяет матрицу $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$.

Легко вычислить n-ю степень конформной матрицы. Действительно, если мы запишем

$$A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

то, учитывая, что $\begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix}$ коммутирует со всеми матрицами 2×2, получим

$$A^n = \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix}^n \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}^n = \begin{pmatrix} r^n & 0 \\ 0 & r^n \end{pmatrix} \begin{pmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{pmatrix}.$$

Итак,

$$\begin{pmatrix} r\cos\theta & -r\sin\theta \\ r\sin\theta & r\cos\theta \end{pmatrix}^n = \begin{pmatrix} r^n\cos n\theta & -r^n\sin n\theta \\ r^n\sin n\theta & r^n\cos n\theta \end{pmatrix}.$$

На языке комплексных чисел это значит, что если

$$z = r(\cos\theta + \mathrm{i}\sin\theta),$$

то

$$z^n = r^n(\cos n\theta + \mathrm{i}\sin n\theta);$$

это равенство известно как теорема Муавра.

Другой способ вычисления матрицы A^n состоит в применении биномиальной формулы. Учитывая, что

$$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} = a\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} 0 & -b \\ b & 0 \end{pmatrix} = b\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

вычисляем

$$A^n = \left(a\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} + b\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\right)^n$$

$$= a^n\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} + na^{n-1}b\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix} + \frac{1}{2}n(n-1)a^{n-2}b^2\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}^2 + \cdots$$

Но $\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}^2 = \begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix}$, поэтому

$$A^n = \left(a^n - \binom{n}{2}a^{n-2}b^2 + \binom{n}{4}a^{n-4}b^4 + \cdots\right)\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$
$$+ \left(\binom{n}{1}a^{n-1}b - \binom{n}{3}a^{n-3}b^3 + \cdots\right)\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}.$$

В следующем параграфе мы покажем процедуру вычисления степеней произвольной матрицы 2×2 (не обязательно конформной). Для этого потребуется понятие собственных значений, играющее ключевую роль в квантовой механике.

2.2. Собственные векторы и собственные значения

Пусть F — линейное преобразование. Зададим вопрос: может ли оно перевести прямую, проходящую через начальную точку, в себя? (Ни один нетривиальный поворот не обладает таким свойством, а вот любое ненулевое сингулярное преобразование переводит свой образ в себя.)

Если $\mathbf{v}$ — ненулевой вектор, лежащий на такой прямой, то должно выполняться равенство

$$F(\mathbf{v}) = \lambda\mathbf{v},$$

где λ — вещественное число. Если это уравнение выполняется для некоторого $\mathbf{v} \neq \mathbf{0}$, то λ называется *собственным значением* преобразования f, соответствующим *собственному вектору* $\mathbf{v}$. Последнее уравнение можно переписать в виде

$$\left[F - \lambda\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right]\mathbf{v} = \mathbf{0}.$$

Так как вектор $\mathbf{v}$ не равен нулю, мы получаем равенство

$$\operatorname{Det}\left(F-\lambda\begin{pmatrix}1&0\\0&1\end{pmatrix}\right)=0,$$

представляющее собой уравнение относительно λ.

Если $F=\begin{pmatrix}a&b\\c&d\end{pmatrix}$, то $F-\lambda\begin{pmatrix}1&0\\0&1\end{pmatrix}=\begin{pmatrix}a-\lambda&b\\c&d-\lambda\end{pmatrix}$,

и предыдущее уравнение можно записать так:

$$\operatorname{Det}\begin{pmatrix}a-\lambda&b\\c&d-\lambda\end{pmatrix}=(a-\lambda)(d-\lambda)-bc=\lambda^2-(a+d)\lambda+(ad-bc)=0,$$

или

$$\boxed{\lambda^2-(\operatorname{Tr}F)\lambda+\operatorname{Det}F=0.}$$

Полином

$$P(X)=X^2-(a+d)X+(ad-bc)$$

называется *характеристическим полиномом* матрицы $\begin{pmatrix}a&b\\c&d\end{pmatrix}$, а уравнение

$$P(\lambda)=0$$

называется *характеристическим уравнением.* Это уравнение будет иметь вещественные корни

$$\lambda=\frac{1}{2}[(a+d)\pm\sqrt{\{(a+d)^2-4(ad-bc)\}}]$$
$$=\frac{1}{2}[(a+d)\pm\sqrt{\{(a-d)^2+4bc\}}]$$

тогда и только тогда, когда

$$(a-d)^2+4bc\geq 0.$$

В этом случае мы можем проверить, что

$$\begin{pmatrix}a-\lambda&b\\c&d-\lambda\end{pmatrix}\begin{pmatrix}d-\lambda\\-c\end{pmatrix}=0$$

и

$$\begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} \begin{pmatrix} b \\ -(a-\lambda) \end{pmatrix} = 0.$$

Поэтому, если $\begin{pmatrix} d-\lambda \\ -c \end{pmatrix}$ или $\begin{pmatrix} b \\ -(a-\lambda) \end{pmatrix}$ не равны нулю, то они являются собственными векторами; они лежат на одной прямой, поскольку

$$\operatorname{Det} \begin{pmatrix} d-\lambda & b \\ -c & -(a-\lambda) \end{pmatrix} = -\operatorname{Det} \begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} = 0.$$

Если же они оба равны нулю, то $a = \lambda$, $b = c = 0$ и $d = \lambda$, поэтому

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}, \qquad F = \lambda\mathbb{I},$$

и каждый (ненулевой) вектор на плоскости является собственным вектором.

Рассмотрим возможные варианты собственных значений линейных преобразований.

Случай 1. Характеристическое уравнение имеет различные вещественные корни. В случае, когда $(a-d)^2 + 4bc > 0$, существуют два *различных* вещественных собственных значения λ_1 и λ_2. Тогда $F \neq \lambda\mathbb{I}$ и, следовательно, есть только две прямые, проходящие через начало координат и не меняющиеся при таком преобразовании, каждая из которых задается собственным вектором $\mathbf{v}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ или $\mathbf{v}_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$. Мы можем записать

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \lambda_1 \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$

и

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \lambda_2 \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}.$$

Возьмем матрицу

$$B = \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}.$$

Она несингулярна, потому что векторы $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ и $\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ принадлежат разным прямым. Тогда уравнение для собственных векторов можно записать в виде

$$FB = B\Lambda,$$

где $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$, или, окончательно,

$$\boxed{F = B\Lambda B^{-1}, \qquad \Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.}$$

Обратное утверждение: если $F = B\Lambda B^{-1}$ при $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$, то

$$F\left[B\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right] = B\Lambda B^{-1}B\begin{pmatrix} 1 \\ 0 \end{pmatrix} = B\Lambda\begin{pmatrix} 1 \\ 0 \end{pmatrix} = B\lambda_1\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \lambda_1 B\begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

так что первый столбец матрицы B является собственным вектором преобразования F, соответствующим собственному значению λ_1, и второй столбец — собственный вектор для собственного значения λ_2.

Случай 2. Характеристическое уравнение имеет кратный вещественный корень. В случае, когда $(a-d)^2 + 4bc = 0$ и полином $P(X) = 0$ имеет два одинаковых корня λ, ситуация немного сложнее. Рассмотрим две матрицы

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Для этих матриц характеристический полином $P(X) = X^2$ имеет два равных корня $\lambda = 0$. Каждый ненулевой вектор на плоскости является собственным вектором первой матрицы, и только векторы $\begin{pmatrix} x \\ 0 \end{pmatrix}$ будут собственными векторами второй. Заметим, однако, что обе матрицы удовлетворяют уравнению $F^2 = 0$, которое можно записать как $P(F) = 0$, т. е. мы заменяем F

(как будто это число) на его характеристический полином и получаем ноль. В общем случае это утверждение называется теоремой Гамильтона–Кэли.

Если $P(X)$ — характеристический полином матрицы F, то

$$P(F) = 0,$$

т. е.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 - (a+d)\begin{pmatrix} a & b \\ c & d \end{pmatrix} + (ad-bc)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 0.$$

В случае матриц 2×2 эта теорема проверяется вычислением. Действительно,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 = \begin{pmatrix} a^2+bc & ab+bd \\ ca+cd & cb+d^2 \end{pmatrix},$$

$$(a+d)\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2+ad & ab+bd \\ ca+cd & ad+d^2 \end{pmatrix}$$

и

$$(ad-bc)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix},$$

поэтому

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 - (a+d)\begin{pmatrix} a & b \\ c & d \end{pmatrix} + (ad-bc)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 0.$$

Если полином $P(X)$ имеет кратный корень, то

$$P(X) = (X-\lambda)^2,$$

следовательно,

$$\left(F - \lambda\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)^2 = 0.$$

При этом существуют две возможности

$$F - \lambda\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 0, \quad \text{тогда} \quad F = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix},$$

или

$$F - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \neq 0.$$

Пусть $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ — собственный вектор преобразования F, а ненулевой вектор $\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ *не* является собственным вектором F. Тогда вектор

$$\left[F - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$$

будет собственным вектором F, потому что $\left(F - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right)^2 = 0$, и он должен быть пропорционален вектору $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$. Действительно, умножая $\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ на соответствующую ненулевую константу, можно получить

$$\left[F - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}.$$

И в этом случае матрица $B = \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}$ несингулярна, а предыдущее уравнение можно записать в виде

$$FB = B \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \quad \text{или} \quad F = B \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} B^{-1}.$$

Обратное утверждение: любая матрица вида

$$F = B \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} B^{-1}$$

обладает свойством $\left[F - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right]^2 = 0$, но при этом

$$F - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \neq 0,$$

что можно легко проверить.

Случай 3. Комплексные корни. Иногда нам приходится работать с преобразованиями, не имеющими вещественных собственных значений и собственных векторов. Наиболее очевидные примеры таких преобразований: повороты на угол, не кратный π; конформное преобразование, которое состоит из поворота плоскости и последующего однородного «растяжения».

Давайте посмотрим, что получится, если мы попытаемся найти собственные значения и собственные векторы конформной матрицы[2]

$$C = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}.$$

Характеристическое уравнение этой матрицы имеет вид

$$(x - \lambda)^2 + y^2 = 0.$$

Далее получаем

$$(\lambda - x)^2 = -y^2,$$
$$\lambda - x = \pm \mathrm{i}y$$

и

$$\lambda = x \pm \mathrm{i}y.$$

В параграфе 2.1 мы отметили, что конформная матрица $\begin{pmatrix} x & -y \\ y & x \end{pmatrix}$ может быть применена для представления комплексного числа $x + \mathrm{i}y$. Сейчас мы видим, что это комплексное число является собственным значением матрицы. Более того, для любой пары комплексно сопряженных чисел $x+\mathrm{i}y$, $x-\mathrm{i}y$ существует вещественная конформная матрица, для которой эти числа являются собственными значениями. Конечно, мы не можем дать геометрическую интерпретацию комплексных собственных значений, потому что соответствующие собственные векторы имеют комплексные компоненты и не могут рассматриваться как векторы на вещественной плоскости.

[2] В дальнейшем предполагается, что $y \neq 0$. — *Прим. ред.*

Покажем, что любая матрица F с собственными значениями $x \pm \mathrm{i}y$ может быть записана в форме

$$\boxed{F = BCB^{-1}, \qquad C = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}.}$$

Доказательство. Рассмотрим матрицу $G = F - x\mathbb{I}$. Согласно теореме Гамильтона–Кэли, имеем

$$(F - x\mathbb{I})^2 + y^2\mathbb{I} = 0.$$

Таким образом,

$$G^2 = -y^2\mathbb{I}.$$

Возьмем произвольный вектор $\mathbf{v}_1 \neq 0$ и определим новый вектор $\mathbf{v}_2$ по формуле

$$\mathbf{v}_2 = y^{-1}G\mathbf{v}_1.$$

Тогда

$$G\mathbf{v}_2 = y^{-1}G^2\mathbf{v}_1 = -y\mathbf{v}_1,$$

а по определению вектора $\mathbf{v}_2$

$$G\mathbf{v}_1 = y\mathbf{v}_2.$$

Пусть B — матрица со столбцами $\mathbf{v}_1$ и $\mathbf{v}_2$. Тогда $\mathbf{v}_1 = B\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ и $\mathbf{v}_2 = B\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, и, следовательно,

$$GB\begin{pmatrix} 1 \\ 0 \end{pmatrix} = yB\begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{и} \quad GB\begin{pmatrix} 0 \\ 1 \end{pmatrix} = -yB\begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

поэтому $GB = B\begin{pmatrix} 0 & -y \\ y & 0 \end{pmatrix}$. Умножая эти соотношения на B^{-1}, получаем, что $G = B\begin{pmatrix} 0 & -y \\ y & 0 \end{pmatrix}B^{-1}$. Итак, матрицу F можно записать в виде

$$\begin{aligned} F = x\mathbb{I} + B\begin{pmatrix} 0 & -y \\ y & 0 \end{pmatrix}B^{-1} &= B(x\mathbb{I})B^{-1} + B\begin{pmatrix} 0 & -y \\ y & 0 \end{pmatrix}B^{-1} \\ &= B\left[x\mathbb{I} + \begin{pmatrix} 0 & -y \\ y & 0 \end{pmatrix}\right]B^{-1}, \end{aligned}$$

а так как $B(x\mathbb{I}) = (x\mathbb{I})B$, то

$$F = B\begin{pmatrix} x & -y \\ y & x \end{pmatrix} B^{-1},$$

что и требовалось доказать.

Например, для простоты выберем $\mathbf{v}_1 = B\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, хотя можно было бы сделать и другой выбор. Тогда

$$\mathbf{v}_2 = y^{-1}G\mathbf{v}_1 = y^{-1}G\begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

т. к. $G\mathbf{v}_1 = y\mathbf{v}_2$. Это значит, что $\mathbf{v}_1$ (первый столбец матрицы B) равен $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, а второй столбец этой матрицы $\mathbf{v}_2$ — это первый столбец матрицы G, деленный на y.

Подведем итог: если $F = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ имеет собственные значения, равные $x \pm \mathrm{i}y$, причем $y \neq 0$, то $F = BCB^{-1}$, где

$$B = \begin{pmatrix} 1 & (a-x)/y \\ 0 & c/y \end{pmatrix} \quad \text{и} \quad C = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}.$$

А матрица $G = F - x\mathbb{I}$ удовлетворяет уравнению $G^2 = -y^2\mathbb{I}$.

В качестве примера рассмотрим матрицу $F = \begin{pmatrix} 7 & -10 \\ 2 & -1 \end{pmatrix}$, имеющую $\operatorname{Tr} F = 6$ и $\operatorname{Det} F = 13$. Характеристическое уравнение $\lambda^2 - 6\lambda + 13 = 0$ имеет корни

$$\lambda = \frac{6 \pm \sqrt{(36-52)}}{2} = 3 \pm 2\mathrm{i}.$$

Таким образом, $x = 3$, $y = 2$. Матрица

$$G = F - 3\mathbb{I} = \begin{pmatrix} 4 & -10 \\ 2 & -4 \end{pmatrix},$$

как и ожидалось, удовлетворяет уравнению $G^2 = -4\mathbb{I}$. Чтобы получить второй столбец матрицы B, нам следует первый столбец G разделить на y:

$$\frac{1}{2}\begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Следовательно, $B = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ и

$$F = BCB^{-1} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & -2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}.$$

Степени матриц. Предположим, что нам задана матрица F, и мы хотим вычислить F^n для разных значений n (или для очень больших n). В следующем параграфе мы дадим пример, когда возникает такая проблема.

Случай 1. Различные вещественные корни. Если

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix},$$

тогда очевидно, что

$$\Lambda^n = \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix}.$$

Это значит, что вычисление степеней диагональной матрицы сводится к вычислению степеней вещественных чисел. Если

$$F = B\Lambda B^{-1},$$

то

$$F^2 = B\Lambda B^{-1} B\Lambda B^{-1} = B\Lambda^2 B^{-1}$$

и тогда по индукции

$$\boxed{F^n = B\Lambda^n B^{-1}.}$$

Случай 2. Равные вещественные корни. Рассмотрим матрицу

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Матрица $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$ коммутирует со всеми матрицами 2×2. Далее мы имеем

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^2 = 0,$$

поэтому можно воспользоваться формулой бинома:

$$\begin{aligned} \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}^n &= \left(\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right)^n = \\ &= \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}^n + n \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}^{n-1} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \end{aligned}$$

т. к. все остальные члены в биномиальном разложении равны нулю. Следовательно,

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}^n = \begin{pmatrix} \lambda^n & n\lambda^{n-1} \\ 0 & \lambda^n \end{pmatrix}.$$

Поэтому если

$$F = B \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} B^{-1},$$

то

$$\boxed{F^n = B \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}^n B^{-1} = B \begin{pmatrix} \lambda^n & n\lambda^{n-1} \\ 0 & \lambda^n \end{pmatrix} B^{-1}.}$$

Случай 3. Комплексные корни. Пусть, наконец,

$$F = BCB^{-1},$$

где $C = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}$ — конформная матрица. Тогда

$$F^n = BC^nB^{-1},$$

где C^n вычисляется одним из двух методов, изложенных в конце параграфа 2.1.

Итак, вычислив собственные значения матрицы F и найдя матрицу B замены базиса, мы получаем простой метод вычисления степеней матрицы F, который применим во всех случаях (разные вещественные собственные значения, два одинаковых вещественных или два комплексно-сопряженных собственных значения).

На самом деле для двух последних случаев у нас нет необходимости вычислять B: если собственные значения вещественны и равны, то матрица $N = (F - \lambda\mathbb{I})$ удовлетворяет условию $N^2 = 0$, следовательно, используя формулу бинома, получаем

$$\boxed{F^n = (\lambda\mathbb{I} + N)^n = \lambda^n\mathbb{I} + n\lambda^{n-1}N.}$$

Для комплексно-сопряженных собственных значений $x \pm \mathrm{i}y$ имеем равенство

$$F - x\mathbb{I} = yH, \quad \text{где} \quad H^2 = -\mathbb{I}.$$

Следовательно,

$$\begin{aligned} F^n &= (x\mathbb{I} + yH)^n = x^n\mathbb{I} + nx^{n-1}yH + \binom{n}{2} x^{n-2}y^2H^2 + \cdots \\ &= \left(x^n - \binom{n}{2} x^{n-2}y^2 + \cdots\right)I + \left(\binom{n}{1} x^{n-1}y - \binom{n}{3} x^{n-3}y^3 + \cdots\right)H. \end{aligned}$$

2.3. Процессы Маркова

В этом параграфе мы покажем, как умножение матриц применяется в вероятностных задачах. При этом мы не собираемся писать подробное введение в теорию вероятностей. Мы просто приведем основные положения этой теории. Каждому «событию» A, B и т. д. ставится в соответствие «вероятность события» — вещественное число $p(A)$, $p(B)$ и т. д., так что выполнены некоторые условия:

$$0 \le p(A) \le 1, \quad 0 \le p(B) \le 1, \quad \ldots$$

в соответствии с определенными правилами:

- вероятность достоверного события равна 1;
- вероятность невозможного события равна 0;

- если событие A может произойти k взаимно исключающими способами $A_1, \ldots, A_k$, что мы записываем как

$$A = A_1 \cup \cdots \cup A_k, \quad A_i \cap A_j = \varnothing, \quad \text{при} \quad i \neq j,$$

то

$$p(A) = p(A_1) + \cdots + p(A_k).$$

В частности, если A^c обозначает «дополнительное событие», означающее, что событие A не произойдет, то $A \cup A^c$ достоверно (событие A либо произойдет, либо нет), и $A \cap A^c = \varnothing$, поэтому $p(A) + p(A^c) = 1$. Кроме того, применяется понятие «условной вероятности»:

$p(B|A)$ — условная вероятность события B при заданном A.

Таким образом, если событие A значит, что «сегодня идет дождь», а событие B значит, что «завтра будет ясно», то $p(B|A)$ есть вероятность того, что завтра будет ясно, если известно, что сегодня идет дождь. Тогда у нас есть правило

$$p\,(A \cap B) = p(B|A)p(A),$$

т. е. вероятность двух событий A и B равна произведению условной вероятности события B при заданном A на вероятность события A.

В частности, если $A_1, \ldots, A_k$ — взаимно исключающие друг друга события, $A_i \cap A_j = \varnothing$, и событие B может произойти только при условии, что произойдет одно из событий A_i, то:

$$B = (B \cap A_1) \cup \cdots \cup (B \cap A_k)\,,$$

тогда

$$p(B) = p\,(B \cap A_1) + \cdots + p\,(B \cap A_k)\,,$$

следовательно,

$$p(B) = p(B|A_1)p(A_1) + \cdots + p(B|A_k)p(A_k).$$

Теперь рассмотрим систему, которая может находиться в одном из двух состояний. Например, прибор выключен или включен; или игра в бадминтон: состояние 1 означает ситуацию, когда подает первый игрок, а в состоянии 2 подает второй игрок. При этом один «шаг» соответствует переходу подачи от одного игрока к другому. Таким образом, во время игры в бадминтон система может находиться в одном состоянии (игрок получил очко и продолжает подавать), потом происходит переход из одного состояния в другое (игрок проиграл очко, и подача переходит к противнику). Итак, пусть подает первый игрок, с вероятностью 0.8 он может выиграть очко, а с вероятностью 0.2 может его проиграть. Когда подает второй игрок, то он выигрывает очко с вероятностью 0.7, а проигрывает с вероятностью 0.3. В реальной игре для каждого игрока вероятность выиграть очко зависит от множества факторов (от его желания победить, от степени его усталости, деморализации и т. д.). Мы же сделаем радикальное упрощающее предположение: развитие игры зависит только от того, кто в данный момент подает, и ни от чего другого. Все перечисленные выше вероятности можно организовать в матрицу

$$\begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix}.$$

Таким образом, 0.8 есть условная вероятность системе находиться в состоянии 1 *после* шага, если *до* выполнения шага она находилась также в состоянии 1; 0.2 есть условная вероятность системе находиться в состоянии 2 после шага, если до него она была в состоянии 1.

Вообще, в случае дискретного времени и стационарной системы с двумя состояниями процесс Маркова — это процесс, когда состояния могут изменяться в дискретные моменты времени, а вероятность перехода из одного состояния в другое зависит только от состояния самой системы и не зависит от прошлой истории системы и от времени перехода между состояниями. Таким образом, все определяется только четырьмя «вероятностями перехода», которые могут быть организованы в матрицу

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

где

$a =$ вероятность перехода из состояния 1 в состояние 1;
$b =$ вероятность перехода из состояния 2 в состояние 1;
$c =$ вероятность перехода из состояния 1 в состояние 2;
$d =$ вероятность перехода из состояния 2 в состояние 2.

Предположим, что мы не знаем, в каком состоянии находится система в данный момент времени. Знаем только, что с вероятностью p система находится в состоянии 1 и с вероятностью $q = 1 - p$ в состоянии 2. Такое распределение вероятностей может быть представлено вектором $\mathbf{v} = \begin{pmatrix} p \\ q \end{pmatrix}$. После одного шага, согласно закону для условной вероятности, мы получаем

$$\left\{\begin{matrix}\text{вероятность быть}\\ \text{в состоянии 1}\\ \text{после одного шага}\end{matrix}\right\} = \left\{\begin{matrix}\text{вероятность пере-}\\ \text{хода из состояния 1}\\ \text{в состояние 1}\end{matrix}\right\} \times \left\{\begin{matrix}\text{вероятность}\\ \text{быть в}\\ \text{состоянии 1}\end{matrix}\right\}$$

$$+ \left\{\begin{matrix}\text{вероятность пере-}\\ \text{хода из состояния 2}\\ \text{в состояние 1}\end{matrix}\right\} \times \left\{\begin{matrix}\text{вероятность}\\ \text{быть в}\\ \text{состоянии 2}\end{matrix}\right\} = ap + bq.$$

Аналогично, вероятность находиться в состоянии 2 после выполнения одного шага равна

$$cp + dq.$$

Другими словами, новый вектор вероятности равен

$$\begin{pmatrix} ap + bq \\ cp + dq \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = A\mathbf{v}.$$

Давайте проиллюстрируем это на примере игры в бадминтон. Предположим, мы знаем, что первый игрок начинает игру. Вектор $\mathbf{v}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ является вектором вероятности в начале игры. После первой подачи вектор вероятности становится

$$\mathbf{v}_1 = A\mathbf{v}_0 = \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix}.$$

После второй подачи он равен

$$\mathbf{v}_2 = A\mathbf{v}_1 = \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix} \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix} = \begin{pmatrix} 0.7 \\ 0.3 \end{pmatrix} = A^2\mathbf{v}_0.$$

После третьей подачи

$$\mathbf{v}_3 = A\mathbf{v}_2 = \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix} \begin{pmatrix} 0.7 \\ 0.3 \end{pmatrix} = \begin{pmatrix} 0.65 \\ 0.35 \end{pmatrix} = A^3\mathbf{v}_0$$

и так далее. В результате n шагов появляется матрица A^n.

Интуиция позволяет сделать вывод, что после большого числа шагов вектор вероятности практически не будет зависеть от начального состояния, т. е. вероятность того, что первый игрок подает, чтобы выиграть пятнадцатое очко, вряд ли сильно зависит от того, кто подавал в первый раз. Эта гипотеза подтверждается вычислением. Действительно,

$$A^2 = \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix} \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix} = \begin{pmatrix} 0.7 & 0.55 \\ 0.3 & 0.45 \end{pmatrix}$$

$$A^4 = \begin{pmatrix} 0.7 & 0.55 \\ 0.3 & 0.45 \end{pmatrix}^2 = \begin{pmatrix} 0.63 & 0.56 \\ 0.37 & 0.44 \end{pmatrix}$$

$$A^8 = \begin{pmatrix} 0.63 & 0.56 \\ 0.37 & 0.44 \end{pmatrix}^2 = \begin{pmatrix} 0.602 & 0.598 \\ 0.398 & 0.402 \end{pmatrix}$$

$$A^{16} = \begin{pmatrix} 0.602 & 0.598 \\ 0.398 & 0.402 \end{pmatrix}^2 = \begin{pmatrix} 0.600006 & 0.599994 \\ 0.399994 & 0.400006 \end{pmatrix}.$$

Теперь можно предположить, что

$$\lim_{n\to\infty} A^n = \begin{pmatrix} 0.6 & 0.6 \\ 0.4 & 0.4 \end{pmatrix}.$$

В общем случае легко показать, что если b и c одновременно не равны 0 или 1, то $\lim_{n\to\infty} A^n$ существует. Надо только найти собственные значения и собственные векторы матрицы A. Учитывая, что $a + c = 1$, $b + d = 1$, можно написать

$$A = \begin{pmatrix} 1-c & b \\ c & 1-b \end{pmatrix}.$$

Так как $\operatorname{Tr} A = 2-(b+c)$ и $\operatorname{Det} A = 1-b-c+bc-bc = 1-(b+c)$, характеристическое уравнение имеет вид

$$\lambda^2 - [2-(b+c)]\lambda + 1 - (b+c) = 0,$$

или

$$(\lambda - 1)(\lambda - (1-b-c)) = 0.$$

Из этого уравнения получаем собственные значения $\lambda_1 = 1$, $\lambda_2 = 1-(b+c)$. Заметим, что $|\lambda_2| \leq 1$, где равенство выполняется только если $b=c=0$ или $b=c=1$.

Собственные векторы легко получаются из рассмотрения матрицы

$$A - \lambda_1 \mathbb{I} = \begin{pmatrix} 1-c & b \\ c & 1-b \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -c & b \\ c & -b \end{pmatrix}.$$

Ядро этой сингулярной матрицы пропорционально собственному вектору, соответствующему $\lambda_1 = 1$. Нормируем этот вектор так, чтобы его компоненты в сумме давали 1. В результате получаем

$$\mathbf{v}_1 = \frac{1}{b+c}\begin{pmatrix} b \\ c \end{pmatrix}.$$

Образ $A - \lambda_1 \mathbb{I}$ пропорционален собственному вектору, соответствующему λ_2, который удобно выбрать в виде $\mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

Операция A может быть наглядно представлена через эти собственные значения и собственные векторы. Проведем отрезок прямой, соединяющий точки $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ и $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Тогда конец вектора $\mathbf{v}_1$ лежит на нем. Любой вектор, конец которого лежит на этом отрезке, может быть записан в виде $\mathbf{v} = \mathbf{v}_1 + \alpha \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Учитывая, что $A\mathbf{v}_1 = \mathbf{v}_1$ и $A\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \lambda_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, получаем (рис. 2.2)

$$A\mathbf{v} = \mathbf{v}_1 + \alpha\lambda_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

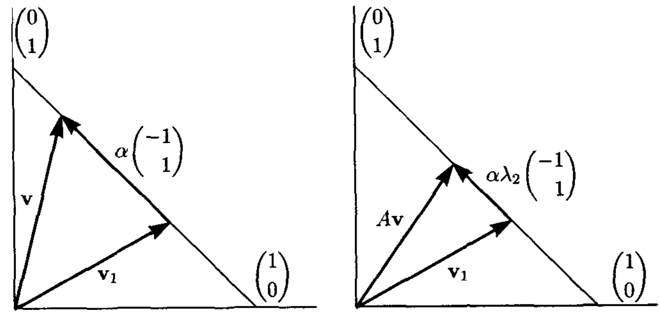


Рис. 2.2

или для произвольной степени n

$$A^n\mathbf{v} = \mathbf{v}_1 + \alpha(\lambda_2)^n \begin{pmatrix}-1\\1\end{pmatrix}.$$

Из того, что $|\lambda_2| < 1$, очевидно, следует, что, *независимо от того, каким был начальный вектор* $\mathbf{v}$, $\lim A^n\mathbf{v} = \mathbf{v}_1$.

Выразим матрицу A явным образом через диагональную. Для этого запишем $A = B\Lambda B^{-1}$, где

$$B = \begin{pmatrix} b/(b+c) & -1 \\ c/(b+c) & 1 \end{pmatrix}, \quad B^{-1} = \begin{pmatrix} 1 & 1 \\ -c/(b+c) & b/(b+c) \end{pmatrix},$$

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Тогда

$$A^n = B\Lambda^n B^{-1}.$$

Учитывая, что

$$\lim_{n\to\infty} \Lambda^n = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

получим

$$\lim_{n\to\infty} A^n = \begin{pmatrix} b/(b+c) & -1 \\ c/(b+c) & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -c/(b+c) & b/(b+c) \end{pmatrix}$$
$$= \begin{pmatrix} b/(b+c) & -1 \\ c/(b+c) & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \frac{1}{b+c}\begin{pmatrix} b & b \\ c & c \end{pmatrix}.$$

Таким образом, если $\mathbf{v} = \begin{pmatrix} p \\ q \end{pmatrix}$ при $p + q = 1$, то

$$\lim_{n\to\infty} A^n \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} b/(b+c) \\ c/(b+c) \end{pmatrix}.$$

Итак, если A — стохастическая матрица 2×2, т. е. $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ с условием $a \geq 0$, $b \geq 0$, $c \geq 0$, $d \geq 0$, $a + c = 1$, $b + d = 1$, то ее собственные значения равны 1 и $1 - (b + c)$, а соответствующие собственные векторы задаются столбцами $\begin{pmatrix} b/(b+c) \\ c/(b+c) \end{pmatrix}$ и $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ соответственно. Если все входящие в матрицу параметры строго положительны, то повторное действие A заставляет систему стремиться к предельному состоянию $\begin{pmatrix} b/(b+c) \\ c/(b+c) \end{pmatrix}$. Переход от состояния в данный момент к этому предельному состоянию осуществляется умножением на λ_2 при каждом шаге. Вернемся к игре в бадминтон. В этом случае матрица $A = \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix}$, $\lambda_2 = 0.5$ и предельное состояние $\begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix}$. Это значит, что после розыгрыша большого числа очков первый игрок имел 60% подач.

Пусть теперь матрица $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ — стохастическая и не удовлетворяет условиям строгой положительности. Очевидно, что

$$A^n = \begin{cases} \mathbb{I}, \text{ если } n \text{ — четное число}, \\ A, \text{ если } n \text{ — нечетное число}. \end{cases}$$

Смысл матрицы A очевиден. Она представляет безусловный переход в другое состояние. В этом случае нет предела A^n при $n \to \infty$. (Однако, в некотором «усредненном» смысле мы ожидаем, что система проводит в каждом состоянии половину времени.)

Несложно сделать обобщение на случай процессов Маркова для систем с большим числом состояний. Процессы с тремя состояниями описываются матрицами 3×3 и т. д. Параметры каждого столбца должны быть неотрицательными и в сумме дают 1.

Такова, например, стохастическая матрица 3×3

$$A = \begin{pmatrix} 0.5 & 0 & 0.1 \\ 0.3 & 0.6 & 0 \\ 0.2 & 0.4 & 0.9 \end{pmatrix}.$$

С некоторыми поправками сохраняются важные свойства матриц 2×2. Например, матрица

$$\begin{pmatrix} 0.5 & 0.3 & 0 & 0 \\ 0.5 & 0.7 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

представляет систему, в которой невозможны переходы из первых двух состояний в два последние и наоборот. Вектор вероятности, ограниченный первыми двумя состояниями, имеет предел. Вектор, ограниченный двумя последними состояниями (т. е. первые две компонеты равны 0) предела при $n \to \infty$ не имеет, и его величина зависит от четности n. Нетрудно понять, как все это будет звучать на языке матричных элементов матрицы A. За исключением этих случаев n-мерная матрица дает то же самое, что и двумерная. Матрица имеет собственное значение, равное 1, соответствующий собственный вектор определяет предельное состояние. Все другие собственные значения меньше единицы, и $\lim_{n\to\infty} A^n$ является сингулярной матрицей, преобразующей любой вектор вероятности в собственный вектор, соответствующий $\lambda = 1$.

Резюме

А. Конформные матрицы

Вы познакомились в этой главе с определением конформной матрицы и с геометрической интерпретацией преобразования, которое она представляет.

Вы должны уметь установить и применять изоморфизм между конформными матрицами и комплексными числами.

B. Собственные значения и собственные векторы

Вы должны уметь писать характеристическое уравнение матрицы 2×2 и использовать его для определения собственных значений матрицы; уметь получать собственные векторы, соответствующие вещественным собственным значениям матрицы 2×2, уметь описывать действие матрицы на языке собственных значений и собственных векторов.

C. Подобие матриц

Для данной матрицы A (2×2) вы должны уметь построить такую матрицу B, что $A = BCB^{-1}$, где C диагональна, если A имеет разные вещественные собственные значения; C конформна, если A имеет комплексно-сопряженные собственные значения и $C = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$, если A имеет два равных собственных значения, но при этом $A \neq \lambda\mathbb{I}$. В каждом случае вы должны дать геометрическую интерпретацию столбцов матрицы B.

D. Процессы Маркова

Вы должны уметь написать матрицу $n \times n$, представляющую процесс Маркова для n состояний.

Для матрицы A (2×2), представляющей процесс Маркова, вы должны уметь связывать собственные значения и собственные векторы с поведением вероятностей двух состояний.

Задачи

2.1. Рассмотрим две конформные матрицы

$$F_1 = \begin{pmatrix} 3 & -4 \\ 4 & 3 \end{pmatrix} \quad \text{и} \quad F_2 = \begin{pmatrix} 4 & 3 \\ -3 & 4 \end{pmatrix}.$$

(a) Напишите комплексные числа z_1 и z_2, соответствующие этим матрицам.

(b) Выразите матрицы F_1 и F_2 в виде произведения числа на единичную матрицу и на матрицу поворота. Используя тождество $e^{i\theta} = \cos\theta + i\sin\theta$, запишите z_1 и z_2 в полярных координатах $z = re^{i\theta}$.

(c) Вычислите F_1^{-1}, потом вычислите z_1^{-1}, домножая числитель и знаменатель на сопряженное число $a - bi$. Сравните полученные результаты.

(d) Вычислите F_1F_2 и F_2F_1. Потом вычислите z_1z_2 и сравните результаты.

2.2. Проверьте теорему Муавра на примере двух конформных матриц F_1 и F_2 из первого упражнения, т. е. вычислите F_1^3 и F_2^3.

2.3. (a) Покажите, что матрица $R = \begin{pmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{pmatrix}$ представляет поворот против часовой стрелки приблизительно на 37°. Вычислите R^{-1}.

(b) Матрица $S = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ представляет сдвиг вдоль оси x. Вычислите S^{-1} и дайте геометрическую интерпретацию результата.

(c) Вычислите $A = RSR^{-1}$ и дайте геометрическую интерпретацию результата. То же самое сделайте для A^{-1}, $B = RS^{-1}R^{-1}$ и B^{-1}.

2.4. Матрицу $F = \begin{pmatrix} -7 & 18 \\ -3 & 8 \end{pmatrix}$ приведите к диагональному виду согласно процедуре:

(a) напишите характеристический полином $P(\lambda)$, приравняйте его к нулю и найдите собственные значения матрицы F. (Ответ: $\lambda_1 = 2, \lambda_2 = -1$.)

(b) Проверьте, что в соответствии с теоремой Гамильтона–Кэли $P(F) = 0$.

(c) Для каждого собственного значения получите собственные векторы. Пусть для каждого собственного вектора $y = 1$.

(d) Постройте матрицы B и B^{-1} так, чтобы $F = B\Lambda B^{-1}$.

2.5. Матрицу лоренцева преобразования $L_2 = \begin{pmatrix} 5/4 & 3/4 \\ 3/4 & 5/4 \end{pmatrix}$ приведите к диагональному виду, записав ее как $L_2 = B\Lambda B^{-1}$, где B есть матрица вращения, а Λ диагональна. Дайте геометрическую интерпретацию результата.

2.6. Получите такую обратимую матрицу B и такую диагональную матрицу D, чтобы выполнялось равенство

$$B\begin{pmatrix} 4 & -3 \\ -1 & 2 \end{pmatrix}B^{-1} = D.$$

2.7. Матрица $F = \begin{pmatrix} -1 & 9 \\ -1 & 5 \end{pmatrix}$ имеет равные собственные значения. Приведите ее к диагональному виду, следуя процедуре:

(a) напишите характеристический полином $P(\lambda)$ и найдите собственные значения;

(b) получите собственный вектор матрицы F в форме $\begin{pmatrix} x_1 \\ 1 \end{pmatrix}$;

(c) постройте матрицу $G = F - \lambda\mathbb{I}$. Покажите, что из теоремы Гамильтона–Кэли следует уравнение $G^2 = 0$. Получите ядро и образ матрицы G;

(d) получите вектор $\begin{pmatrix} x_2 \\ 1 \end{pmatrix}$, такой, что $G\begin{pmatrix} x_2 \\ 1 \end{pmatrix} = \begin{pmatrix} x_1 \\ 1 \end{pmatrix}$. Постройте матрицы B и B^{-1} так, чтобы выполнялось равенство $F = B\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} B^{-1}$.

2.8. Диагонализуйте матрицу

$$F = \begin{pmatrix} 3/5 & 1/5 \\ -4/5 & 7/5 \end{pmatrix},$$

имеющую равные собственные значения. Получите образ и ядро матрицы $G = F - \lambda\mathbb{I}$, дайте геометрическую интерпретацию преобразования F.

2.9. Пусть A есть матрица с собственными значениями $\lambda_1 > \lambda_2 > 0$.

(a) Диагонализуя матрицу A, опишите процедуру вычисления матрицы $G_n = \lambda_1^{-n} A^n$. Покажите, что матрица $F = \lim_{n\to\infty} G_n$ сингулярна.

(b) Проделайте эту процедуру для матрицы $A = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix}$, вычисляя явно матрицы G_n и F. Найдите собственные значения и собственные векторы матрицы A, ядро и образ преобразования F и свяжите их с собственными векторами матрицы A.

2.10. Для любой матрицы A след ($\mathrm{Tr}\, A$) определяется как сумма матричных элементов, стоящих на главной диагонали. Таким образом, если $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, то $\mathrm{Tr}\, A = a + d$.

(a) Докажите, что для двух матриц A и B (2×2) $\mathrm{Tr}\,(AB) = \mathrm{Tr}\,(BA)$, даже если A и B не коммутируют друг с другом.

(b) Докажите, что $\operatorname{Tr} A$ равен сумме собственных значений матрицы A. Докажите, что если $A = SBS^{-1}$, то $\operatorname{Tr} A = \operatorname{Tr} B$.

(c) Используя результат пункта (a), докажите, что

$$\operatorname{Tr}(ABC) = \operatorname{Tr}(BCA) = \operatorname{Tr}(CAB).$$

2.11. Матрицу $F = \begin{pmatrix} 3 & -1 \\ 5 & -1 \end{pmatrix}$, имеющую комплексно-сопряженные собственные значения, запишите в форме $F = BCB^{-1}$, следуя процедуре, описанной ниже.

(a) Найдите собственные значения матрицы F.

(b) Постройте конформную матрицу C, имеющую те же собственные значения, что и F.

(c) Постройте матрицу B в виде $\begin{pmatrix} 1 & c \\ 0 & d \end{pmatrix}$.

2.12. Пусть $A = \begin{pmatrix} 4 & -5 \\ 1 & 0 \end{pmatrix}$. Постройте такую конформную матрицу C и такую матрицу сдвига S, чтобы $A = SCS^{-1}$.

2.13. Пусть $F = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ — матрица с различными вещественными собственными значениями λ_1 и λ_2, и пусть

$$x = \frac{1}{2}(\lambda_1 + \lambda_2), \quad y = \frac{1}{2}(\lambda_1 - \lambda_2).$$

(a) Покажите, что $H = F - x\mathbb{I}$ удовлетворяет уравнению

$$H^2 = y^2\mathbb{I}.$$

(b) Покажите, что матрица $S = \begin{pmatrix} x & y \\ y & x \end{pmatrix}$ имеет те же собственные значения, что и F.

(c) Постройте матрицу B, у которой первый столбец $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. При этом должны выполняться условия $H = B\begin{pmatrix} 0 & y \\ y & 0 \end{pmatrix}B^{-1}$ и $F = B\begin{pmatrix} x & y \\ y & x \end{pmatrix}B^{-1}$.

(d) Постройте такую матрицу R, чтобы выполнялось равенство

$$\begin{pmatrix} x & y \\ y & x \end{pmatrix} = R\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}R^{-1}.$$

Докажите, что $BR\begin{pmatrix}1\\0\end{pmatrix}$ и $BR\begin{pmatrix}0\\1\end{pmatrix}$ являются собственными векторами матрицы F.

2.14. Пусть F есть матрица 2×2 с различными вещественными собственными значениями λ_1 и λ_2. Определим

$$P_1 = \frac{F - \lambda_2 I}{\lambda_1 - \lambda_2}, \quad P_2 = \frac{F - \lambda_1 I}{\lambda_2 - \lambda_1}.$$

Докажите следующие свойства P_1 и P_2:

(a) P_1 и P_2 являются проекциями: $P_1^2 = P_1, P_2^2 = P_2$;

(b) $P_1P_2 = P_2P_1 = 0$;

(c) $F = \lambda_1 P_1 + \lambda_2 P_2$;

(d) $F^n = \lambda_1^n P_1 + \lambda_2^n P_2$.

(e) Вычислите P_1 и P_2 для случая $F = \begin{pmatrix}3 & 4\\-1 & -2\end{pmatrix}$. Воспользовавшись этим результатом, вычислите F^7.

2.15. Пусть F — матрица 2×2, характеристическое уравнение которой имеет корни $\lambda = x \pm \mathrm{i}y$. Можно переписать это в полярных координатах $x \pm \mathrm{i}y = re^{\pm \mathrm{i}\theta}$, где $r = \sqrt{(x^2 + y^2)}$ и $e^{\pm \mathrm{i}\theta} = \cos\theta \pm \mathrm{i}\sin\theta$. Если F — конформная матрица, то она поворачивает плоскость на угол θ и однородно растягивает ее с коэффициентом r. В этой задаче нет необходимости требовать конформность матрицы F.

(a) Покажите, что матрица F^n пропорциональна единичной матрице для целых чисел n тогда и только тогда, когда $n\theta = m\pi$ для целого числа m. В этом случае

$$F^n = (-1)^m r^n \mathbb{I}.$$

(Подсказка: $F = BCB^{-1}$, где C — конформная матрица.)

(b) Запишите матрицу $F = \begin{pmatrix}3 & 7\\-1 & -1\end{pmatrix}$ в виде BCB^{-1}. Найдите наименьшее число n, для которого F^n пропорциональна единичной матрице. Проверьте ответ прямым вычислением.

(c) Покажите, что этот ответ соответствует теореме Гамильтона–Кэли.

(d) Получите «квадратный корень» матрицы $G = \begin{pmatrix}-2 & -15\\3 & 10\end{pmatrix}$, т. е. найдите такую матрицу A, чтобы выполнялось условие

$A^2 = G$. Напоминаем формулы:

$$\cos^2\frac{\theta}{2} = \frac{1}{2}(1+\cos\theta), \quad \sin^2\frac{\theta}{2} = \frac{1}{2}(1-\cos\theta).$$

2.16. Композитор-модернист Але А. Тори[3] создает музыкальные произведения из двух нот, следуя теории процессов Маркова.

1. Если $(n-1)$-я нота есть Фа, тогда вероятность p_n того, что n-я нота тоже Фа, будет равна $3/4$, а вероятность q_n того, что n-я нота Соль, равна $1/4$.
2. Если $(n-1)$-я нота есть Соль, то вероятность p_n того, что n-я нота Фа, равна $1/2$, а вероятность q_n того, что n-я нота Соль, тоже равна $1/2$.

   (a) Постройте матрицу A (2×2), которая преобразует вероятности $\begin{pmatrix} p_{n-1} \\ q_{n-1} \end{pmatrix}$ в вероятности $\begin{pmatrix} p_n \\ q_n \end{pmatrix}$.

   (b) Допустим, что первая нота есть Фа. С помощью матрицы A найдите, чему равна вероятность третьей ноте тоже быть Фа.

   (c) Получите собственные значения и соответствующие собственные векторы матрицы A.

   (d) Допустим, что первая нота есть Фа. Это значит, что

$$\begin{pmatrix} p_1 \\ q_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

   Покажите на диаграмме последовательность векторов

$$\begin{pmatrix} p_2 \\ q_2 \end{pmatrix}, \quad \begin{pmatrix} p_3 \\ q_3 \end{pmatrix}, \quad \ldots$$

   и получите предел этой последовательности, интерпретируя его через собственные значения матрицы A.

2.17. Один студент, успешно изучавший теорию вероятностей и увлекавшийся футболом, придумал игру со следующими правилами.

1. Если в $(n-1)$-й раз некоторый игрок дал пас другому игроку, то в n-й раз он делает пас с вероятностью $p_n = 1/6$ и ведет мяч сам с вероятностью $q_n = 5/6$.

[3]Непереводимая игра слов: aleatory — случайный. — *Прим. ред.*

2. Если в $(n-1)$-й раз игрок повел мяч сам, то вероятность паса на n-м ходу равна $p_n = 2/3$, а вероятность самостоятельного прохода равна $q_n = 1/3$.
   (a) Постройте матрицу A, которая преобразует вероятности $\begin{pmatrix} p_{n-1} \\ q_{n-1} \end{pmatrix}$ в вероятности $\begin{pmatrix} p_n \\ q_n \end{pmatrix}$.
   (b) Для матрицы A найдите собственные значения и соответствующие им собственные векторы. Покажите на диаграмме, как A действует на собственные векторы.
   (c) Никто не знает, что капитан сделает, получив мяч в первый раз. Однако, видеозапись игры показала, что при втором ходе в половине случаев он отдавал пас, а в другой половине случаев вел мяч сам. Каковы вероятности $\begin{pmatrix} p_1 \\ q_1 \end{pmatrix}$ для первого хода?

2.18. Профессор Константин Байес много лет читает курс «Стохастические методы в классической археологии». Всем известно, что Байес подбирает экзаменационные вопросы, вытаскивая разноцветные шарики из древнегреческих урн, стоящих в его кабинете, но содержание урн он держит в секрете. Однако, анализируя вопросы Байеса, заданные им на выпускных экзаменах в течение многих лет, а все они хранятся в библиотеке Ламонта, студенты выяснили следующее:

1. Если на выпускном экзамене в $(N-1)$-м году профессор спрашивал о статуях, то на выпускном экзамене N-го года вероятность получить вопрос о статуях равнялась $1/2$, и вероятность получить вопрос о керамике тоже равнялась $1/2$;
2. Если на выпускном экзамене в $(N-1)$-м году профессор задавал вопрос о керамике, то в N-м году вопрос о статуях на выпускном экзамене можно было получить с вероятностью $1/4$, а вопрос о керамике — с вероятностью $3/4$.
   (a) Напишите матрицу M, преобразующую распределение вероятностей вопросов о статуях и о керамике $\begin{pmatrix} p_{N-1} \\ q_{N-1} \end{pmatrix}$ в распределение вероятностей вопросов на выпускных экзаменах следующего года $\begin{pmatrix} p_N \\ q_N \end{pmatrix}$.
   (b) Найдите собственные значения и собственные векторы матрицы M и запишите ее в форме SDS^{-1}, где D — диагональная матрица.

(c) Заходя регулярно в кабинет Байеса, студенты, наконец, поняли метод профессора. У Байеса было две урны, но он пользовался ими дважды: один раз для себя, предварительно, а потом еще раз для окончательного экзаменационного вопроса, так что матрица M представляет собой *два* шага марковского процесса. Диагонализуя матрицу M, найдите две возможные матрицы N для *одного* шага процесса.

2.19. Джон и Илья играют в мяч, который может быть либо в лунке H, либо в лунке Y. Джон предпочитает мяч в лунке H, а Илья — в лунке Y. Джон получает ход. Если мяч лежит в H, он не делает ничего. Но если мяч оказывается в Y, то он пытается перекатить его в H. Это не так просто. Вероятность того, что Джону удается перекатить мяч, равна $2/3$, но с вероятностью $1/3$ мяч скатывается обратно в лунку Y. Когда Илья получает ход, он ничего не делает, если мяч лежит в лунке Y, и пытается перекатить, если мяч находится в H. Илья играет хуже Джона, и вероятность удачи равна только $1/2$.

(a) Мяч лежит в лунке Y и играет Джон. Чему равна вероятность мячу быть в лунке H после *второго* хода Джона?

(b) Джон играл n раз, а Илья $(n-1)$ раз. Чему равна вероятность того, что после n-го хода Джона мяч окажется в H?

(c) Допустим, что игра продолжается «очень долго». Вы смотрите на игру после хода Ильи. Чему равна вероятность увидеть мяч в лунке H? Сколько надо сделать ходов, если мы хотим быть уверенными в этой величине вероятности с точностью до 0.001?

2.20. Банк разработал правила, согласно которым очередь в окно не должна быть длиннее двух человек. Если появляется третий клиент, то все трое сопровождаются в кабинет менеджера и обслуживаются по высшему разряду. Перед окном нет никакой очереди. Более того, вооруженный охранник на входе в банк не пропускает более одного клиента в минуту (именно столько времени требуется для тщательной проверки клиента). В результате длина очереди в окно определяется процессом Маркова, описывающим, что происходит в течение одной минуты.

1. Если нет очереди, то с вероятностью $1/2$ клиент появится в течение одной минуты и с вероятностью $1/2$ не появится.
2. Перед окном стоит один человек. С вероятностью $1/6$ этого клиента обслужат, и он уйдет; с вероятностью $1/3$ появится второй клиент и с вероятностью $1/2$ ничего не изменится.

3. Перед окном в очереди два человека. С вероятностью $1/6$ одного клиента обслужат, и он уйдет; с вероятностью $1/6$ появится третий клиент, и все трое уйдут в кабинет менеджера, и с вероятностью $2/3$ ничего не изменится.

   (a) Постройте матрицу M, преобразующую распределение вероятностей $\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$ в момент t в распределение вероятностей в момент $(t+1)$.

   (b) Когда банк открывается в 9 часов утра, $p_1 = 1$. Каково распределение вероятностей $\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$ через три минуты?

   (c) Получите собственные значения и собственные векторы матрицы M.

   (d) Какое будет распределение вероятностей $\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$ после многих часов работы банка? Оцените, через какое время вероятности p_1, p_2 и p_3 будут совпадать с предельными значениями с точностью до 0.001.

   (e) Сколько, в среднем, посетителей обслуживается в течение одной минуты? Сколько клиентов приглашаются в кабинет менеджера?

Модель Эренфеста. Предположим, что у нас есть два ящика и N шаров. Пусть i шаров лежат в первом ящике и $N - i$ во втором. Состояние системы определяется целым числом i, т. е. система имеет $n+1$ состояние: $i = 0$, $i = 1$, $\ldots$, $i = N$. В каждый временной интервал из одного ящика вынимают шар (с вероятностью $1/N$) и перекладывают его в другой. Таким образом, i принимает значения $i-1$ или $i+1$, в зависимости от того, берут шар из первого ящика или кладут в него. Вероятности таких переходов равны iN^{-1} и $(N-i)N^{-1}$ соответственно. Итак,

$$\begin{aligned} p_{i-1,i} &= iN^{-1}, \\ p_{i+1,i} &= (N-i)N^{-1}, \\ p_{j,i} &= 0, \quad j \neq i \pm 1. \end{aligned}$$

Например, если $N = 4$, то матрица 5×5 равна

$$P = \begin{pmatrix} 0 & 1/4 & 0 & 0 & 0 \\ 1 & 0 & 1/2 & 0 & 0 \\ 0 & 3/4 & 0 & 3/4 & 0 \\ 0 & 0 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 1/4 & 0 \end{pmatrix}.$$

Заметим, что P преобразует состояние с четным i в состояние с нечетным i и наоборот. А преобразование P^2 переводит четные состояния в четные и нечетные состояния в нечетные, так как матрица P^2 имеет вид

$$\begin{aligned} p_{i-2,i} &= i(i-1)N^{-2}, \\ p_{i,i} &= [i(N-i+1) + (N-i)(i+1)]N^{-2}, \\ p_{i+2,i} &= (N-i)(N-i-1)N^{-2}, \\ p_{j,i} &= 0, \quad j \neq i - \pm 2, i. \end{aligned}$$

В случае, когда $N = 4$, соответствующая матрица равна

$$P^2 = \begin{pmatrix} 1/4 & 0 & 1/8 & 0 & 0 \\ 0 & 5/8 & 0 & 3/8 & 0 \\ 3/4 & 0 & 3/4 & 0 & 3/4 \\ 0 & 3/8 & 0 & 5/8 & 0 \\ 0 & 0 & 1/8 & 0 & 1/4 \end{pmatrix}.$$

Так как P^2 осуществляет переходы только между состояниями с одинаковой четностью, можно рассматривать отдельно состояния с четными значениями $i = 0, 2, 4, \ldots$ и с нечетными $i = 1, 3, 5, \ldots$ Это значит, что из выписанной матрицы можно выделить четно-четные позиции и нечетно-нечетные в две отдельные матрицы

$$Q = \begin{pmatrix} 1/4 & 1/8 & 0 \\ 3/4 & 3/4 & 1/4 \\ 0 & 1/8 & 3/4 \end{pmatrix} \quad \text{и} \quad R = \begin{pmatrix} 5/8 & 3/8 \\ 3/8 & 5/8 \end{pmatrix}.$$

2.21. (a) Для матрицы Q покажите, что $\begin{pmatrix} 1 \\ 6 \\ 1 \end{pmatrix}$ является собственным вектором, соответствующим собственному значению 1; для матрицы R покажите, что $\begin{pmatrix}1\\1\end{pmatrix}$ является собственным вектором, соответствующим собственному значению 1.

(b) Повторите предыдущее упражнение для случая $N = 5$. Покажите, что «четный» собственный вектор для собственного значения 1 пропорционален $\begin{pmatrix}1\\10\\5\end{pmatrix}$ и «нечетный» собственный вектор пропорционален $\begin{pmatrix}5\\10\\1\end{pmatrix}$.

(c) Докажите в общем случае, что «четные» и «нечетные» собственные векторы для собственного значения 1 пропорциональны вектору, где элементы столбца суть четные и нечетные биномиальные коэффициенты соответственно с четными и нечетными номерами. Другими словами, докажите, что

$$\underset{\text{собственные векторы}}{\underset{\text{четные}}{\begin{pmatrix}\binom{N}{0}\\ \binom{N}{2}\\ \binom{N}{4}\\ \vdots\end{pmatrix}}} \quad \text{и} \quad \underset{\text{собственные векторы.}}{\underset{\text{нечетные}}{\begin{pmatrix}\binom{N}{1}\\ \binom{N}{3}\\ \binom{N}{5}\\ \vdots\end{pmatrix}}}$$

2.22. Числа Фибоначчи.

Числовая последовательность $0, 1, 1, 2, 3, 5, 8, 13\ldots$ называется последовательностью Фибоначчи. Эти числа появляются при изучении многих интересных физических и математических задач, от роста растений до механики небесных тел (см. например книгу D'Arcy Thompson *On Growth and Form*). Рекуррентная формула для этой последовательности имеет вид

$$x_{n+2} = x_{n+1} + x_n, \quad x_0 = 0, \quad x_1 = 1.$$

(a) Вычислите отношение x_{n+1}/x_n для n от 1 до 8. Имеет ли эта последовательность предел?

(b) Найдите такую матрицу A, чтобы выполнялось равенство

$$\begin{pmatrix}x_{n+2}\\x_{n+1}\end{pmatrix} = A\begin{pmatrix}x_{n+1}\\x_n\end{pmatrix}.$$

С помощью матрицы A выразите $\begin{pmatrix} x_{n+2} \\ x_{n+1} \end{pmatrix}$ через $\begin{pmatrix} x_1 \\ x_0 \end{pmatrix}$.

(c) Получите явное выражение для x_n через x_1 и x_0. (Подсказка: приведите матрицу A к диагональному виду.)

(d) Покажите, что существует предел $\lim\limits_{n\to\infty} x_{n+1}/x_n$ и вычислите его.

(e) Используя величину этого предела, что можно сказать о бесконечной дроби

$$1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}}}?$$

(f) Существуют ли такие x_0 и x_1, что $\lim\limits_{n\to\infty} x_{n+1}/x_n$ отличается от значения, полученного в упражнении (d)?

Глава 3

Линейные дифференциальные уравнения на плоскости

Основная цель этой главы — объяснить, что система однородных дифференциальных уравнений с постоянными коэффициентами может быть записана в форме $d\mathbf{u}/dt = A\mathbf{u}$, где A — матрица и $\mathbf{u}$ — вектор, и что решение этого уравнения имеет вид $e^{At}\mathbf{u}_0$, где $\mathbf{u}_0$ определяется начальными условиями. При этом, конечно, надо объяснить, что такое экспонента, показателем которой является матрица. Кроме того, обсуждается качественное поведение решений в неоднородном случае, в том числе явление резонанса.

3.1. Функции матриц

Обсуждая теорему Гамильтона–Кэли, мы уже встречались с матричным полиномом. В общем случае, пусть $Q(X)$ — любой полином. Если

$$Q(X) = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0,$$

то можно определить матрицу $Q(F)$ как

$$Q(F) = a_n F^n + a_{n-1} F^{n-1} + \cdots + a_1 F + a_0.$$

Мы можем перемножить два полинома и получить третий полином: $(Q_1 Q_2)(X) = Q_1(X) Q_2(X)$. Аналогично, для матриц

$$(Q_1 Q_2)(F) = Q_1(F) Q_2(F).$$

Здесь не возникает проблемы из-за некоммутативности матричного умножения, так как степени фиксированной матрицы всегда коммутируют друг с другом в соответствии с ассоциативным законом:

$$F^k F^l = F^{k+l} = F^l F^k.$$

Аналогично,

$$(Q_1 + Q_2)(F) = Q_1(F) + Q_2(F).$$

Короче говоря, при вычислении матричного полинома действуют обычные законы алгебры.

Мы хотим перейти к изучению функций от матриц более общего вида. С этой целью сделаем небольшое отклонение и поговорим о степенных рядах.

Выражение вида

$$R(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n + \cdots ,$$

где коэффициенты $a_i, i = 0, 1, \ldots$ — вещественные числа, а X — некий символ, называется *формальным степенным рядом*. Сложение степенных рядов производится по правилу

$$\begin{aligned}(a_0 + a_1 X + a_2 X^2 + \cdots) &+ (b_0 + b_1 X + b_2 X^2 + \cdots)\\ &= (a_0 + b_0) + (a_1 + b_1)X + (a_2 + b_2)X^2 + \cdots ,\end{aligned}$$

т. е. мы складываем коэффициенты при одинаковых степенях X. Для умножения степенных рядов воспользуемся правилом $X^k X^l = X^{k+l}$. Тогда, собирая члены с одинаковыми степенями X, получаем

$$\begin{aligned}(a_0 + a_1 X + a_2 X^2 + \cdots)&(b_0 + b_1 X + b_2 X^2 + \cdots)\\ &= a_0 b_0 + (a_1 b_0 + a_0 b_1)X + (a_2 b_0 + a_1 b_1 + a_0 b_2)X^2 + \cdots .\end{aligned}$$

Таким образом, например,

$$(1 + X + X^2 + \cdots)(1 + X + X^2 + \cdots) = 1 + 2X + 3X^2 + \cdots .$$

Легко проверить, что обычные свойства сложения и умножения полиномов сохраняются и для степенных рядов.

Пусть t — любое вещественное число. По определению, формальный степенной ряд для $\exp(tX)$ имеет вид

$$\exp(tX) = 1 + tX + \frac{1}{2!}t^2X^2 + \frac{1}{3!}t^3X^3 + \frac{1}{4!}t^4X^4 + \cdots . \tag{3.1}$$

Тогда

$$\begin{aligned}\exp(sX)\exp(tX) &= \left(1+sX+\frac{1}{2!}s^2X^2+\cdots\right)\left(1+tX+\frac{1}{2!}t^2X^2+\cdots\right)\\ &= 1 + (s+t)X + \frac{1}{2!}(s^2+2st+t^2)X^2 + \cdots ,\end{aligned}$$

где коэффициент при X^n определяется формулой

$$\frac{1}{n!}\left(s^n + ns^{n-1}t + n\frac{(n-1)}{2}s^{n-2}t^2 + \cdots + t^n\right),$$

которая, по формуле бинома Ньютона, равна просто $(1/n!)(s+t)^n$. Так как по определению

$$1 + (s+t)X + \frac{1}{2!}(s+t)^2X^2 + \frac{1}{3!}(s+t)^3X^3 + \cdots = \exp(s+t)X,$$

то можно сделать вывод, что равенство

$$\exp(sX)\exp(tX) = \exp(s+t)X \tag{3.2}$$

выполняется тождественно для формальных степенных рядов.

В отличие от полиномов, мы в общем случае не можем вычислить $R(X)$ для какого-нибудь числа r или матрицы F. Действительно, если в формальный степенной ряд

$$R(X) = a_0 + a_1X + a_2X^2 + \cdots$$

вместо символа X подставить вещественное число r, то получится бесконечная сумма

$$a_0 + a_1r + a_2r^2 + \cdots ,$$

которая в таком виде не имеет смысла. Один способ придать такой сумме смысл сводится к обрезанию бесконечного ряда на каком-то слагаемом. Может оказаться, что такая конечная сумма не будет сильно зависеть от места обрыва ряда, если мы оставим достаточное количество слагаемых. Величину, полученную как «предельное значение» конечной суммы, обозначим $R(r)$. Объясним эту процедуру более аккуратно. Пусть из бесконечной суммы мы оставили M слагаемых. Обозначим $R^M(r)$ соответствующую конечную сумму:

$$R^M(r) = a_0 + a_1 r + a_2 r^2 + \cdots + a_M r^M.$$

Мы будем говорить, что степенной ряд $R(X)$ *сходится* для данного числа r, если при сколь угодно малом положительном ε можно найти достаточно большое число M_0, такое, что для любых целых M и N, бóльших M_0, выполняется неравенство

$$|R^M(r) - R^N(r)| < \varepsilon.$$

Другими словами, если мы учтем достаточно большое число слагаемых, то все значения $R^M(r)$ будут отличаться на малое значение ε. Чем больше членов мы учитываем, тем ближе конечная сумма $R^M(r)$ к предельному значению, обозначаемому $R(r)$[1].

Точно такое же определение можно ввести для матриц. Для любой матрицы F конечная сумма

$$R^M(F) = a_0 + a_1 F + \cdots + a_M F^M$$

имеет смысл. Разность $R^M(F) - R^N(F)$ — тоже матрица. Тогда условие

$$|R^M(F) - R^N(F)| < \varepsilon$$

означает, что все матричные элементы матрицы $R^M(F) - R^N(F)$ по модулю меньше ε. Итак, мы говорим, что $R(X)$ сходится для

[1]Авторы проявляют здесь некоторую непоследовательность. Предположив в начале книги, что читатели владеют одномерным математическим анализом, было бы естественно считать, что они знают понятие предела числовой последовательности. Тогда сумму $R(r)$ ряда можно определить как $\lim\limits_{M\to\infty} R^M(r)$, если, конечно, этот предел существует. Рассуждения авторов здесь нечетким образом используют «критерий Коши» (теорему о равносильности сходимости последовательности и ее фундаментальности). Кстати, предел последовательности чисел (и матриц!) уже использовался авторами в главе 2. — *Прим. ред.*

матрицы F, если при любом $\varepsilon > 0$ существует такое значение M_0, что условие $|R^M(F) - R^N(F)| < \varepsilon$ выполняется для всех M и N, бóльших M_0. В этом случае каждый матричный элемент $R^M(F)$ стремится к предельному значению[2]. Таким образом, мы получаем матрицу предельных значений матричных элементов, которую обозначаем $R(F)$.

Очевидно, что если $R_1(X)$, $R_2(X)$ и $R_3(X)$ — формальные степенные ряды, для которых справедливо равенство

$$R_1(X)R_2(X) = R_3(X),$$

и, если все три степенных ряда сходятся для матрицы F, то

$$R_1(F)R_2(F) = R_3(F).^3$$

Аналогично для сложения: если $R_1(X) + R_2(X) = R_3(X)$, и все ряды[4] сходятся для матрицы F, то $R_1(F) + R_2(F) = R_3(F)$.

3.2. Экспонента от матрицы

Мы уже ввели формулой (3.1) формальный степенной ряд для экспоненциальной функции и показали, что для нее выполняется равенство (3.2). Сейчас докажем, что степенной ряд для $\exp(tX)$ будет сходиться в случае, когда символ X заменяется матрицей 2×2, обозначаемой буквой A. В качестве первого шага проведем доказательство того, что ряд сходится абсолютно[5], если вместо X подставить произвольное вещественное число k. Рассмотрим степенной ряд

$$\exp y = 1 + y + \frac{1}{2!}y^2 + \frac{1}{3!}y^3 + \cdots, \tag{3.3}$$

[2]Это следует из упомянутого в предыдущем примечании критерия Коши. — *Прим. ред.*

[3]Вряд ли стоит называть очевидным утверждение, являющееся следствием одной тонкой теоремы Абеля. — *Прим. ред.*

[4]Здесь достаточно предположить сходимость первых двух рядов. — *Прим. ред.*

[5]Говорят, что ряд сходится *абсолютно*, если сходится другой ряд, составленный из модулей слагаемых первого ряда. В этом случае первый ряд тоже сходится. — *Прим. ред.*

где $y = |tk|$. При положительных значениях y все слагаемые этого ряда положительны. Если мы покажем, что ряд (3.3) сходится, будет доказана *абсолютная* сходимость ряда для $\exp tk$.

Для доказательства сходимости ряда (3.3) покажем, что фрагмент $r_{m,n}$, получающийся при суммировании членов с номерами от m до $(m+n)$, может быть сколь угодно малым, если выбрано достаточное большое значение m. Это легко сделать, сравнивая фрагмент

$$r_{m,n} = \frac{1}{m!}y^m + \frac{1}{m!\,(m+1)}y^{m+1} + \frac{1}{m!\,(m+1)(m+2)}y^{m+2} + \cdots$$
$$+ \frac{1}{m!\,(m+1)\cdots(m+n)}y^{m+n}$$

с геометрической прогрессией

$$s_{m,n} = \frac{1}{m!}y^m + \frac{1}{m!\,m}y^{m+1} + \frac{1}{m!\,m^2}y^{m+2} + \cdots + \frac{1}{m!\,m^n}y^{m+n}.$$

Очевидно, что $r_{m,n} < s_{m,n}$ для всех m. Сумму $s_{m,n}$ геометрической прогрессии можно вычислить явно:

$$\begin{aligned} s_{m,n} &= \frac{y^m}{m!}\left(1 + \frac{y}{m} + \left(\frac{y}{m}\right)^2 + \left(\frac{y}{m}\right)^3 + \cdots + \left(\frac{y}{m}\right)^n\right) \\ &= \left(1 - \left(\frac{y}{m}\right)^{n+1}\right) s_m, \end{aligned}$$

где

$$s_m = \frac{y^m}{m!}\,\frac{1}{1 - y/m}.$$

Итак, мы доказали, что

$$r_{m,n} < s_m.$$

Предположим, что мы выбрали $m > 2y$. В этом случае

$$\frac{1}{1 - y/m} < \frac{1}{1 - 1/2} = 2.$$

Тогда $s_m < 2y^m/m!$. Каждый раз, увеличивая m на единицу, мы умножаем s_m на множитель, меньший $1/2$. Очевидно, что выбрав достаточно большое значение m, можно сделать величину s_m сколь угодно малой. А так как $r_{m,n} < s_m$, то и значение $r_{m,n}$ может быть сделано сколь угодно малым. А это означает, что ряд

$$\exp tk = 1 + tk + \frac{1}{2!}(tk)^2 + \frac{1}{3!}(tk)^3 + \cdots$$

сходится абсолютно. Между прочим, хорошо известное доказательство абсолютной сходимости степенного ряда

$$a_0 + a_1 y + a_2 y^2 + a_3 y^3 + \cdots$$

с учетом оценки предела

$$\lim_{m\to\infty} \frac{a_{m+1}y^{m+1}}{a_m y^m} < 1$$

основывается на тех же аргументах, которые мы только что использовали[6].

Перейдем теперь к матричному случаю. Возьмем матрицу $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, в которой каждый элемент по модулю меньше, чем $k/2$. Матричные элементы матрицы $A^2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ состоят из двух слагаемых, причем эти слагаемые по модулю меньше, чем $(k/2)^2 = k^2/4$. Таким образом, матричные элементы A^2 по модулю меньше, чем $k^2/2$. Легко показать, что все матричные элементы A^3 по модулю меньше, чем $k^3/2$, и далее, по индукции, *элементы матрицы* A^m по модулю меньше, чем $k^m/2$. Теперь, если мы будем суммировать ряд

$$\exp(tA) = \mathbb{I} + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \cdots,$$

то все слагаемые в четырех матричных элементах получающейся матрицы меньше, чем соответствующие слагаемые ряда

$$1 + t\frac{k}{2} + \frac{t^2}{2!}\frac{k^2}{2} + \frac{t^3}{3!}\frac{k^3}{2} + \cdots.$$

[6]В русскоязычной литературе последний прием составляет содержание признака Даламбера. — *Прим. ред.*

Из этого следует, что для любого вещественного числа t и любой матрицы A (2×2) ряд $\exp(tA)$ сходится. Используя те же аргументы, заменив только $k/2$ на k/n, можно показать, что экспоненциальный ряд сходится и в случае, когда матрица A имеет размер $n \times n$.

Итак, для экспоненциальной функции от матрицы мы получили фундаментальное тождество

$$\exp(s+t)A = \exp(sA)\exp(tA).$$

Вы можете спросить: а как обстоит дело с тождеством

$$\exp(A+B) \stackrel{?}{=} (\exp A)(\exp B),$$

где A и B — произвольные матрицы? Чтобы понять, что происходит в этом случае, разложим в ряд обе части этого равенства:

$$\begin{aligned}\exp(A+B) &= \mathbb{I} + A + B + \frac{1}{2}(A+B)^2 + \cdots \\ &= \mathbb{I} + A + B + \frac{1}{2}(A^2 + AB + BA + B^2) + \cdots\end{aligned}$$

и

$$\begin{aligned}(\exp A)(\exp B) &= \left(\mathbb{I} + A + \frac{1}{2}A^2 + \cdots\right)\left(\mathbb{I} + B + \frac{1}{2}B^2 + \cdots\right) \\ &= \mathbb{I} + A + B + \left(\frac{1}{2}A^2 + AB + \frac{1}{2}B^2\right) + \cdots,\end{aligned}$$

где многоточия обозначают слагаемые более высоких степеней по A и B. Если мы сравним слагаемые второй степени в этих двух выражениях, то увидим, что они равны *только* при условии

$$\frac{1}{2}(AB+BA) = AB,$$

т. е. если

$$AB = BA.$$

Итак, если матрицы A и B не коммутируют, то нет оснований ожидать, что выполняется тождество

$$\exp(A+B) = (\exp A)(\exp B),$$

и на самом деле оно не всегда верно. Рассмотрим конкретный пример. Возьмем матрицу

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

В этом случае $A^2 = 0$ и, следовательно, все слагаемые более высоких степеней в разложении $\exp A$ равны нулю. Получаем простую формулу:

$$\exp A = \mathbb{I} + A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Возьмем теперь другую матрицу $B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, для которой тоже $B^2 = 0$. Тогда

$$\exp B = \mathbb{I} + B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

А теперь перемножим две получившиеся матрицы:

$$(\exp A)(\exp B) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

С другой стороны, $A+B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ и, соответственно, $(A+B)^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Отсюда получаем, что

$$\begin{aligned} \exp(A+B) &= \mathbb{I} + (A+B) + \frac{1}{2}\mathbb{I} + \frac{1}{3!}(A+B) + \frac{1}{4!}\mathbb{I} + \cdots \\ &= \left(1 + \frac{1}{2!} + \frac{1}{4!} + \cdots\right)\mathbb{I} + \left(1 + \frac{1}{3!} + \frac{1}{5!} + \cdots\right)(A+B). \end{aligned}$$

Множители в скобках можно упростить[7]

$$1 + \frac{1}{2!} + \frac{1}{4!} + \cdots = \frac{1}{2}(e + e^{-1})$$

---

[7]Это делается с использованием тождеств

$$1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots = \exp 1 = e,$$
$$1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots = \exp(-1) = e^{-1}.$$

и

$$1 + \frac{1}{3!} + \frac{1}{5!} + \frac{1}{7!} + \cdots = \frac{1}{2}(e - e^{-1}).$$

Окончательно, получаем, что

$$\exp(A+B) = \frac{1}{2}\begin{pmatrix} e + e^{-1} & e - e^{-1} \\ e - e^{-1} & e + e^{-1} \end{pmatrix} \neq (\exp A)(\exp B).$$

Справедливость формулы

$$\exp(s+t)A = \exp(sA)\exp(tA)$$

связана с тем, что матрицы sA и tA коммутируют.

Показав, что $\exp(tA)$ определяется степенным рядом, мы теперь обобщим на матричный случай хорошо известную формулу

$$\frac{d}{dt}[\exp(tk)] = k\exp(tk). \tag{3.4}$$

Определим производную функции $\exp(tA)$ по вещественной переменной t следующим образом

$$\frac{d}{dt}[\exp(tA)] = \lim_{h\to 0}\frac{1}{h}[\exp((t+h)A) - \exp(tA)].$$

Предел в правой части равенства означает матрицу, элементы которой являются пределами соответствующих элементов зависящей от t матрицы, стоящей под знаком $\lim\limits_{h\to 0}$. Поскольку

$$\exp((t+h)A) = \exp(hA)\exp(tA),$$

получаем

$$\frac{d}{dt}[\exp(tA)] = \lim_{h\to 0}\frac{1}{h}[\exp(hA) - \mathbb{I}]\exp(tA).$$

В то же время

$$\exp(hA) - \mathbb{I} = hA + \frac{h^2A^2}{2!} + \frac{h^3A^3}{3!} + \cdots,$$

следовательно,

$$\frac{1}{h}(\exp(hA) - \mathbb{I}) = A + \frac{hA^2}{2!} + \frac{h^2A^3}{3!} + \cdots$$

и

$$\lim_{h\to 0}\frac{1}{h}[\exp(hA)-\mathbb{I}]=A.$$[8]

$$\frac{d}{dt}[\exp(tA)]=A\exp(tA).$$

Эта формула похожа на (3.4), но справа подразумевается матричное умножение.

Выберем на плоскости фиксированный вектор $\mathbf{v}_0=\begin{pmatrix}x_0\\y_0\end{pmatrix}$ и определим зависящий от времени вектор $\mathbf{v}(t)$ формулой

$$\mathbf{v}(t)=\exp(tA)\mathbf{v}_0,\quad \text{где}\quad A=\begin{pmatrix}a & b\\ c & d\end{pmatrix}.$$

Определим производную по времени $\dot{\mathbf{v}}(t)$ вектора $\mathbf{v}(t)$:

$$\dot{\mathbf{v}}(t)=\lim_{h\to 0}\frac{1}{h}[\mathbf{v}(t+h)-\mathbf{v}(t)].$$

Если $\mathbf{v}(t)=\begin{pmatrix}x(t)\\y(t)\end{pmatrix}$, то производная этого вектора по времени, очевидно, может быть записана в виде $\dot{\mathbf{v}}(t)=\begin{pmatrix}\dot{x}(t)\\\dot{y}(t)\end{pmatrix}$. Так как $\mathbf{v}_0$ — постоянный вектор, то

$$\dot{\mathbf{v}}(t)=\lim_{h\to 0}\frac{1}{h}[\exp((t+h)A)-\exp(tA)]\mathbf{v}_0,$$

т. е.

$$\dot{\mathbf{v}}(t)=\frac{d}{dt}\exp(tA)\mathbf{v}_0.$$

Отсюда

$$\dot{\mathbf{v}}(t)=A\exp(tA)\mathbf{v}_0=A\mathbf{v}(t).$$

Итак, мы показали, что $\mathbf{v}(t)$ удовлетворяет дифференциальному уравнению

$$\dot{\mathbf{v}}(t)=A\mathbf{v}(t),\quad \mathbf{v}(0)=\mathbf{v}_0.$$

[8]Здесь следовало сослаться на теорему о почленном переходе к пределу, которая требует проверки некоторых условий, например, равномерной сходимости ряда (это понятие авторы даже не упоминают). — *Прим. ред.*

Записав, как и выше, $\mathbf{v}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$, получаем

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix},$$

т. е.

$$\begin{aligned} \dot{x}(t) &= ax(t) + by(t), \\ \dot{y}(t) &= cx(t) + dy(t). \end{aligned}$$

Итак, формула

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \exp(tA) \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

дает решение *системы линейных дифференциальных уравнений первого порядка*, написанной выше.

Легко доказать, что *любое* решение дифференциального уравнения $\dot{\mathbf{v}}(t) = A\mathbf{v}(t)$ имеет вид $\mathbf{v}(t) = \exp(tA)\mathbf{v}_0$. Для доказательства возьмем вектор $\mathbf{w}(t) = \exp(-tA)\mathbf{v}(t)$. Тогда[9]

$$\dot{\mathbf{w}}(t) = \frac{d}{dt}(\exp(-tA))\mathbf{v}(t) + \exp(-tA)\dot{\mathbf{v}}(t).$$

В то же время

$$\frac{d}{dt}\exp(-tA) = -A\exp(-tA)$$

и, кроме того,

$$\dot{\mathbf{v}}(t) = A\mathbf{v}(t).$$

Следовательно,

$$\dot{\mathbf{w}}(t) = -A\exp(-tA)\mathbf{v}(t) + \exp(-tA)A\mathbf{v}(t) = \mathbf{0},$$

так как матрицы A и $\exp(-tA)$ коммутируют. Из этого следует, что $\mathbf{w}$ — постоянный вектор (обозначенный $\mathbf{v}_0$), и мы получаем

$$\mathbf{v}_0 = \exp(-tA)\mathbf{v}(t)$$

[9]Следующая далее выкладка использует упоминавшуюся ранее формулу дифференцирования переменного вектора $\mathbf{w}(t)$, заданного как образ другого переменного вектора $\mathbf{v}(t)$ при преобразовании, также зависящем от t. — *Прим. ред.*

или

$$\mathbf{v}(t) = \exp(tA)\mathbf{v}_0.$$

Итак, мы показали, что последняя формула с экспонентой $\exp(tA)$ определяет *общее решение* системы дифференциальных уравнений с постоянными коэффициентами. Поэтому важно обсудить различные методы вычисления $\exp(tA)$.

3.3. Вычисление экспоненты от матрицы

Предположим, что матрицы F и G связаны между собой соотношением

$$F = BGB^{-1}.$$

Тогда при любом k

$$F^k = BG^kB^{-1}$$

и из разложения $\exp(tA)$ в степенной ряд следует, что

$$\exp(tF) = B\exp(tG)B^{-1}.$$

Случай 1. Матрица F имеет разные вещественные собственные значения. Мы уже научились приводить матрицу 2×2 к диагональному виду. Допустим, что матрица F имеет разные вещественные собственные значения λ_1 и λ_2. Тогда можно записать, что при некоторой матрице B

$$F = B\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}B^{-1}.$$

Очевидно, что

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}^n = \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix},$$

и из разложения экспоненциальной функции в степенной ряд следует, что

$$\exp\begin{pmatrix} \lambda_1 t & 0 \\ 0 & \lambda_2 t \end{pmatrix} = \begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix}.$$

Тогда

$$\exp(tF) = B\begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix}B^{-1}.$$

Приведем конкретный пример. Характеристический полином матрицы $F = \begin{pmatrix} 7 & 4 \\ -8 & -5 \end{pmatrix}$ имеет вид $\lambda^2 - 2\lambda - 3 = 0$, откуда получаем собственные значения $\lambda_1 = 3$, $\lambda_2 = -1$. Рассматривая $(F - 3\mathbb{I}) = \begin{pmatrix} 4 & 4 \\ -8 & -8 \end{pmatrix}$, найдем собственные векторы $\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ (ядро матрицы $F - 3\mathbb{I}$) и $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ (образ матрицы $F - 3\mathbb{I}$). Таким образом, можно взять

$$B = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \quad B^{-1} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

и записать матрицу F в виде

$$F = \begin{pmatrix} -1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Отсюда следует, что

$$\exp(tF) = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} e^{3t} & 0 \\ 0 & e^{-t} \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Перемножив эти матрицы, получаем

$$\exp(tF) = \begin{pmatrix} 2e^{3t} - e^{-1} & e^{3t} - e^{-t} \\ -2e^{3t} + 2e^{-t} & -e^{3t} + 2e^{-t} \end{pmatrix}.$$

Выбрав вектор $\mathbf{v}_0$, определяющий *начальные условия*, т. е. значение $\mathbf{v}$ в момент $t = 0$, мы можем написать частное решение уравнения $\dot{\mathbf{v}}(t) = F\mathbf{v}(t)$. Пусть, например, $\mathbf{v}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Тогда

$$\mathbf{v}(t) = \exp(tF) \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3e^{3t} - 2e^{-t} \\ -3e^{3t} + 4e^{-t} \end{pmatrix}.$$

Дифференцирование этого вектора дает

$$\dot{\mathbf{v}}(t) = \begin{pmatrix} 9e^{3t} + 2e^{-t} \\ -9e^{3t} - 4e^{-t} \end{pmatrix}.$$

Прямым вычислением легко проверяется равенство

$$F\mathbf{v}(t) = \begin{pmatrix} 7 & 4 \\ -8 & -5 \end{pmatrix} \begin{pmatrix} 3e^{3t} - 2e^{-t} \\ -3e^{3t} + 4e^{-t} \end{pmatrix} = \begin{pmatrix} 9e^{3t} + 2e^{-t} \\ -9e^{3t} - 4e^{-t} \end{pmatrix},$$

т. е. справедливость уравнения $\dot{\mathbf{v}} = F\mathbf{v}$.

Случай 2. Одинаковые собственные значения. Метод, описанный выше, работает, когда F имеет *разные* собственные значения. Допустим теперь, что F имеет два одинаковых собственных значения λ. Тогда либо $F = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$, когда можно писать $\exp(tF) = \begin{pmatrix} e^{\lambda t} & 0 \\ 0 & e^{\lambda t} \end{pmatrix}$, либо $F = B \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} B^{-1}$.

Пусть в показателе экспоненты стоит матрица $\begin{pmatrix} \lambda t & t \\ 0 & \lambda t \end{pmatrix}$. Сначала преобразуем ее к виду

$$\begin{pmatrix} \lambda t & t \\ 0 & \lambda t \end{pmatrix} = \lambda t \mathbb{I} + \begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix}.$$

Если матрицы C и D коммутируют, то, как мы знаем,

$$\exp(C + D) = (\exp C)(\exp D).$$

Поскольку $\lambda \mathbb{I}$ коммутирует с любой матрицей, мы имеем

$$\exp \begin{pmatrix} \lambda t & t \\ 0 & \lambda t \end{pmatrix} = \exp(\lambda t \mathbb{I}) \exp \begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix}.$$

В то же время $\begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, поэтому степенной ряд для соответствующей экспоненты сводится к

$$\exp \begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix} t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix} + \text{слагаемые, равные нулю},$$

т. е. $\exp \begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ и, следовательно,

$$\exp \begin{pmatrix} \lambda t & t \\ 0 & \lambda t \end{pmatrix} = \begin{pmatrix} e^{\lambda t} & 0 \\ 0 & e^{\lambda t} \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{\lambda t} & te^{\lambda t} \\ 0 & e^{\lambda t} \end{pmatrix}.$$

Итак, мы получили, что

$$\text{если} \quad F = B\begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix}B^{-1}, \quad \text{то} \quad \exp(tF) = B\begin{pmatrix}e^{\lambda t} & te^{\lambda t}\\ 0 & e^{\lambda t}\end{pmatrix}B^{-1}.$$

В качестве примера рассмотрим систему уравнений

$$\begin{aligned}\dot{x} &= x + y,\\ \dot{y} &= -x + 3y.\end{aligned}$$

Матрица $F = \begin{pmatrix}1 & 1\\ -1 & 3\end{pmatrix}$ имеет характеристическое уравнение $\lambda^2 - 4\lambda + 4 = 0$, имеющее два равных корня $\lambda = 2$. Рассматривая $(F - 2\mathbb{I}) = \begin{pmatrix}-1 & 1\\ -1 & 1\end{pmatrix}$, получаем собственный вектор $\mathbf{v}_1 = \begin{pmatrix}1\\1\end{pmatrix}$ и вектор $\mathbf{v}_2 = \begin{pmatrix}0\\1\end{pmatrix}$, для которого $(F - 2\mathbb{I})\mathbf{v}_2 = \mathbf{v}_1$. Отсюда

$$F = \begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}\begin{pmatrix}2 & 1\\ 0 & 2\end{pmatrix}\begin{pmatrix}1 & 0\\ -1 & 1\end{pmatrix}.$$

Поскольку $\exp\begin{pmatrix}2t & t\\ 0 & 2t\end{pmatrix} = \begin{pmatrix}e^{2t} & te^{2t}\\ 0 & e^{2t}\end{pmatrix}$, мы имеем

$$\exp(tF) = \begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}\begin{pmatrix}e^{2t} & te^{2t}\\ 0 & e^{2t}\end{pmatrix}\begin{pmatrix}1 & 0\\ -1 & 1\end{pmatrix}$$

или

$$\exp(tF) = \begin{pmatrix}(1-t)e^{2t} & te^{2t}\\ -te^{2t} & (1+t)e^{2t}\end{pmatrix}.$$

Если, например, мы хотим найти решение нашей системы дифференциальных уравнений с начальными условиями $x_0 = 2$, $y_0 = 1$, то следует построить вектор

$$\exp(tF)\begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}2e^{2t} - te^{2t}\\ e^{2t} - te^{2t}\end{pmatrix}.$$

Легко проверить, что

$$\begin{aligned}\dot{x} &= 4e^{2t} - e^{2t} - 2te^{2t} = x + y,\\ \dot{y} &= 2e^{2t} - e^{2t} - 2te^{2t} = -x + 3y.\end{aligned}$$

На самом деле в случае равных собственных значений λ необязательно записывать матрицу F в виде $F = B\begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix}B^{-1}$. Согласно теореме Гамильтона–Кэли $(F - \lambda\mathbb{I})^2 = 0$, поэтому матрица $G = F - \lambda\mathbb{I}$ является нильпотентом. Поскольку $\lambda\mathbb{I}$ и G коммутируют, из равенства $F = G + \lambda\mathbb{I}$ следует:

$$\exp(tF) = \exp(\lambda t\mathbb{I})\exp(tG).$$

Но $\exp(tG)$ легко вычисляется разложением в степенной ряд:

$$\exp(tG) = \mathbb{I} + tG,$$

поскольку $(tG)^2$ и все последующие слагаемые ряда равны нулю. Следовательно,

$$\exp(tF) = \begin{pmatrix}e^{\lambda t} & 0\\ 0 & e^{\lambda t}\end{pmatrix}(\mathbb{I} + tG).$$

В предыдущем примере, где $F = \begin{pmatrix}1 & 1\\ -1 & 3\end{pmatrix}$, получаем

$$G = F - 2\mathbb{I} = \begin{pmatrix}-1 & 1\\ -1 & 1\end{pmatrix}$$

и

$$\exp(tF)=\begin{pmatrix}e^{2t} & 0\\ 0 & e^{2t}\end{pmatrix}\left[\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}+\begin{pmatrix}-t & t\\ -t & t\end{pmatrix}\right]=\begin{pmatrix}(1-t)e^{2t} & te^{2t}\\ -te^{2t} & (1+t)e^{2t}\end{pmatrix},$$

что и следовало ожидать.

Случай 3. Комплексно-сопряженные собственные значения. Ранее мы доказали, что если матрица F имеет комплексно-сопряженные собственные значения $\alpha \pm \mathrm{i}\beta$, то ее можно представить в форме $F = BCB^{-1}$, где $C = \begin{pmatrix}\alpha & -\beta\\ \beta & \alpha\end{pmatrix}$ — конформная матрица. Перед нами стоит задача получить экспоненту, в показателе которой стоит C.

Заметим, что конформная матрица C может быть записана в виде $C = \alpha\mathbb{I} + \beta J$, где $J = \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$. Матрица J удовлетворяет уравнению $J^2 = -\mathbb{I}$ и соответствует комплексному числу i. Так как $\alpha\mathbb{I}$ и βJ коммутируют, мы имеем равенство

$$\exp(tC) = \exp(t\alpha\mathbb{I})\exp(t\beta J),$$

где

$$\exp(t\alpha\mathbb{I}) = \begin{pmatrix} e^{\alpha t} & 0 \\ 0 & e^{\alpha t} \end{pmatrix}.$$

Чтобы вычислить $\exp(t\beta J)$, воспользуемся разложением в степенной ряд:

$$\exp(t\beta J) = \mathbb{I} + t\beta J + \frac{1}{2!}(t\beta J)^2 + \frac{1}{3!}(t\beta J)^3 + \frac{1}{4!}(t\beta J)^4 + \cdots .$$

Поскольку $J^2 = -\mathbb{I}$, $J^3 = -J$, $J^4 = \mathbb{I}$, $J^5 = J$ и т. д., получаем

$$\begin{aligned}\exp(t\beta J) &= \mathbb{I} + t\beta J - \frac{1}{2!}(t\beta)^2\mathbb{I} - \frac{1}{3!}(t\beta)^3 J + \frac{1}{4!}(t\beta)^4\mathbb{I} + \cdots \\ &= \left(1 - \frac{(t\beta)^2}{2!} + \frac{(t\beta)^4}{4!} + \cdots\right)\mathbb{I} + \left(t\beta - \frac{(t\beta)^3}{3!} + \cdots\right)J.\end{aligned}$$

Воспользуемся тем, что коэффициент при матрице $\mathbb{I}$ в последнем выражении — степенной ряд для $\cos\beta t$, а коэффициент при матрице J — степенной ряд для $\sin\beta t$. Следовательно, можно записать

$$\exp(t\beta J) = \cos\beta t\ \mathbb{I} + \sin\beta t\ J = \begin{pmatrix} \cos\beta t & -\sin\beta t \\ \sin\beta t & \cos\beta t \end{pmatrix}.$$

Полученная матрица является матрицей вращения, зависящей от времени. Отождествляя матрицу J с комплексным числом i, мы можем получить известную формулу

$$e^{i\beta t} = \cos\beta t + i\sin\beta t.$$

Итак, если $C = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$, то

$$\exp(tC) = \begin{pmatrix} e^{\alpha t} & 0 \\ 0 & e^{\alpha t} \end{pmatrix}\begin{pmatrix} \cos\beta t & -\sin\beta t \\ \sin\beta t & \cos\beta t \end{pmatrix}.$$

И если $F = BCB^{-1}$, то $\exp(tF)$ можно вычислить по формуле

$$\exp(tF) = B\exp(tC)B^{-1}.$$

В качестве примера возьмем матрицу $\begin{pmatrix} 3 & -10 \\ 2 & -5 \end{pmatrix}$. Характеристическое уравнение этой матрицы $\lambda^2 + 2\lambda + 5 = 0$ имеет корни $\lambda = -1 \pm 2\mathrm{i}$. Тогда $F = BCB^{-1}$, где $C = \begin{pmatrix} -1 & -2 \\ 2 & -1 \end{pmatrix}$ — конформная матрица. Как описано в параграфе 2.2, в качестве первого столбца матрицы B можно взять $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, тогда второй столбец B равняется первому столбцу $F - \alpha\mathbb{I}$, деленному на β, т. е. столбцу $\frac{1}{2}\begin{pmatrix} 3+2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Следовательно, $B = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ и $F = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -1 & -2 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$. Воспользовавшись изложенной выше процедурой, вычислим $\exp(tC)$:

$$\exp(tC) = \begin{pmatrix} e^{-t} & 0 \\ 0 & e^{-t} \end{pmatrix}\begin{pmatrix} \cos 2t & -\sin 2t \\ \sin 2t & \cos 2t \end{pmatrix}.$$

Из равенства $\exp(tF) = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\exp(tC)\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$, перемножая матрицы, получаем

$$\exp(tF) = \begin{pmatrix} e^{-t}(\cos 2t + 2\sin 2t) & -5e^{-t}\sin 2t \\ e^{-t}\sin 2t & e^{-t}(\cos 2t - 2\sin 2t) \end{pmatrix}.$$

Можете самостоятельно проверить, что

$$\frac{d}{dt}\exp(tF) = F\exp(tF).$$

И в этом случае нет необходимости явно выписывать разложение на множители $F = BCB^{-1}$. Пусть F имеет собственные значения $\alpha \pm \mathrm{i}\beta$, так что характеристическое уравнение имеет вид

$$(\lambda - \alpha)^2 + \beta^2 = 0.$$

Согласно теореме Гамильтона–Кэли

$$(F - \alpha\mathbb{I})^2 + \beta^2\mathbb{I} = 0,$$

поэтому матрица $G = F - \alpha\mathbb{I}$, след которой равен нулю, удовлетворяет уравнению $G^2 = -\beta^2\mathbb{I}$. Из равенства $F = \alpha\mathbb{I} + G$ получаем, что

$$\exp(tF) = \exp(\alpha t\mathbb{I})\exp(tG).$$

Экспоненту, в показателе которой стоит tG, запишем в виде ряда

$$\exp(tG) = \mathbb{I} + tG + \frac{t^2G^2}{2!} + \frac{t^3G^3}{3!} + \frac{t^4G^4}{4!} + \cdots .$$

Используя уравнение $G^2 = -\beta^2\mathbb{I}$, получаем

$$\exp(tG) = \mathbb{I} + tG - \frac{\beta^2t^2}{2!}\mathbb{I} - \frac{\beta^3t^3G}{3!} + \frac{\beta^4t^4}{4!}\mathbb{I} + \cdots .$$

Коэффициент при матрице $\mathbb{I}$ — это степенной ряд для $\cos\beta t$, а коэффициент при матрице G равен

$$t - \frac{\beta^2t^3}{3!} + \frac{\beta^4t^5}{5!} = \frac{1}{\beta}\left(\beta t - \frac{\beta^3t^3}{3!} + \frac{\beta^5t^5}{5!} + \cdots\right) = \frac{\sin\beta t}{\beta}.$$

Следовательно,

$$\exp(tG) = \cos\beta t\,\mathbb{I} + [(\sin\beta t)/\beta]G,$$

и

$$\exp(tF) = \begin{pmatrix} e^{\alpha t} & 0 \\ 0 & e^{\alpha t} \end{pmatrix}\exp(tG).$$

Возвращаясь к матрице $F = \begin{pmatrix} 3 & -10 \\ 2 & -5 \end{pmatrix}$, для которой $\alpha = -1$, $\beta = 2$, построим матрицу $G = F + \mathbb{I} = \begin{pmatrix} 4 & -10 \\ 2 & -4 \end{pmatrix}$, имеющую след, равный нулю и удовлетворяющую уравнению $G^2 = -4\mathbb{I}$. Тогда

$$\exp(tG) = \begin{pmatrix} \cos 2t & 0 \\ 0 & \cos 2t \end{pmatrix} + \frac{1}{2}\sin 2t\begin{pmatrix} 4 & -10 \\ 2 & -4 \end{pmatrix},$$

$$\exp(tG) = \begin{pmatrix} \cos 2t + 2\sin 2t & -5\sin 2t \\ \sin 2t & \cos 2t - 2\sin 2t \end{pmatrix}.$$

Умножив последнее равенство на $\exp(\alpha t\mathbb{I}) = \begin{pmatrix} e^{-t} & 0 \\ 0 & e^{-t} \end{pmatrix}$, получим ранее вычисленную $\exp(tF)$, но затратив гораздо меньше усилий.

3.4. Дифференциальные уравнения и фазовые портреты

Мы уже показали, что дифференциальное уравнение $\dot{\mathbf{v}}(t) = A\mathbf{v}$ с начальным условием $\mathbf{v}(0) = \mathbf{v}_0$ имеет единственное решение

$$\mathbf{v}(t) = \exp(tA)\mathbf{v}_0.$$

Это решение $\mathbf{v}(t)$ определяет функцию от *времени* в двумерном векторном пространстве. Поскольку $\exp(tA)$ определена как для положительных t, так и для отрицательных, область определения векторной функции $\mathbf{v}(t)$ — вся вещественная ось: $-\infty < t < \infty$. Изображая на графике положение конца вектора $\mathbf{v}$ для всех значений t, мы получаем *интегральную кривую* дифференциального уравнения (рис. 3.1). Эта кривая изображает траекторию частицы, движущейся в плоскости, вектор $\dot{\mathbf{v}}(t) = A\mathbf{v}(t)$ касается траектории и изображает скорость частицы в момент времени t. Через каждую точку плоскости проходит единственное решение уравнения, а преобразование $\exp(tA)$ задает движение по этой кривой в зависимости от времени t.

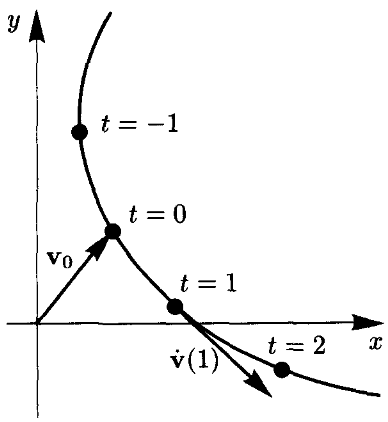


Рис. 3.1

Нарисовав семейство интегральных кривых, мы можем создать *фазовый портрет*, отражающий важные свойства решений дифференциального уравнения. Хотя матрицы A, задающие дифференциальное уравнение $\dot{\mathbf{v}}(t) = A\mathbf{v}$, могут быть самыми разными, но они порождают ограниченное число различных *типов* фазовых портретов. Скажем конкретнее: если матрицы A и F — сопряженные, т.е. $A = BFB^{-1}$, то интегральные кривые уравнения $\dot{\mathbf{v}}(t) = A\mathbf{v}$ получаются из решений уравнения $\dot{\mathbf{w}}(t) = F\mathbf{w}$ линейным преобразованием $\mathbf{v}(t) = B\mathbf{w}(t)$. Доказательство очень простое: матрица B не зависит от времени, поэтому $\dot{\mathbf{v}}(t) = B\dot{\mathbf{w}}(t)$. Следовательно, если $\dot{\mathbf{w}}(t) = F\mathbf{w}$, то

$$\dot{\mathbf{v}} = B\dot{\mathbf{w}} = BF\mathbf{w} = BFB^{-1}\mathbf{v} = A\mathbf{v}.$$

Таким образом, если A и F — сопряженные матрицы, то фазовые портреты для уравнения $\dot{\mathbf{v}}(t) = A\mathbf{v}$, по существу, совпадают с фазовыми портретами $\dot{\mathbf{w}}(t) = F\mathbf{w}$. Поэтому можно (с точностью до линейного преобразования) определить фазовые портреты, рассматривая различные возможные собственные значения матрицы A.

Сначала отметим, что если $\mathbf{v}_0$ — собственный вектор A, соответствующий собственному значению λ, то

$$\begin{aligned}\mathbf{v}(t) = \exp(tA)\mathbf{v}_0 &= \left[1 + tA + \frac{1}{2!}(tA)^2 + \cdots\right]\mathbf{v}_0 \\ &= \left[1 + t\lambda + \frac{1}{2!}(t\lambda)^2 + \cdots\right]\mathbf{v}_0 = e^{\lambda t}\mathbf{v}_0.\end{aligned}$$

В этом случае интегральная кривая представляет собой прямую, проходящую через начало координат по направлению вектора $\mathbf{v}_0$. Если λ положительно, то с ростом t конец вектора $\mathbf{v}(t)$ экспоненциально удаляется от начала координат. Если же λ отрицательно, то частица, движущаяся по закону $\mathbf{v}(t)$, при $t \to +\infty$ экспоненциально приближается к началу координат. Наконец, когда $\lambda = 0$, $\mathbf{v}(t) = \mathbf{v}_0$ для всех t. Поэтому все точки на прямой, проходящей через начало координат и $\mathbf{v}_0$, остаются неподвижными.

Теперь мы можем перечислить все возможные случаи.

Случай 1. Матрица A пропорциональна единичной матрице.

1a. $A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, $\exp(tA) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Поэтому все точки неподвижны.

1b. $A = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$, $\exp(tA) = \begin{pmatrix} e^{\mu} & 0 \\ 0 & e^{\mu} \end{pmatrix}$. Каждый вектор является собственным, поэтому все интегральные кривые — это прямые линии, проходящие через начало координат. Если $\lambda > 0$, то при $t \to +\infty$ каждая точка плоскости экспоненциально удаляется от начала координат (рис. 3.2(a)); если $\lambda < 0$, то точки движутся в обратном направлении, к началу координат (рис. 3.2(b)).

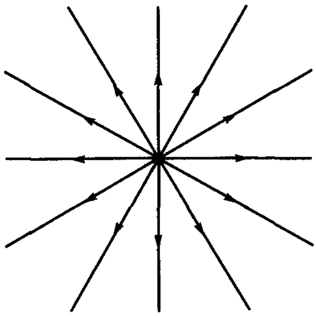


Рис. 3.2(a). $\lambda > 0$

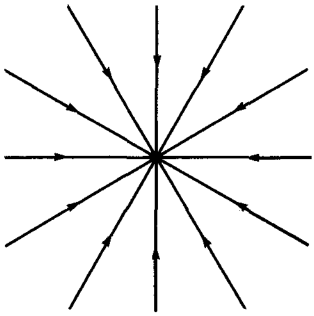


Рис. 3.2(b). $\lambda < 0$

Случай 2. Матрица A имеет разные вещественные собственные значения λ_1 и λ_2. Тогда

$$A = B \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} B^{-1}.$$

2a. λ_1 и λ_2 положительны и $\lambda_1 > \lambda_2$. В частном случае, если $B = \mathbb{I}$, интегральные кривые изображены на рис. 3.3(a). Оси x и y являются интегральными кривыми. Далее, поскольку решения задаются формулами

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e^{\lambda_1 t} x_0 \\ e^{\lambda_2 t} y_0 \end{pmatrix}$$

и $\lambda_1 - \lambda_2 > 0$, отношение $x/y = e^{(\lambda_1 - \lambda_2)t}$ растет при $t \to +\infty$, а при $t \to -\infty$ кривые входят в начало координат, касаясь оси y.

В случае произвольной матрицы B пусть $\mathbf{v}_1$ и $\mathbf{v}_2$ — собственные векторы матрицы A. Прямые, идущие вдоль векторов $\mathbf{v}_1$ и $\mathbf{v}_2$, являются интегральными кривыми. Поскольку $\lambda_1 > \lambda_2$, другие интегральные кривые становятся почти параллельными $\mathbf{v}_1$

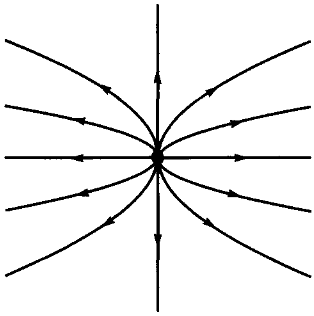


Рис. 3.3(a). $A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$

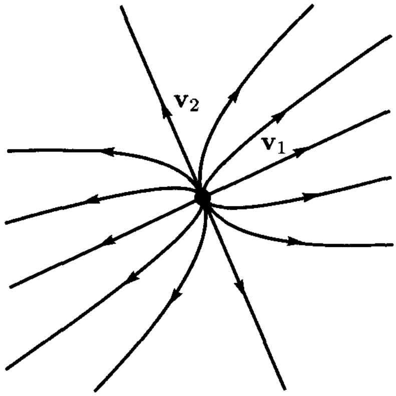


Рис. 3.3(b). $A = B\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} B^{-1}$

при $t \to +\infty$ и касаются $\mathbf{v}_2$ при $t \to -\infty$, как это показано на рис. 3.3(b).

2b. λ_1 и λ_2 отрицательны и $|\lambda_1| < |\lambda_2|$. Этот случай похож на случай 2a, но все стрелки направлены в противоположную сторону. При $t \to +\infty$ все интегральные кривые движутся к началу координат.

2c. λ_1 положительно, а λ_2 отрицательно. В частном случае, когда $B = \mathbb{I}$, оси x и y опять являются интегральными кривыми. С ростом t значение $|x|$ растет, а значение $|y|$ убывает.

Другие интегральные кривые стремятся к оси x при $t \to +\infty$ и к оси y при $t \to -\infty$, как это показано на рис. 3.4(a).

В случае произвольной матрицы B пусть вектор $\mathbf{v}_1$ соответствует собственному значению $\lambda_1 > 0$, а вектор $\mathbf{v}_2$ — собственному значению $\lambda_2 < 0$. Точки, расположенные на прямой, проходящей через $\mathbf{v}_1$, удаляются от начала координат, а точки на прямой, порожденной вектором $\mathbf{v}_2$, движутся к началу координат. Кривые решений стремятся к прямой, проходящей через $\mathbf{v}_1$, при $t \to \infty$, и к прямой, проходящей через $\mathbf{v}_2$, при $t \to -\infty$, как это изображено на рис. 3.4(b).

2d. λ_1 положительно, λ_2 равно нулю. В частном случае, когда $B = \mathbb{I}$, координата y на каждом решении не зависит от времени. Точки оси y неподвижны. Интегральными кривыми являются

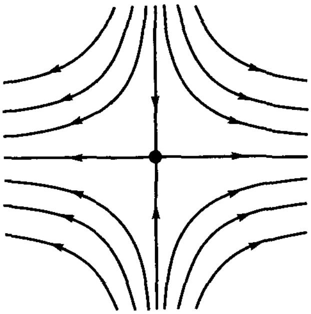


Рис. 3.4(a). $A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$, $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e^{\lambda_1 t} x_0 \\ e^{\lambda_2 t} y_0 \end{pmatrix}$

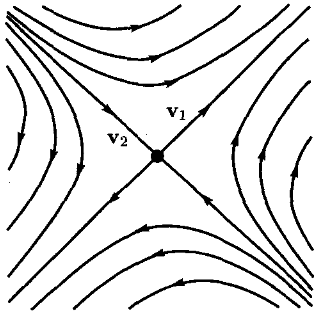


Рис. 3.4(b). $A = B \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} B^{-1}$

прямые, параллельные оси x. При $t \to +\infty$ точки удаляются от оси y, а при $t \to -\infty$ приближаются к ней (рис. 3.5(a)).

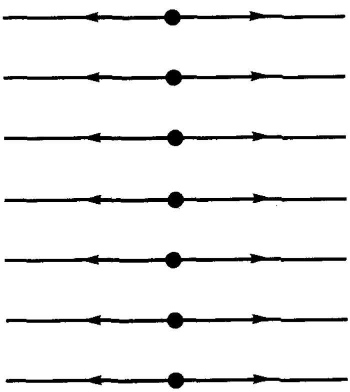


Рис. 3.5(a). $A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e^{\lambda_1 t} x_0 \\ y_0 \end{pmatrix}$

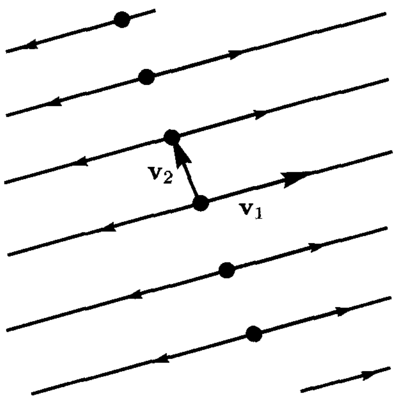


Рис. 3.5(b). $A = B \begin{pmatrix} \lambda_1 & 0 \\ 0 & 0 \end{pmatrix} B^{-1}$

В случае произвольной матрицы B, пусть вектор $\mathbf{v}_1$ соответствует собственному значению $\lambda_1 > 0$, а $\mathbf{v}_2$ — собственному значению $\lambda_2 = 0$. Прямые, проходящие через начало координат вдоль вектора $\mathbf{v}_2$, состоят из неподвижных точек. Интегральными кривыми являются прямые, параллельные вектору $\mathbf{v}_1$. Точки удаляются от прямой, проходящей через $\mathbf{v}_2$, при $t \to \infty$ и приближаются к ней при $t \to -\infty$ (рис 3.5(b)).

2e. λ_1 отрицательно, λ_2 равно нулю. Все стрелки имеют направление, противоположное предыдущему случаю. При $t \to +\infty$ вся плоскость проектируется на прямую, проходящую через вектор $\mathbf{v}_2$.

Случай 3. Собственные значения матрицы A равны, но $A \neq \lambda\mathbb{I}$. При этом $A = B\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}B^{-1}$, $\exp(tA) = Be^{\lambda_1}\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}B^{-1}$. В этом случае преобразование A имеет только один собственный вектор $\mathbf{v}_1$.

3a. $\lambda > 0$. Если $B = \mathbb{I}$, то собственный вектор $\mathbf{v}_1$ лежит на оси x. Ось x является интегральной кривой. На любом решении при $t \to +\infty$ $|x|$ становится много больше $|y|$; соответственно интегральные кривые, задаваемые формулой

$$\begin{pmatrix} x \\ y \end{pmatrix} = e^{\lambda t}\begin{pmatrix} x_0 + ty_0 \\ y_0 \end{pmatrix},$$

становятся «почти параллельными» оси x (рис. 3.6(a)). При $t \to -\infty$ знак x противоположен знаку y. Все кривые пересекают ось y при $t = -x_0/y_0$.

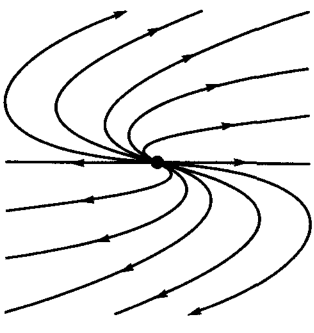


Рис. 3.6(a). $A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$

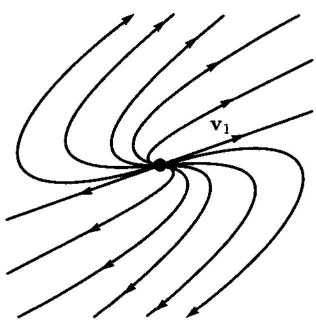


Рис. 3.6(b). $A = B\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}B^{-1}$

В случае произвольной матрицы B интегральной кривой является прямая, проходящая через начало координат вдоль вектора $\mathbf{v}_1$. При $t \to \infty$ все другие кривые становятся «параллельными» этой прямой, причем пересекаются они в начале координат.

3b. $\lambda < 0$. Стрелки меняют направление на противоположное по сравнению с предыдущим случаем.

3c. $\lambda = 0$. В частном случае, когда $B = \mathbb{I}$, ось x состоит из неподвижных точек. Интегральными кривыми являются прямые, параллельные этой оси. Для произвольного значения t функция $\exp(tA)$ является преобразованием сдвига (см. рис. 3.7(a)).

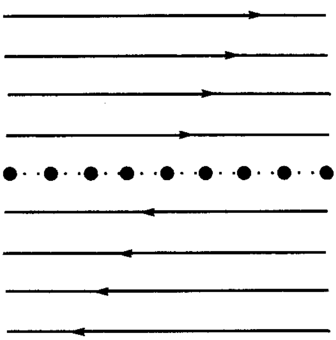


Рис. 3.7(a). $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\exp(tA) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$

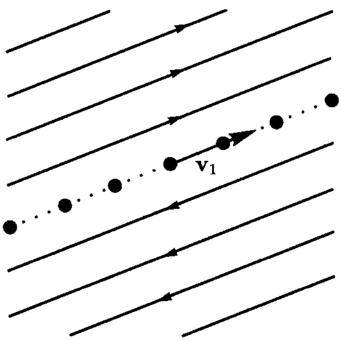


Рис. 3.7(b). $A = B \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} B^{-1}$

В случае произвольной матрицы B прямая, проходящая через начало координат вдоль вектора $\mathbf{v}_1$, остается неподвижной. Все точки плоскости сдвигаются параллельно этой прямой (рис. 3.7(b)).

Случай 4. Матрица A имеет комплексно-сопряженные собственные значения $\alpha \pm \mathrm{i}\beta$. В этом случае собственные векторы отсутствуют.

4a. $\alpha = 0$. В частном случае, когда $B = \mathbb{I}$, интегральными кривыми будут окружности с центром в начале координат. Одно из решений — неподвижное начало координат (рис. 3.8(a)). Преобразование $\exp(tA)$ является вращением:

$$\exp(tA) = \begin{pmatrix} \cos\beta t & -\sin\beta t \\ \sin\beta t & \cos\beta t \end{pmatrix}.$$

Решения — периодические с периодом $2\pi/\beta$.

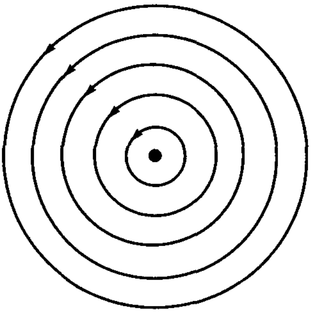


Рис. 3.8(a). $A = \begin{pmatrix} 0 & -\beta \\ \beta & 0 \end{pmatrix}$, $\exp(tA) = \begin{pmatrix} \cos\beta t & -\sin\beta t \\ \sin\beta t & \cos\beta t \end{pmatrix}$

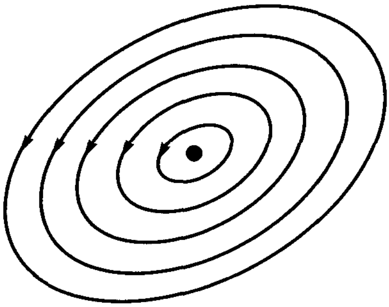


Рис. 3.8(b). $A = B\begin{pmatrix} 0 & -\beta \\ \beta & 0 \end{pmatrix}B^{-1}$

В общем случае интегральные кривые являются эллипсами. Решения — периодические с периодом $2\pi/\beta$ (рис. 3.8(b)).

4b. $\alpha > 0$. Интегральные кривые — это спирали, закрученные против часовой стрелки (рис. 3.9(a), (b)). При $t \to +\infty$ точки удаляются от начала координат, а при $t \to -\infty$ приближаются к нему.

4c. $\alpha < 0$. То же самое, что и в предыдущем случае, но спирали закручиваются к центру при $t \to +\infty$.

3.5. Применения дифференциальных уравнений

На рис. 3.10 изображена физическая система, поведение которой описывается дифференциальным уравнением только что рассмотренного типа.

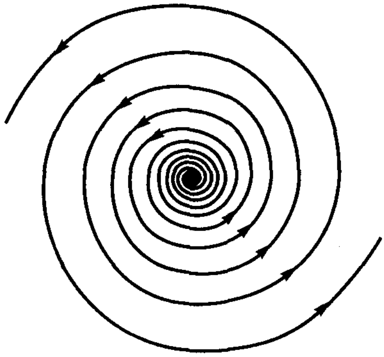


Рис. 3.9(a). $A = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$, $\exp(tA) = e^{\alpha t} \begin{pmatrix} \cos\beta t & -\sin\beta t \\ \sin\beta t & \cos\beta t \end{pmatrix}$

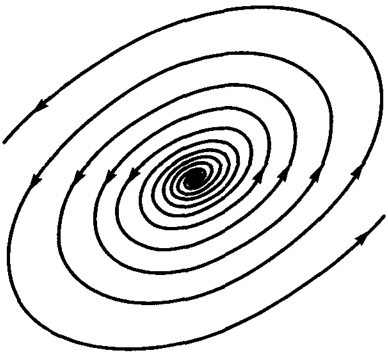


Рис. 3.9(b). $A = B \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix} B^{-1}$

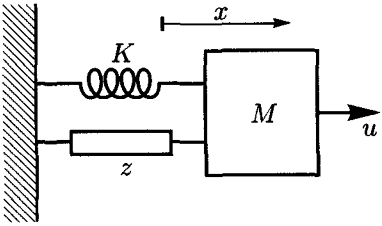


Рис. 3.10

Масса M движется под действием пружины, создающей силу $-kx$, пропорциональную смещению x; линейный буфер создает силу $-zu$, пропорциональную скорости u, с которой движется эта масса. Второй закон Ньютона говорит, что $M\dot{u} = -kx - zu$, где скорость определена как $\dot{x} = u$. Состояние системы полностью описывается вектором $\mathbf{w} = \begin{pmatrix} x \\ u \end{pmatrix}$, компонентами которого являются положение и скорость массы M. Поскольку

$$\begin{aligned} \dot{x} &= u, \\ \dot{u} &= -\frac{k}{M}x - \frac{z}{M}u, \end{aligned}$$

можно написать дифференциальное уравнение

$$\dot{\mathbf{w}} = A\mathbf{w},$$

где

$$A = \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & -\Gamma \end{pmatrix}, \quad \omega_0^2 = \frac{k}{M}, \quad \Gamma = \frac{z}{M}.$$

Движение системы определяется собственными значениями матрицы A. Характеристическое уравнение матрицы имеет вид

$$\lambda^2 + \Gamma\lambda + \omega_0^2 = 0,$$

решая которое, получаем следующие собственные значения:

$$\lambda = \frac{-\Gamma \pm \sqrt{\Gamma^2 - 4\omega_0^2}}{2}.$$

Таким образом, существуют четыре возможности.

1. Если $\Gamma = 0$ (отсутствует трение), то $\lambda = \pm i\omega_0$. Такая система движется, *не затухая*, совершая колебания около положения равновесия (*осциллятор без затухания*). Фазовый портрет соответствует случаю 4а из предыдущего параграфа. Кривыми решения являются эллипсы в плоскости xu (рис. 3.11(а)). Уравнение для каждого из эллипсов имеет вид $\frac{1}{2}kx^2 + \frac{1}{2}Mu^2 = \mathrm{const}$, что означает сохранение энергии. Период движения равен $2\pi/\omega_0$.

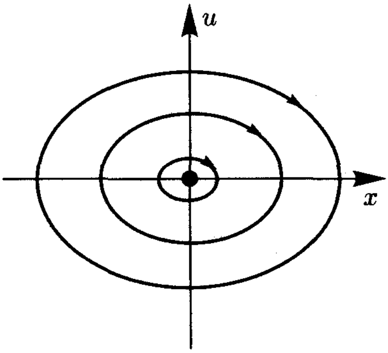


Рис. 3.11(a)

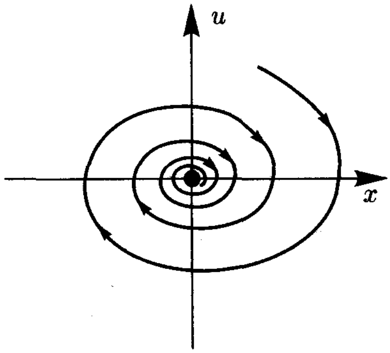


Рис. 3.11(b)

2. Если $\Gamma > 0$ и при этом $\Gamma^2 - 4\omega_0^2 < 0$, то собственные значения равны

$$\lambda = -\frac{\Gamma}{2} \pm \mathrm{i}\sqrt{\left(\omega_0^2 - \frac{1}{4}\Gamma^2\right)}.$$

Такая система называется *осциллятором с затухающими колебаниями*. Фазовый портрет соответствует случаю 4c. Интегральными кривыми являются спирали, закручивающиеся к началу координат (рис. 3.11(b)).

3. Если $\Gamma^2 = 4\omega_0^2$, то характеристическое уравнение имеет два одинаковых корня $\lambda = -\Gamma/2$. Такую систему называют *осциллятором с критическим затуханием*, его фазовый портрет соответствует случаю 3b (рис. 3.11(c)).

Запишем явно матрицу

$$A - \lambda\mathbb{I} = \begin{pmatrix} \Gamma/2 & 1 \\ -\Gamma^2/4 & -\Gamma/2 \end{pmatrix}.$$

Отсюда видно, что $\mathbf{w}_1 = \begin{pmatrix} 1 \\ -\Gamma/2 \end{pmatrix}$ — один из собственных векторов матрицы A. Это значит, что если

$$\begin{pmatrix} x_0 \\ u_0 \end{pmatrix} = \begin{pmatrix} x_0 \\ -(\Gamma/2)x_0 \end{pmatrix},$$

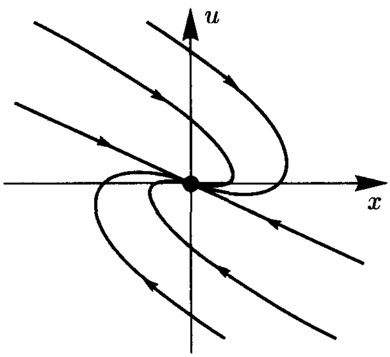


Рис. 3.11(c)

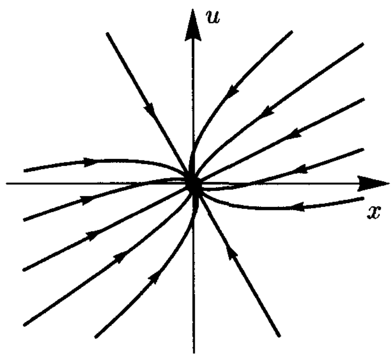


Рис. 3.11(d)

то в процессе движения отношение координат x и u остается постоянным.

4. Если $\Gamma^2 > 4\omega_0^2$, то характеристическое уравнение имеет отрицательные вещественные корни

$$\lambda = -\Gamma/2 \pm \sqrt{\Gamma^2/4 - \omega_0^2}.$$

Такое поведение системы называется *апериодическим затуханием*; его фазовый портрет соответствует случаю 2b (рис. 3.11(d)).

Чтобы лучше понять эти четыре типа движения, давайте для каждого из этих случаев найдем и нарисуем зависимость координаты x от времени t.

Случай 1. Осциллятор без затухания. Мы имеем:

$$A = \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & 0 \end{pmatrix} = \begin{pmatrix} \omega_0^{-1/2} & 0 \\ 0 & \omega_0^{1/2} \end{pmatrix} \begin{pmatrix} 0 & \omega_0 \\ -\omega_0 & 0 \end{pmatrix} \begin{pmatrix} \omega_0^{1/2} & 0 \\ 0 & \omega_0^{-1/2} \end{pmatrix},$$

следовательно,

$$\exp\begin{pmatrix} 0 & 1 \\ -\omega_0^2 & 0 \end{pmatrix} = \begin{pmatrix} \omega_0^{-1/2} & 0 \\ 0 & \omega_0^{1/2} \end{pmatrix} \begin{pmatrix} \cos\omega_0 t & \sin\omega_0 t \\ -\sin\omega_0 t & \cos\omega_0 t \end{pmatrix} \begin{pmatrix} \omega_0^{1/2} & 0 \\ 0 & \omega_0^{-1/2} \end{pmatrix}.$$

Отсюда зависимость x от t имеет вид

$$x = m\cos\omega_0 t + n\sin\omega_0 t,$$

где m и n определяются из начальных условий. Введем параметр ρ, удовлетворяющий равенству

$$m^2 + n^2 = \rho^2$$

и определим ϕ из соотношений

$$m = \rho\sin\phi, \quad n = \rho\cos\phi.$$

Тогда можно написать

$$x(t) = \rho\cos(\omega_0 t + \phi).$$

Такая кривая представляет собой синусоиду с частотой ω_0. Фаза ϕ и амплитуда ρ зависят от начальных условий (рис. 3.12).

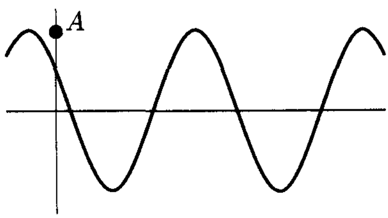


Рис. 3.12

Случай 2. Осциллятор с затухающими колебаниями. Нам известно, что в этом случае

$$\exp(tA) = B\begin{pmatrix} e^{-\Gamma t}\cos\omega_1 t & -e^{-\Gamma t}\sin\omega_1 t \\ e^{-\Gamma t}\sin\omega_1 t & e^{-\Gamma t}\cos\omega_1 t \end{pmatrix} B^{-1},$$

где

$$\omega_1 = \frac{1}{2}\sqrt{4\omega_0^2 - \Gamma^2}.$$

Таким образом, зависимость x от t имеет вид

$$x(t) = \rho e^{-\Gamma t}\cos(\omega_1 t + \phi),$$

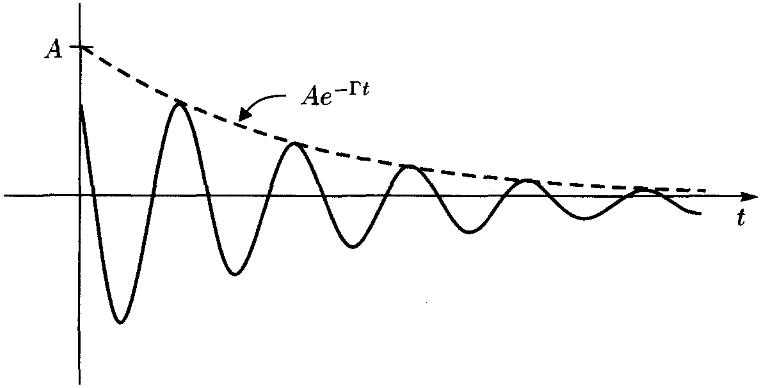


Рис. 3.13

где ρ и ϕ зависят от начальных условий. Графиком этой зависимости является *экспоненциально затухающая синусоида* (рис. 3.13).

Случаи 3 и 4. Здесь у графика $x(t)$ не существует синусоидальной составляющей в траектории, координата x экспоненциально затухает со временем t и может пересекать линию $x = 0$ не больше одного раза (рис. 3.14 и 3.15).

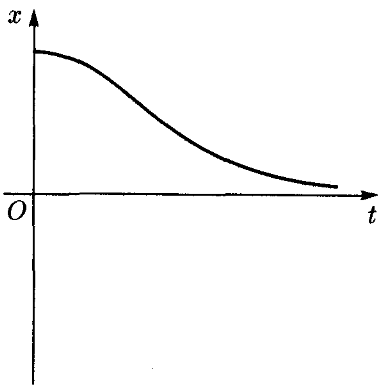


Рис. 3.14

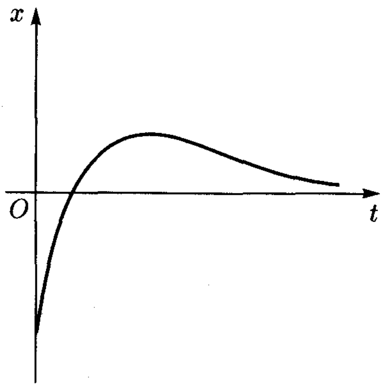


Рис. 3.15

Вынужденные колебания. Очень часто дифференциальные уравнения, возникающие при решении физических задач, являются *неоднородными* уравнениями типа

$$\dot{\mathbf{v}} - A\mathbf{v} = \mathbf{b}(t),$$

где $\mathbf{b}(t)$ не равняется тождественно нулю. Чтобы понять, как появляется член $\mathbf{b}(t)$, рассмотрим управляемый осциллятор: на массу M помимо пружины и буфера действует внешняя вынуждающая сила $F(t)$ (рис. 3.16).

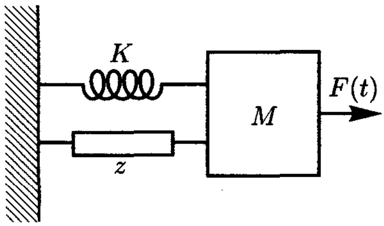


Рис. 3.16

Тогда уравнения $\dot{x} = u$ и $m\dot{u} = -kx - Zu + F(t)$ можно записать в форме

$$\begin{pmatrix} \dot{x} \\ \dot{u} \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ -k/M & -Z/M \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} = \begin{pmatrix} 0 \\ F(t)/M \end{pmatrix}.$$

Чтобы получить решение неоднородного уравнения $\dot{\mathbf{v}} - A\mathbf{v} = \mathbf{b}(t)$, обобщим метод, называемый *вариацией постоянных.* Этот метод хорошо известен для решения линейных дифференциальных уравнений от одной переменной. В случае одной неизвестной функции для решения уравнения

$$\dot{x} - kx = b(t)$$

пользуются тем, что функция e^{kt} является решением однородного уравнения $\dot{x} - kx = 0$. Тогда решение неоднородного уравнения можно искать в виде $x(t) = e^{kt}u(t)$. Подставляя это выражение в неоднородное уравнение $\dot{x} - kx = b(t)$, получаем

$$ke^{kt}u(t) + e^{kt}\dot{u}(t) - ke^{kt}u(t) = b(t),$$

откуда

$$\dot{u}(t) = e^{-kt}b(t).$$

Интегрируя это равенство по времени, получаем в качестве одного из решений

$$u(t) = \int\limits_0^t e^{-ks} b(s)\, ds, \quad x(t) = e^{kt} u(t) = e^{kt} \int\limits_0^t e^{-ks} b(s)\, ds$$

или

$$x(t) = \int\limits_0^t e^{k(t-s)} b(s)\, ds.$$

Это решение получено для начального условия $x(0) = 0$. Чтобы получить решение для произвольного начального условия x_0, мы просто добавляем соответствующее решение однородного уравнения $e^{kt}x_0$. Тогда решение уравнения $\dot{x} - kx = b(t)$ с начальным условием $x(0) = x_0$ можно записать в виде

$$x(t) = \int\limits_0^t e^{k(t-s)} b(s)\, ds + e^{kt} x_0.$$

Этот метод можно применить и для решения векторного уравнения $\dot{\mathbf{v}} - A\mathbf{v} = \mathbf{b}(t)$. Действительно, функция $\exp(tA)\mathbf{v}_0$ является решением однородного уравнения $\dot{\mathbf{v}} - A\mathbf{v} = \mathbf{0}$. Мы можем заменить постоянный вектор $\mathbf{v}_0$ на функцию $\mathbf{w}(t)$ и искать решение неоднородного уравнения в виде $\mathbf{v}(t) = \exp(tA)\mathbf{w}(t)$. Подстановка этой функции дает

$$A \exp(tA)\mathbf{w}(t) + \exp(tA)\dot{\mathbf{w}}(t) - A \exp(tA)\mathbf{w}(t) = \mathbf{b}(t).$$

Отсюда следует, что

$$\dot{\mathbf{w}}(t) = \exp(-tA)\mathbf{b}(t).$$

Интегрируя по времени, получаем, как одно из решений,

$$\mathbf{w}(t) = \int\limits_0^t \exp(-sA)\mathbf{b}(s)\, ds.$$

Итак,

$$\mathbf{v}(t) = \exp(tA)\mathbf{w}(t),$$

$$\mathbf{v}(t) = \int_0^t \exp[(t-s)A]\mathbf{b}(s)\,ds.$$

Очевидно, это решение удовлетворяет условию $\mathbf{v}(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Чтобы получить решение с произвольными начальными условиями, добавим общее решение однородного уравнения. Тогда получаем векторную функцию

$$\mathbf{v}(t) = \int_0^t \exp[(t-s)A]\mathbf{b}(s)\,ds + \exp(tA)\mathbf{v}_0,$$

которая удовлетворяет условию $\mathbf{v}(0) = \mathbf{v}_0$.

В учебниках по физике при обсуждении осциллятора с вынуждающей силой обычно рассматривают в качестве функции $\mathbf{b}(t)$ синусоиду с постоянной частотой ω. Например,

$$\mathbf{b}(t) = \begin{pmatrix} \sin\omega t + \cos\omega t \\ 3\cos\omega t \end{pmatrix}.$$

Тогда $\mathbf{b}(t)$ удовлетворяет уравнению $\ddot{\mathbf{b}} = -\omega^2\mathbf{b}$. В этом случае мы можем в общем решении, полученном выше, провести интегрирование по частям. Применим тот же трюк, который используют для вычисления неопределенного интеграла $\int e^{-ks}\sin s\,ds$: проинтегрируем два раза по частям и получим уравнение для неизвестного интеграла. Пусть

$$\mathbf{v} = \int_0^t \exp[(t-s)A]\mathbf{b}(s)\,ds.$$

Интегрируем один раз по частям и, предполагая, что A — несингулярная матрица, получаем

$$\mathbf{v} = \Big[-\exp[(t-s)A]A^{-1}\mathbf{b}(s)\Big]_0^t + \int_0^t \exp[(t-s)A]A^{-1}\dot{\mathbf{b}}(s)\,ds.$$

Еще раз интегрируем по частям:

$$\mathbf{v} = \left[-\exp[(t-s)A]A^{-1}\mathbf{b}(s)\right]_0^t - \left[\exp(t-s)A^{-2}\dot{\mathbf{b}}(s)\right]_0^t + \int_0^t \exp[(t-s)A]A^{-2}\ddot{\mathbf{b}}(s)\,ds.$$

Заменив $\ddot{\mathbf{b}}$ на $-\omega^2\mathbf{b}$, видим, что последний член равен $-A^{-2}\omega^2\mathbf{v}$. Итак,

$$\mathbf{v} + A^{-2}\omega^2\mathbf{v} = -\left[\exp[(t-s)A]\left[A^{-1}\mathbf{b}(s) + A^{-2}\dot{\mathbf{b}}(s)\right]\right]_0^t$$

или

$$(A^2+\omega^2 I)\mathbf{v} = -\left[\exp[(t-s)A]\left[\dot{\mathbf{b}}(s) + A\mathbf{b}(s)\right]\right]_0^t.$$

Кроме случая, когда A имеет собственные значения $\pm i\omega$ (тогда матрица $(A^2+\omega^2\mathbb{I})$ сингулярна), обе части последнего равенства умножаем на $(A^2+\omega^2\mathbb{I})^{-1}$ и получаем решение уравнения в явном виде

$$\mathbf{v} = -(A^2+\omega^2\mathbb{I})^{-1}\left[\dot{\mathbf{b}}(t) + A\mathbf{b}(t) - \exp(tA)\left(\dot{\mathbf{b}}(0) + A\mathbf{b}(0)\right)\right].$$

Член, содержащий $\exp(tA)$, обеспечивает выполнение начального условия $\mathbf{v}(0) = 0$. Если $\Gamma > 0$, то умножение на $\exp(tA)$ приводит к тому, что любой вектор стремится к нулю при $t \to +\infty$. Таким образом, при больших значениях t этот член можно опустить и мы получаем *установившееся* решение:

$$\mathbf{v} = -(A^2+\omega^2\mathbb{I})^{-1}\left[\dot{\mathbf{b}}(t) + A\mathbf{b}(t)\right].$$

Чтобы проверить этот результат, заметим, что

$$\dot{\mathbf{v}} = -(A^2+\omega^2\mathbb{I})^{-1}\left[\ddot{\mathbf{b}}(t) + A\dot{\mathbf{b}}(t)\right]$$

и

$$A\mathbf{v} = -(A^2+\omega^2\mathbb{I})^{-1}\left[A\dot{\mathbf{b}}(t) + A^2\mathbf{b}(t)\right].$$

Тогда

$$\dot{\mathbf{v}} - A\mathbf{v} = -(A^2 + \omega^2\mathbb{I})^{-1}\left[\omega^2\mathbf{b}(t) + A^2\mathbf{b}(t)\right] = \mathbf{b}(t).$$

Эта проверка показывает, что результат правильный, даже если матрица A сингулярна!

Например, мы хотим найти решение для

$$\begin{pmatrix}\dot{x}\\ \dot{y}\end{pmatrix} - \begin{pmatrix}-1 & -2\\ 2 & -1\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}\sin 3t\\ \cos 3t\end{pmatrix}.$$

Здесь

$$A = \begin{pmatrix}-1 & -2\\ 2 & -1\end{pmatrix}, \qquad \mathbf{b}(t) = \begin{pmatrix}\sin 3t\\ \cos 3t\end{pmatrix}$$

и $\ddot{\mathbf{b}}(t) = -3^2\mathbf{b}(t)$, т. е. $\omega = 3$. Тогда

$$A^2 + \omega^2\mathbb{I} = \begin{pmatrix}-1 & -2\\ 2 & -1\end{pmatrix}\begin{pmatrix}-1 & -2\\ 2 & -1\end{pmatrix} + \begin{pmatrix}9 & 0\\ 0 & 9\end{pmatrix} = \begin{pmatrix}6 & 4\\ -4 & 6\end{pmatrix},$$

$$(A^2 + \omega^2\mathbb{I})^{-1} = \frac{1}{52}\begin{pmatrix}6 & -4\\ 4 & 6\end{pmatrix} = \frac{1}{26}\begin{pmatrix}3 & -2\\ 2 & 3\end{pmatrix}$$

и

$$\begin{aligned}\mathbf{v} &= -\frac{1}{26}\begin{pmatrix}3 & -2\\ 2 & 3\end{pmatrix}\left[\begin{pmatrix}3\cos 3t\\ -3\sin 3t\end{pmatrix} + \begin{pmatrix}-\sin 3t - 2\cos 3t\\ 2\sin 3t - \cos 3t\end{pmatrix}\right]\\ &= -\frac{1}{26}\begin{pmatrix}3 & -2\\ 2 & 3\end{pmatrix}\begin{pmatrix}\cos 3t - \sin 3t\\ -\cos 3t - \sin 3t\end{pmatrix} = -\frac{1}{26}\begin{pmatrix}\sin 3t - 5\cos 3t\\ 5\sin 3t + \cos 3t\end{pmatrix}.\end{aligned}$$

Это есть установившееся решение исходного дифференциального уравнения. Заметим, что компоненты вектора $\mathbf{v}(t)$ — опять синусоидальные функции от t с той же частотой, что и вынуждающий член. Однако, амплитуды компонент $\mathbf{v}$ и соответствующие фазы (положение гребней и впадин волны) изменились.

Рассмотрим установившееся решение для физической системы, описанной в начале параграфа, с синусоидальной вынуждающей силой. Пусть

$$A = \begin{pmatrix}0 & 1\\ -\omega_0^2 & -\Gamma\end{pmatrix}, \qquad \mathbf{b}(t) = \begin{pmatrix}0\\ \sin\omega t\end{pmatrix}.$$

Тогда

$$A^2 = \begin{pmatrix} -\omega_0^2 & -\Gamma \\ \Gamma\omega_0^2 & \Gamma^2 - \omega_0^2 \end{pmatrix}.$$

Далее,

$$A^2 + \omega^2\mathbb{I} = \begin{pmatrix} \omega^2 - \omega_0^2 & -\Gamma \\ \Gamma\omega_0^2 & \Gamma^2 + \omega^2 - \omega_0^2 \end{pmatrix}$$

и

$$\mathrm{Det}(A^2+\omega^2\mathbb{I}) = (\omega^2-\omega_0^2)^2+\Gamma^2(\omega^2-\omega_0^2)+\Gamma^2\omega_0^2 = (\omega^2-\omega_0^2)^2+\Gamma^2\omega^2.$$

Поэтому

$$(A^2+\omega^2\mathbb{I})^{-1} = \frac{1}{(\omega^2-\omega_0^2)^2+\Gamma^2\omega^2}\begin{pmatrix} \Gamma^2+\omega^2-\omega_0^2 & \Gamma \\ -\Gamma\omega_0^2 & \omega^2-\omega_0^2 \end{pmatrix},$$

и мы получаем формулу для установившегося поведения системы

$$-(A^2+\omega^2\mathbb{I})^{-1}(\dot{\mathbf{b}}(t)+A\mathbf{b}(t)) = -\frac{1}{(\omega^2-\omega_0^2)^2+\Gamma^2\omega^2}$$
$$\times\begin{pmatrix} \Gamma^2+\omega^2-\omega_0^2 & \Gamma \\ -\Gamma\omega_0^2 & \omega^2-\omega_0^2 \end{pmatrix}\left(\begin{pmatrix} 0 \\ \omega\cos\omega t \end{pmatrix} + \begin{pmatrix} \sin\omega t \\ -\Gamma\sin\omega t \end{pmatrix}\right).$$

Таким образом, закон движения определяется формулой

$$\begin{aligned} x(t) &= -\frac{1}{(\omega^2-\omega_0^2)^2+\Gamma^2\omega^2}\left[\omega\Gamma\cos\omega t + (\omega^2-\omega_0^2)\sin\omega t\right] \\ &= \rho\sin(\omega t+\phi), \end{aligned}$$

где

$$\rho = \frac{1}{\sqrt{(\omega^2-\omega_0^2)^2+\Gamma^2\omega^2}}, \qquad \phi = \arcsin(-\Gamma\omega/\rho) = \arccos(\omega_0^2-\omega^2)/\rho.$$

Заметим, что если Γ мало, то для значения ω, близкого к ω_0, в амплитуде появляется большой множитель. Это явление называется *резонансом* (рис. 3.17).

Обратите также внимание, как изменяется *фазовый сдвиг* ϕ от нуля при малых значениях ω^2 до $-\pi/2$ для $\omega^2 = \omega_0^2$. Для больших значений ω^2 фазовый сдвиг равен $\arccos(-1) = -\pi$ (рис. 3.18).

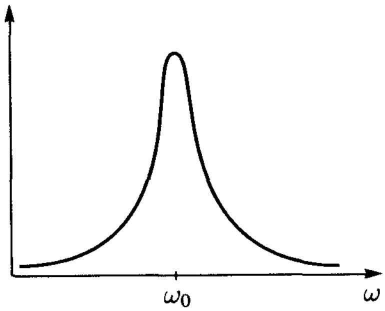


Рис. 3.17. Резонансная кривая

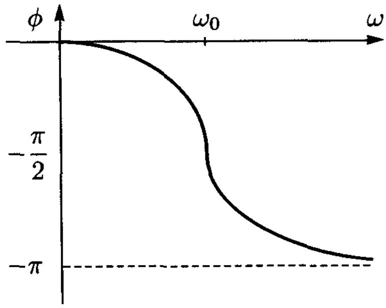


Рис. 3.18. Фазовый сдвиг

Резюме

А. Экспонента от матрицы

Изучив эту главу, вы должны научиться выражать экспоненциальную функцию $\exp tX$ в виде формального степенного ряда, уметь вычислить эту функцию в тех случаях, когда степени матрицы X имеют простую форму.

Вы должны уметь объяснить смысл производной $(d/dt)e^{tX}$ и показать, что она равна Xe^{tX}.

В. Линейные дифференциальные уравнения

Вы должны уметь показать, что все решения уравнения $\dot{\mathbf{v}} = A\mathbf{v}$ имеют вид $\mathbf{v} = e^{At}\mathbf{v}_0$.

Вы должны уметь вычислять e^{At} для любой матрицы A (2×2) и решать уравнение $\dot{\mathbf{v}} = A\mathbf{v}$ с заданными начальными условиями.

Вычислив собственные значения матрицы A (2×2), вы должны уметь рисовать фазовый портрет, представляющий интегральные кривые для уравнения $\dot{\mathbf{v}} = A\mathbf{v}$.

С. Неоднородные уравнения и гармонический осциллятор

Вы должны уметь преобразовывать дифференциальное уравнение второго порядка, описывающее осциллятор, к виду $\dot{\mathbf{v}} - A\mathbf{v} = \mathbf{b}$, где A — матрица 2×2 и $\mathbf{b}$ — вектор, зависящий от времени.

Вы должны уметь решать это уравнение и связывать решение с поведением осциллятора при наличии затухания и резонанса.

Задачи

3.1. (a) Напишите разложение в степенной ряд для $(1-X)^{-1}$ и для $(1-X)^{-2}$.

(b) Перемножьте два полученных ряда и сравните результат с разложением в степенной ряд $(1-X)^{-3}$.

3.2. (a) Пусть $F = \begin{pmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix}$. Докажите, что $F^2 = F/2$ и $F^n = F/2^{n-1}$. Используя этот результат, получите разложение в ряд матрицы $(\mathbb{I}-F)^{-1}$. Получите обратную матрицу прямым вычислением и сравните результаты.

(b) Вычислите матрицу, обратную $\begin{pmatrix} 3/2 & 1/2 \\ 1/2 & 3/2 \end{pmatrix}^{-1}$, записав ее в виде $(\mathbb{I}+P)^{-1}$, где P — проекция $\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$, и воспользовавшись разложением в ряд матрицы $(1+X)^{-1}$. Обратите внимание на то, что ряд расходится, несмотря на существование обратной матрицы.

3.3. (a) Для матрицы $N_{\pi/4} = \begin{pmatrix} -1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix}$ справедливо условие $N_{\pi/4}^2 = 0$. Используя это, вычислите $F(t) = \exp(tN_{\pi/4})$ и проверьте, что $F'(t) = N_{\pi/4}F(t)$.

(b) Матрица $P_{\pi/4} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ удовлетворяет условию $P_{\pi/4}^2 = P_{\pi/4}$. Вычислите $G(t) = \exp(tP_{\pi/4})$, воспользовавшись этим свойством, и проверьте, что $G'_{\pi/4} = P_{\pi/4}G(t)$.

3.4. Пусть матрица P удовлетворяет уравнению $P^2 = 3P$.

(a) Какие собственные значения имеет P? Приведите свои аргументы.

(b) Используя разложение экспоненты в степенной ряд, покажите, что $\exp(tP)$ может быть записана в виде

$$\exp(tP) = \mathbb{I} + g(t)P.$$

Найдите явное выражение для функции $g(t)$.

3.5. Допустим, что B есть матрица 2×2 с одинаковыми собственными значениями λ.

(a) Покажите, что матрица $N = B - \lambda\mathbb{I}$ является нильпотентом, т. е. $N^2 = 0$.

(b) Записав $B = N + \lambda\mathbb{I}$ и воспользовавшись разложением экспоненты в степенной ряд, покажите, что

$$\exp(tB) = (\mathbb{I} + tN)\exp(t\lambda\mathbb{I}).$$

3.6. Воспользовавшись результатом задачи 3.5, решите систему уравнений

$$\begin{aligned}\dot{x} &= x(t) - y(t),\\ \dot{y} &= x(t) + 3y(t)\end{aligned}$$

с произвольными начальными условиями $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$.

3.7. Вычислите $\exp(tA)$ для следующих матриц и проверьте, что $(d/dt)\exp(tA) = A\exp(tA)$:

(a) $A = \begin{pmatrix} -4 & 5 \\ -2 & 3 \end{pmatrix}$.

(b) $A = \begin{pmatrix} -1 & 9 \\ -1 & 5 \end{pmatrix}$. (Подсказка: $A = 2\mathbb{I} + N$, где N — нильпотентная матрица.)

(c) $A = \begin{pmatrix} 3 & -1 \\ 5 & -1 \end{pmatrix}$.

3.8. Пусть A — матрица 2×2, имеющая два разных вещественных собственных значения λ_1 и λ_2, соответствующие собственным векторам $\mathbf{v}_1$ и $\mathbf{v}_2$.

(a) Покажите, что матрица $P_1 = (A - \lambda_2\mathbb{I})/(\lambda_1 - \lambda_2)$ задает проекцию на прямую, задаваемую собственным вектором $\mathbf{v}_1$: $P_1^2 = P_1$, образ P_1 есть $\lambda\mathbf{v}_1$, а ядро P_1 есть $\lambda\mathbf{v}_2$.

(b) Покажите, что матрица $P_2 = (A - \lambda_1\mathbb{I})/(\lambda_2 - \lambda_1)$ задает проекцию на прямую, задаваемую $\mathbf{v}_2$. Покажите, что

$$P_1P_2 = P_2P_1 = 0, \quad P_1 + P_2 = \mathbb{I}, \quad \lambda_1P_1 + \lambda_2P_2 = A.$$

(c) Используя разложение экспоненты в степенной ряд, покажите, что

$$\exp(t\lambda_1P_1 + t\lambda_2P_2) = e^{\lambda_1 t}P_1 + e^{\lambda_2 t}P_2.$$

(d) Воспользовавшись этим результатом, решите систему уравнений

$$\begin{aligned}\dot{x}(t) &= -4x(t) + 5y(t),\\ \dot{y}(t) &= -2x(t) + 3y(t)\end{aligned}$$

с произвольным начальным условием $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$.

3.9. Пусть A — матрица 2×2, у которой след равен нулю и определитель равен 1.

(a) Напишите характеристическое уравнение матрицы A. Что можно сказать о матрице A^2 в этом случае?

(b) Используя разложение $\exp(tA)$ в степенной ряд, получите формулу
$$\exp(tA) = F(t)\mathbb{I} + G(t)A,$$
где матрицы $F(t)$ и $G(t)$ содержат тригонометрические функции от t.

(c) Кривой решения уравнения $\dot{\mathbf{v}} = A\mathbf{v}$ с начальным условием $\mathbf{v} = \mathbf{v}_0$ является эллипс, изображенный на рис. 3.19. Докажите, что все хорды, соединяющие $\exp(tA)\mathbf{v}_0$ с $\exp(-tA)\mathbf{v}_0$ параллельны $A\mathbf{v}_0$, и что средняя точка каждой хорды лежит на диаметре эллипса, на котором лежит вектор $\mathbf{v}_0$.

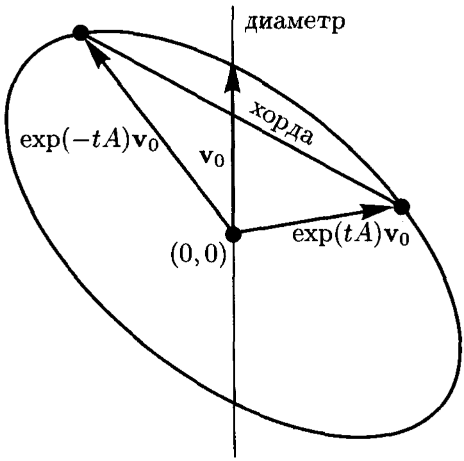


Рис. 3.19

3.10. Допустим, что матрица G имеет след, равный нулю и определитель, равный $-\beta^2$.

(a) Чему равняется G^2 (примените теорему Гамильтона–Кэли)?

(b) Используя разложение экспоненты в степенной ряд, покажите, что матрица $\exp G + \exp(-G)$ пропорциональна единичной матрице. Найдите такую функцию f, чтобы выполнялось равенство

$$\exp(G) + \exp(-G) = f(\beta)\mathbb{I}.$$

(c) Умножив это тождество на $\exp G$ и применив теорему Гамильтона–Кэли, покажите, что $\mathrm{Det}(\exp(G)) = 1$, и получите выражение для следа матрицы $\exp G$.

(d) Пусть $F = \lambda\mathbb{I} + G$. Используя результаты, полученные выше, покажите, что $\mathrm{Det}(\exp F) = \exp(\mathrm{Tr}\, F)$.

3.11. Для каждой системы дифференциальных уравнений, представленных ниже, найдите, какой из фазовых портретов, полученных в параграфе 3.4 (случаи 1(a)–4(c)), наилучшим образом описывает поведение общего решения. Решите эти системы для начального условия $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ при $t = 0$.

(a) $\dot{x} = -4y,$
$\dot{y} = x - 4y;$

(b) $\dot{x} = x - 2y,$
$\dot{y} = -2x + 4yx - 4y;$

(c) $\dot{x} = 4x - 5y,$
$\dot{y} = 4x - 4y;$

(d) $\dot{x} = 2x + y,$
$\dot{y} = -x + 4y;$

(e) $\dot{x} = x - 5y,$
$\dot{y} = 2x - 5y;$

(f) $\dot{x} = -2x + 4y,$
$\dot{y} = -x + 2y.$

3.12. Решите системы уравнений, приведенные ниже, с начальными условиями $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$. В каждом случае определите, какой фазовый портрет наилучшим образом представляет общее решение системы.

(a) $\dot{x} = 3y,$
$\dot{y} = x - 2y;$

(b) $\dot{x} = -x + y,$
$\dot{y} = -5x + 3y;$

(c) $\dot{x} = 3x + y,$
$\dot{y} = -x + y;$

(d) $\dot{x} = -5x + 4y,$
$\dot{y} = -8x + 7y;$

(e) $\dot{x} = -4x - 2y,$
$\dot{y} = 5x + 2y;$

(f) $\dot{x} = x + 2y,$
$\dot{y} = 2x - 4y.$

3.13. Обобщив на случай матрицы 3×3 то, что вам известно о вычислении и применении экспоненты, в показателе которой стоит матрица 2×2, решите систему дифференциальных уравнений

$$\begin{aligned} \dot{x} &= y, \\ \dot{y} &= z, \\ \dot{z} &= -6x - 11y - 6z \end{aligned}$$

с начальными условиями $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$ при $t = 0$. (Примечание: возникнет проблема вычисления обратной матрицы 3×3. Это можно сделать «в лоб», если рассмотреть задачу решения трех систем линейных уравнений.)

3.14. Обобщив известную технику, решите систему уравнений

$$\begin{aligned} \dot{x} &= x + y - z, \\ \dot{y} &= -x + 5y + z, \\ \dot{z} &= -2x + 2y + 4z \end{aligned}$$

с начальными условиями $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ при $t = 0$.

3.15. Введя величину $v = \dot{x}$, преобразуйте дифференциальное уравнение второго порядка

$$\ddot{x} + 4\dot{x} + 5x = 0$$

в систему уравнений первого порядка. Решите эту систему для произвольных начальных условий $\begin{pmatrix} x_0 \\ v_0 \end{pmatrix}$.

3.16. (a) Дифференциальное уравнение для гармонического осциллятора с критическим затуханием (единицы выбраны так, что $\omega^2 = 1$) имеет вид

$$\ddot{x} + 2\dot{x} + x = 0.$$

Решите это уравнение матричным методом, вводя $v = \dot{x}$ в качестве новой переменной. Запишите решение для начальных условий $\begin{pmatrix} x \\ v \end{pmatrix} = \begin{pmatrix} x_0 \\ 0 \end{pmatrix}$ и $\begin{pmatrix} x \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ v_0 \end{pmatrix}$, нарисуйте фазовые портреты этих и других решений. Покажите, что каждое из равенств $x = 0$ или $v = 0$ может выполняться не больше, чем в одной точке.

(b) Есть другой способ решения этого уравнения. Сначала ищутся решения уравнения $\ddot{x}+2\dot{x}+(1-\varepsilon^2)x = 0$ при $\varepsilon > 0$, что приводит к матрице с разными вещественными собственными значениями, а затем рассматривается предельный переход $\varepsilon \to 0$. (Физически это соответствует использованию слегка ослабленной пружины.) Выполните всю процедуру для начальных условий $\begin{pmatrix} x_0 \\ 0 \end{pmatrix}$ и $\begin{pmatrix} 0 \\ v_0 \end{pmatrix}$. Покажите, что произойдет с фазовыми портретами при $\varepsilon \to 0$.

(c) Еще один способ решения того же уравнения. Сначала решается уравнение $\ddot{x}+2\dot{x}+(1+\varepsilon^2)x = 0$, что приводит к матрице с комплексно-сопряженными собственными значениями, после этого $\varepsilon \to 0$. Проделайте вычисления для начальных условий $\begin{pmatrix} x_0 \\ 0 \end{pmatrix}$ и $\begin{pmatrix} 0 \\ v_0 \end{pmatrix}$, нарисуйте фазовые портреты решений и покажите, что происходит с фазовыми портретами при $\varepsilon \to 0$.

3.17. Рассмотрим функцию $\cos tx$.

(a) Разложите функцию в формальный степенной ряд и покажите, что

$$\frac{d^2}{dt^2}(\cos tx) = -x^2 \cos tx$$

и

$$\frac{d}{dt}(\cos tx) = 0 \quad \text{при} \quad t = 0.$$

(b) Пусть $\begin{pmatrix} \ddot{x} \\ \ddot{y} \end{pmatrix} = -B \begin{pmatrix} x \\ y \end{pmatrix}$, где матрица B имеет квадратный корень A (т. е. $A^2 = B$). Покажите, что $\cos(tA) \begin{pmatrix} x(0) \\ y(0) \end{pmatrix}$ есть решение дифференциального уравнения второго порядка

$$\frac{d^2}{dt^2}v(t) = A^2 v(t)$$

с начальными условиями $v(0) = v_0$ и $dv/dt(0) = 0$.[10]

(c) Пусть $B = \begin{pmatrix} 5/2 & -3/2 \\ -3/2 & 5/2 \end{pmatrix}$. Найдите такую матрицу A, которая имеет положительные собственные значения и удовлетворяет уравнению $A^2 = B$. (Указание: диагонализуйте матрицу B.)

(d) Для построенной таким образом матрицы A вычислите $\cos(tA)$. (Указание: Матрица A уже приведена к диагональному виду. Воспользуйтесь той же процедурой, что и для вычисления $\exp(tA)$.)

(e) Используйте полученные выше результаты для решения системы уравнений

$$\ddot{x} = -\frac{5}{2}x + \frac{3}{2}y,$$
$$\ddot{y} = \frac{3}{2}x - \frac{5}{2}y$$

с начальными условиями

$$\dot{x}(0) = \dot{y}(0) = 0, \quad x(0) = x_0, \quad y(0) = y_0.$$

3.18. Рассмотрим систему дифференциальных уравнений

$$\dot{x} = 4\beta x - y$$
$$\dot{y} = 9x + \beta,$$

где β — вещественный параметр.

(a) Решите систему с произвольными начальными условиями при $\beta = 0$.

(b) Найдите два критических значения параметра $\beta_1 < 0$ и $\beta_2 > 0$, при которых меняется характер решения. Обсудите случаи $\beta = \beta_1$ и $\beta = \beta_2$.

(c) Нарисуйте фазовые портреты, которые качественно описывают природу решений для $\beta < \beta_1$, $\beta_1 < \beta < \beta_2$ и $\beta > \beta_2$.

3.19. Пусть $A = \begin{pmatrix} -2 & 1 \\ 2 & -1 \end{pmatrix}$.

[10]Авторы имеют в виду, что матрица $\cos(tA)$ должна быть получена как сумма формального степенного ряда для $\cos tx$, где вместо числа x подставлена матрица A. Разумеется, необходимо убедиться в сходимости полученного ряда. (Можно доказать, что он сходится для всех t при любой матрице A.) См. также указания авторов к задаче 3.17 (d). — *Прим. ред.*

(a) Найдите такие матрицы D и B, чтобы выполнялось соотношение $A = BDB^{-1}$ (D — диагональна).

(b) Постройте решение дифференциального уравнения $\dot{\mathbf{v}} = A\mathbf{v}$ для произвольных начальных условий $\mathbf{v}_0 = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ при $t = 0$.

(c) Нарисуйте фазовый портрет решения уравнения $\dot{\mathbf{v}} = A\mathbf{v}$. Получите образ и ядро матрицы

$$F = \lim_{t\to\infty} \exp(At),$$

объясните их значения в связи с фазовыми портретами.

(d) Выбрав пробное решение в виде $\mathbf{v} = \exp(At)\mathbf{w}$, получите решение дифференциального уравнения $\dot{\mathbf{v}} - A\mathbf{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

3.20. (a) Введя $u = \dot{x}$ в качестве новой переменной, преобразуйте

$$\ddot{x} + 2\dot{x} - 3x = 3\sin 2t + 2\cos 2t$$

к уравнению вида

$$\begin{pmatrix} \dot{x} \\ \dot{u} \end{pmatrix} - A\begin{pmatrix} x \\ u \end{pmatrix} = \mathbf{b}(t).$$

(b) Решите это уравнение с начальными условиями $x(0) = 0$, $u(0) = 0$, воспользовавшись результатами, полученными в параграфе 3.4.

3.21. Когда к незатухающему осциллятору приложена вынуждающая сила, частота которой равна собственной частоте осциллятора, обычные методы решения не работают. В этом случае они не дают установившегося решения. Однако, мы получили формулу

$$\mathbf{v}(t) = \int_0^t \exp[(t-s)A]\mathbf{b}(s)ds,$$

которая прекрасно работает. С помощью этой формулы получите решение уравнения

$$\begin{pmatrix} \dot{x} \\ \dot{u} \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ -\omega^2 & 0 \end{pmatrix}\begin{pmatrix} x \\ u \end{pmatrix} = \begin{pmatrix} 0 \\ \sin\omega t \end{pmatrix}$$

для начальных условий $x(0) = 0$, $u(0) = 0$.

Глава 4

Скалярные произведения

Глава 4 посвящена изучению скалярных произведений и квадратичных форм. В ней много физических приложений, в том числе обсуждение нормальных мод и подробное изложение специальной теории относительности.

4.1. Скалярные произведения в евклидовом пространстве

Как вы помните, на аффинной плоскости мы располагаем очень ограниченным понятием длины: мы можем сравнивать длины отрезков на параллельных прямых, но для прямых, которые не параллельны, это сделать нельзя. Например, можно говорить, что на рис. 4.1 длина отрезка QR (или $Q'R'$) в два раза больше длины отрезка PQ, но мы не имеем права сравнивать длины отрезков PQ' и PQ.

Евклидова плоскость — это аффинная плоскость, где определена *функция расстояния*, которая каждой паре точек приписывает неотрицательное вещественное число $D(P,Q)$, называемое *расстоянием* между этими точками. Эта функция расстояния совместима с ограниченным понятием длины в аффинной геометрии, например, на рис. 4.1 $D(Q',R') = 2D(P,Q)$. Кроме того, она позволяет сравнивать длины непараллельных отрезков, таких как PQ и PQ'. На евклидовой плоскости $\mathbb{R}^2$ расстояние

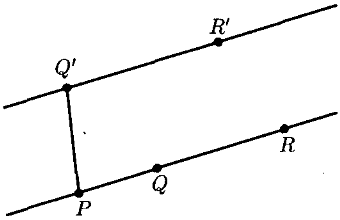


Рис. 4.1

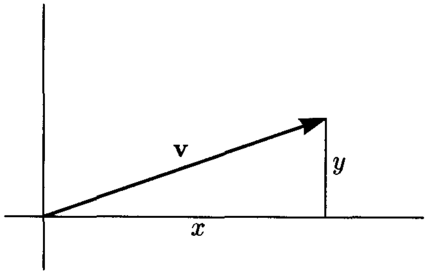


Рис. 4.2

между двумя точками определяется хорошо известной формулой

$$D(P,Q) = \sqrt{(x_Q - x_P)^2 + (y_Q - y_P)^2}.$$

Евклидовым преобразованием $f : \mathbb{R}^2 \to \mathbb{R}^2$ называется аффинное преобразование, сохраняющее функцию расстояния, т. е. удовлетворяющее условию $D(f(P), f(Q)) = D(P,Q)$.

Обратим внимание на векторное пространство смещений на евклидовой плоскости. Мы видим, что функция расстояния позволяет определить длину каждого вектора. Действительно, длина вектора — это просто расстояние от его «головы» до «хвоста». Длину вектора $\mathbf{v}$ будем обозначать $\|\mathbf{v}\|$. Очевидно, что если $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$, то $\|\mathbf{v}\| = \sqrt{x^2 + y^2}$ (рис. 4.2).

Вообще говоря, линейные преобразования векторного пространства $\mathbb{R}^2$ не сохраняют длины векторов. Линейные преобразования, сохраняющие длину, называются *ортогональными преобразованиями*. Это могут быть вращения вокруг начала координат или отражения относительно прямых, проходящих через начало координат.

Поскольку евклидово преобразование плоскости сохраняет длины, то оно каждый треугольник переводит в конгруэнтный треугольник (рис. 4.3) и, следовательно, сохраняет не только длины, но и *углы*. В частности, в евклидовой геометрии приобретает смысл понятие «перпендикулярности» (это понятие не определено в аффинной геометрии). Мы будем говорить, что два вектора $\mathbf{v}$ и $\mathbf{w}$ *перпендикулярны* или *ортогональны* друг другу,

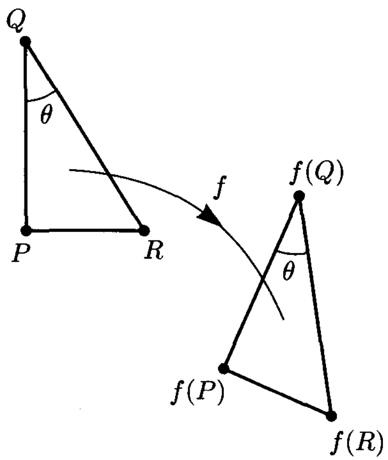


Рис. 4.3

если для треугольника, который они определяют, справедлива теорема Пифагора, т. е. $\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 = \|\mathbf{v} - \mathbf{w}\|^2$ (рис. 4.4).

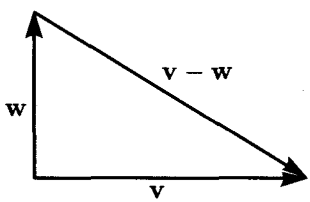


Рис. 4.4

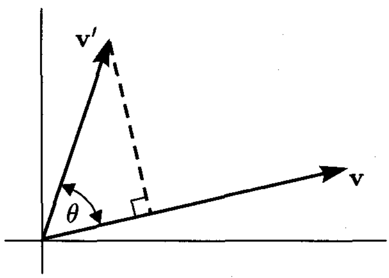


Рис. 4.5

Теперь, имея понятия длины и угла в евклидовом пространстве, мы можем определить *скалярное произведение* двух векторов. Пусть даны два вектора, $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$ и $\mathbf{v}' = \begin{pmatrix} x' \\ y' \end{pmatrix}$. Их скалярное произведение, обозначаемое $(\mathbf{v}, \mathbf{v}')$, определяется равенством

$$(\mathbf{v}, \mathbf{v}') = \|\mathbf{v}\| \cdot \|\mathbf{v}'\| \cos\theta,$$

где θ — угол между этими векторами.

Геометрически это можно истолковать так (рис. 4.5): мы берем проекцию вектора $\mathbf{v}'$ на прямую, на которой расположен вектор $\mathbf{v}$. Затем мы умножаем полученную проекцию на длину вектора $\mathbf{v}$. Берем это произведение со знаком плюс, если направление спроектированного вектора совпадает с направлением вектора $\mathbf{v}$, и со знаком минус, если проекция имеет противоположное направление. Поскольку скалярное произведение определено в терминах евклидовой геометрии плоскости, то любое линейное преобразование, сохраняющее длину, должно сохранять и скалярное произведение, т. е. линейное преобразование, сохраняющее длину векторов и угол между ними, сохраняет скалярное произведение этих векторов. Другими словами, если есть ортогональное преобразование M, то $(M\mathbf{v}, M\mathbf{v}') = (\mathbf{v}, \mathbf{v}')$ для любой пары векторов $\mathbf{v}'$ и $\mathbf{v}$. Ниже мы приведем алгебраическое доказательство этого утверждения. Обратно, поскольку $(\mathbf{v}, \mathbf{v}) = \|\mathbf{v}\|^2$, то любое преобразование M, сохраняющее скалярное произведение любых векторов $\mathbf{v}'$ и $\mathbf{v}$, является ортогональным.

Допустим, что мы зафиксировали вектор $\mathbf{v}$. Рассмотрим $(\mathbf{v}, \mathbf{v}')$ как функцию вектора $\mathbf{v}'$. Мы утверждаем, что $(\mathbf{v}, \mathbf{v}')$ является линейной функцией от $\mathbf{v}'$, т. е. $(\mathbf{v}, a\mathbf{v}' + b\mathbf{w}') = a(\mathbf{v}, \mathbf{v}') + b(\mathbf{v}, \mathbf{w}')$ *для любых чисел* a и b и *для любых векторов* $\mathbf{v}'$ и $\mathbf{w}'$. Самый простой способ доказать это состоит в следующем. Сначала рассмотрим частный случай, когда вектор $\mathbf{v} = \begin{pmatrix} c \\ 0 \end{pmatrix}$ лежит на оси x. Тогда для вектора $\mathbf{v}' = \begin{pmatrix} x' \\ y' \end{pmatrix}$ скалярное произведение $(\mathbf{v}, \mathbf{v}') = cx'$. Очевидно, что это выражение зависит линейно от $\mathbf{v}'$, что доказывает наше утверждение в частном случае. Теперь пусть у нас вектор $\mathbf{v}$ произволен. Можно найти преобразование вращения M, совмещающее вектор $\mathbf{v}$ с осью x. При этом $(\mathbf{v}, \mathbf{v}') = (M\mathbf{v}, M\mathbf{v}')$ линейно зависит от $M\mathbf{v}'$ и $M\mathbf{v}'$ линейно зависит от $\mathbf{v}'$, что и требовалось доказать. Сделаем это доказательство более подробно.

$$\begin{aligned}(\mathbf{v}, a\mathbf{v}' + b\mathbf{w}') &\overset{(a)}{=} (M\mathbf{v}, M(a\mathbf{v}' + b\mathbf{w}')) \overset{(b)}{=} (M\mathbf{v}, aM\mathbf{v}' + bM\mathbf{w}') \\ &\overset{(c)}{=} a(M\mathbf{v}, M\mathbf{v}') + b(M\mathbf{v}, M\mathbf{w}') \overset{(d)}{=} a(\mathbf{v}, \mathbf{v}') + b(\mathbf{v}, \mathbf{w}'),\end{aligned}$$

где (a) следует из ортогональности M, (b) — из линейности M,

(c) — доказано в частном случае, что $M\mathbf{v}$ лежит на оси x, (d) также следует из ортогональности M. Так как скалярное произведение $(\mathbf{v},\mathbf{v}')$ симметрично по $\mathbf{v}$ и $\mathbf{v}'$, то $(\mathbf{v},\mathbf{v}')$ будет линейной функцией $\mathbf{v}$ при фиксированном $\mathbf{v}'$. Это позволяет нам получить формулу для скалярного произведения. Пусть мы имеем два вектора:

$$\mathbf{v}=\begin{pmatrix}x\\y\end{pmatrix}=x\begin{pmatrix}1\\0\end{pmatrix}+y\begin{pmatrix}0\\1\end{pmatrix} \quad \text{и} \quad \mathbf{v}'=\begin{pmatrix}x'\\y'\end{pmatrix}=x'\begin{pmatrix}1\\0\end{pmatrix}+y'\begin{pmatrix}0\\1\end{pmatrix}.$$

Базисные вектора $\begin{pmatrix}1\\0\end{pmatrix}$ и $\begin{pmatrix}0\\1\end{pmatrix}$ ортогональны друг другу и имеют длину, равную единице. Тогда

$$\begin{aligned}(\mathbf{v},\mathbf{v}') &\overset{(a)}{=} x\left(\begin{pmatrix}1\\0\end{pmatrix},\mathbf{v}'\right)+y\left(\begin{pmatrix}0\\1\end{pmatrix},\mathbf{v}'\right)\\ &\overset{(b)}{=} xx'\left(\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}1\\0\end{pmatrix}\right)+xy'\left(\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}\right)\\ &\quad + yx'\left(\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}\right)+yy'\left(\begin{pmatrix}0\\1\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}\right)\\ &\overset{(c)}{=} xx'+yy'.\end{aligned}$$

Обоснования: (a) следует из линейности по $\mathbf{v}$, (b) следует из линейности по $\mathbf{v}'$, (c) — использованы соотношения

$$\left(\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}1\\0\end{pmatrix}\right)=\left(\begin{pmatrix}0\\1\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}\right)=1$$

и

$$\left(\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}\right)=\left(\begin{pmatrix}0\\1\end{pmatrix},\begin{pmatrix}1\\0\end{pmatrix}\right)=0.$$

Таким образом, мы получили удобную формулу для скалярного произведения двух векторов, лежащих в одной плоскости:

$$(\mathbf{v},\mathbf{v}')=xx'+yy'.$$

Перечислим важные свойства скалярного произведения в евклидовом пространстве:

1. симметрия: $(\mathbf{v}, \mathbf{v}') = (\mathbf{v}', \mathbf{v})$;
2. линейность: $(\mathbf{v}, a\mathbf{v}' + b\mathbf{w}') = a(\mathbf{v}, \mathbf{v}') + b(\mathbf{v}, \mathbf{w}')$;
3. положительная определенность: $(\mathbf{v}, \mathbf{v}) \geq 0$ и $(\mathbf{v}, \mathbf{v}) = 0$, только если $\mathbf{v} = 0$.

Воспользовавшись этими свойствами, можно выразить скалярное произведение через длины векторов. Рассмотрим тождество:

$$\|\mathbf{v} - \mathbf{w}\|^2 = (\mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w}).$$

Скалярное произведение линейно по обоим векторам, поэтому

$$\|\mathbf{v} - \mathbf{w}\|^2 = (\mathbf{v}, \mathbf{v}) - (\mathbf{v}, \mathbf{w}) - (\mathbf{w}, \mathbf{v}) + (\mathbf{w}, \mathbf{w}).$$

Так как $(\mathbf{v}, \mathbf{v}) = \|\mathbf{v}\|^2$, $(\mathbf{w}, \mathbf{w}) = \|\mathbf{w}\|^2$ и $(\mathbf{v}, \mathbf{w}) = (\mathbf{w}, \mathbf{v})$, мы получаем

$$2(\mathbf{v}, \mathbf{w}) = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - \|\mathbf{v} - \mathbf{w}\|^2.$$

Следовательно,

$$\boxed{(\mathbf{v}, \mathbf{w}) = \frac{1}{2}\{\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - \|\mathbf{v} - \mathbf{w}\|^2\}.} \tag{4.1}$$

Эта формула показывает, что евклидово скалярное произведение определяется евклидовым понятием длины. Если мы напишем $(\mathbf{v}, \mathbf{w}) = \|\mathbf{v}\| \cdot \|\mathbf{w}\| \cos\theta$ и посмотрим на рис. 4.6, то увидим, что формула 4.1 — это просто завуалированная теорема косинусов.

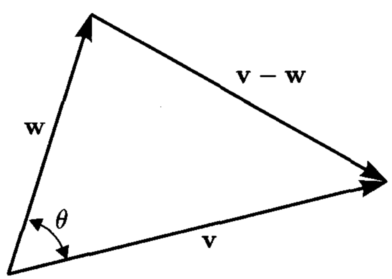


Рис. 4.6

4.2. Процесс Грама–Шмидта

Пусть V — абстрактное двумерное векторное пространство. Допустим, что в нем дано положительно определенное скалярное произведение $(\cdot,\cdot)_V$. Другими словами, допустим, что у нас есть функция, которая каждой паре векторов $\mathbf{v}_1$, $\mathbf{v}_2$ из V приписывает вещественное число $(\mathbf{v}_1,\mathbf{v}_2)_V$, удовлетворяющее условиям симметричности, билинейности и положительной определенности. Тогда мы утверждаем, что существует линейный изоморфизм[1] $L : V \to \mathbb{R}^2$ такой, что

$$(\mathbf{v}_1,\mathbf{v}_2)_V = (L\mathbf{v}_1, L\mathbf{v}_2).$$

Другими словами, выбрав правильный базис, можно сделать так, чтобы скалярное произведение $(\cdot,\cdot)_V$ в пространстве V выглядело так же, как евклидово скалярное произведение $(\cdot,\cdot)$ в пространстве $\mathbb{R}^2$. Для доказательства этого утверждения выберем в пространстве V ненулевой вектор $\mathbf{w}$. Так как $\|\mathbf{w}\|_V^2 = (\mathbf{w},\mathbf{w})_V > 0$, длина вектора

$$\mathbf{e}_1 = \frac{\mathbf{w}}{\|\mathbf{w}\|_V}$$

равна единице, т. е.

$$\|\mathbf{e}_1\|^2 = (\mathbf{e}_1,\mathbf{e}_2)_V = \frac{1}{\|\mathbf{w}\|_V^2}(\mathbf{w},\mathbf{w})_V = 1.$$

Теперь пусть u — вектор в пространстве V, линейно независимый от $\mathbf{e}_1$. Так как это пространство двумерно, то такой вектор существует. Пусть

$$\mathbf{u}_2 = \mathbf{u} - (\mathbf{u},\mathbf{e}_1)_V\mathbf{e}_1.$$

Вектор $\mathbf{u}_2$ перпендикулярен $\mathbf{e}_1$, т. е. $(\mathbf{u}_2,\mathbf{e}_1)_V = 0$. Действительно,

$$\begin{aligned}(\mathbf{u}_2,\mathbf{e}_1)_V &= (\mathbf{u} - (\mathbf{u},\mathbf{e}_1)_V\mathbf{e}_1, \mathbf{e}_1)_V\\ &= (\mathbf{u},\mathbf{e}_1)_V - (\mathbf{u},\mathbf{e}_1)_V(\mathbf{e}_1,\mathbf{e}_1)_V\\ &= (\mathbf{u},\mathbf{e}_1)_V - (\mathbf{u},\mathbf{e}_1)_V = 0,\end{aligned}$$

[1] Напоминаем, что в параграфе 1.12 мы определили понятие изоморфизма: L — линейное, взаимно-однозначное, сюрьективное преобразование, и, следовательно, имеющее обратное линейное преобразование.

так как $(\mathbf{e}_1, \mathbf{e}_1)_V = 1$. Кроме того, $\mathbf{u}_2 \neq 0$, так как в противном случае вектора $\mathbf{e}_1$ и $\mathbf{u}$ были бы линейно зависимы. Определим вектор

$$\mathbf{e}_2 = \frac{1}{\|\mathbf{u}_2\|_V}\mathbf{u}_2.$$

Тогда $(\mathbf{e}_2, \mathbf{e}_1)_V = (1/\|\mathbf{u}_2\|_V)(\mathbf{u}_2, \mathbf{e}_1)_V = 0$ и $\|\mathbf{e}_2\| = 1$. Выберем эти векторы $\mathbf{e}_1$, $\mathbf{e}_2$ в качестве базиса в пространстве V. Тогда любой вектор в этом пространстве можно представить в форме

$$\mathbf{v} = x\mathbf{e}_1 + y\mathbf{e}_2.$$

Заметим, что $x = (\mathbf{v}, \mathbf{e}_1)$, так как $(\mathbf{e}_2, \mathbf{e}_1) = 0$ и $(\mathbf{e}_1, \mathbf{e}_1) = 1$. Аналогично, $y = (\mathbf{v}, \mathbf{e}_2)$.

Пусть

$$\mathbf{v}_1 = x_1\mathbf{e}_1 + y_1\mathbf{e}_2,$$

$$\mathbf{v}_2 = x_2\mathbf{e}_1 + y_2\mathbf{e}_2.$$

Возьмем отображение $L : V \to \mathbb{R}^2$ в нашем базисе

$$L\mathbf{v}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \quad \text{и} \quad L\mathbf{v}_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}.$$

Тогда

$$(\mathbf{v}_1, \mathbf{v}_2)_V = (x_1\mathbf{e}_1 + y_1\mathbf{e}_2,\ x_2\mathbf{e}_1 + y_2\mathbf{e}_2)_V = x_1x_2 + y_1y_2 = (L\mathbf{v}_1, L\mathbf{v}_2),$$

т. к. $(\mathbf{e}_1, \mathbf{e}_2)_V = (\mathbf{e}_2, \mathbf{e}_1)_V = 0$ и $(\mathbf{e}_1, \mathbf{e}_1)_V = (\mathbf{e}_2, \mathbf{e}_2)_V = 1$. Наше утверждение доказано.

Определим евклидово скалярное произведение в пространстве $\mathbb{R}^3$ формулой

$$\left(\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}\right) = x_1x_2 + y_1y_2 + z_1z_2.$$

Как и в двумерном случае, если $\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, то $\|\mathbf{v}\|^2 = (\mathbf{v}, \mathbf{v})$ является квадратом евклидовой длины вектора $\mathbf{v}$. Приведенные выше соображения показывают, что, воспользовавшись формулой 4.1, по длине вектора можно восстановить скалярное произведение:

$$(\mathbf{v}, \mathbf{w}) = \frac{1}{2}(\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - \|\mathbf{v} - \mathbf{w}\|^2).$$

Таким образом, вращения в трехмерном пространстве сохраняют скалярное произведение. В частности, если нам даны два вектора $\mathbf{v}$ и $\mathbf{w}$, то можно повернуть плоскость, в которой они расположены, так, чтобы она совпала с плоскостью $z = 0$. Для векторов в такой плоскости скалярное произведение сводится к скалярному произведению в пространстве $\mathbb{R}^2$. А для таких векторов мы уже получили, что

$$(\mathbf{v}, \mathbf{w}) = \|\mathbf{v}\| \cdot \|\mathbf{w}\| \cos\theta,$$

и, следовательно, это справедливо для любой пары векторов (поскольку обе части равенства инвариантны относительно вращений).

Векторное пространство V называется трехмерным, если любые четыре вектора линейно зависимы, но существуют три линейно независимых. Это значит, что для любых четырех векторов $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, $\mathbf{v}_4$ можно найти четыре числа a_1, a_2, a_3, a_4, не все равные нулю, при которых выполняется равенство

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + a_4\mathbf{v}_4 = 0,$$

и при этом существуют три вектора $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ такие, что равенство

$$a\mathbf{u} + b\mathbf{v} + c\mathbf{w} = 0$$

выполняется только при условии $a = b = c = 0$.

Допустим, что в пространстве V есть положительно определенное скалярное произведение $(\cdot, \cdot)_V$. Повторим построение, приведенное выше для двумерного случая. Выберем ненулевой вектор. Обозначим его $\mathbf{e}_1$. Выберем другой вектор $\mathbf{u}$, линейно независимый с вектором $\mathbf{e}_1$. Положим

$$\mathbf{u}_2 = \mathbf{u} - (\mathbf{u}, \mathbf{e}_1)_V \mathbf{e}_1,$$
$$\mathbf{e}_2 = \frac{1}{\|\mathbf{u}_2\|_V}\mathbf{u}_2.$$

Векторы $\mathbf{e}_1$ и $\mathbf{e}_2$ удовлетворяют условиям

$$\|\mathbf{e}_1\|_V = \|\mathbf{e}_2\|_V = 1, \qquad (\mathbf{e}_1, \mathbf{e}_2)_V = 0.$$

Множество векторов вида $x\mathbf{e}_1 + y\mathbf{e}_2$ изоморфно $\mathbb{R}^2$ и, следовательно, образует двумерное векторное пространство. Оно не заполняет все пространство V, так как в этом множестве мы не можем найти три линейно независимых вектора. Таким образом, в пространстве V существует вектор $\mathbf{w}$, который нельзя записать в форме $x\mathbf{e}_1 + y\mathbf{e}_2$. Значит, вектор

$$\mathbf{w}_3 = \mathbf{w} - (\mathbf{w}, \mathbf{e}_1)_V \mathbf{e}_1 - (\mathbf{w}, \mathbf{e}_2)_V \mathbf{e}_2$$

не равен нулю. Определим вектор

$$\mathbf{e}_3 = \frac{1}{\|\mathbf{w}_3\|_V} \mathbf{w}_3.$$

Тогда

$$\|\mathbf{e}_1\|_V = \|\mathbf{e}_2\|_V = \|\mathbf{e}_3\|_V = 1,$$
$$(\mathbf{e}_1, \mathbf{e}_2)_V = (\mathbf{e}_1, \mathbf{e}_3)_V = (\mathbf{e}_2, \mathbf{e}_3)_V = 0.$$

Если $\mathbf{v}$ — произвольный вектор в пространстве V, то мы утверждаем, что

$$\mathbf{v} - (\mathbf{v}, \mathbf{e}_1)_V \mathbf{e}_1 - (\mathbf{v}, \mathbf{e}_2)_V \mathbf{e}_2 - (\mathbf{v}, \mathbf{e}_3)_V \mathbf{e}_3 = \mathbf{0}.$$

Действительно, воспользовавшись теми же соображениями, что и раньше, положим

$$\mathbf{v}_4 = \mathbf{v} - (\mathbf{v}, \mathbf{e}_1)\mathbf{e}_1 - (\mathbf{v}, \mathbf{e}_2)\mathbf{e}_2 - (\mathbf{v}, \mathbf{e}_3)\mathbf{e}_3.$$

Тогда

$$(\mathbf{v}_4, \mathbf{e}_1)_V = (\mathbf{v}_4, \mathbf{e}_2)_V = (\mathbf{v}_4, \mathbf{e}_3)_V = 0.$$

Если предположить, что $\mathbf{v}_4 \neq \mathbf{0}$, то можно доказать, что векторы $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{v}_4$ линейно независимы. В самом деле, пусть выполнено равенство

$$a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3 + a_4\mathbf{v}_4 = \mathbf{0}.$$

Домножим обе его части на $\mathbf{e}_1$, $\mathbf{e}_2$ и $\mathbf{e}_3$. В результате получим, что $a_1 = 0$, $a_2 = 0$, $a_3 = 0$. Следовательно, если $\mathbf{v}_4 \neq \mathbf{0}$, то

$a_4 = 0$. Но это противоречит исходному предположению, что пространство V трехмерное.

Итак, любой вектор в пространстве V может быть записан в форме

$$\mathbf{v} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3, \quad \text{где} \quad x = (\mathbf{v}, \mathbf{e}_1), \ \ y = (\mathbf{v}, \mathbf{e}_2), \ \ z = (\mathbf{v}, \mathbf{e}_3).$$

Точно также, как и в двумерном случае, определим отображение $L : V \to \mathbb{R}^3$ равенством

$$L(\mathbf{v}) = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \text{если} \quad \mathbf{v} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3.$$

Это отображение является линейным, взаимно-однозначным отображением V на пространство $\mathbb{R}^3$ и

$$(\mathbf{u}, \mathbf{v})_V = (L\mathbf{u}, L\mathbf{v})_{\mathbb{R}^3}.$$

Аналогичное утверждение можно доказать и для случаев четырех, пяти, ..., n измерений.

В пространстве $\mathbb{R}^n$ определим *евклидово скалярное произведение*

$$\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}\right) = x_1 w_1 + \cdots + x_n w_n.$$

Векторное пространство V называется n-мерным, если существует n линейно независимых векторов, при этом любой набор из $n+1$ векторов является линейно зависимым. Общую теорию n-мерных векторных пространств мы будем изучать в главе 10. Если V — n-мерное пространство, имеющее положительно определенное скалярное произведение, то в нем можно найти *ортогональный базис* $\mathbf{e}_1, \ldots, \mathbf{e}_n$, т. е. можно найти n векторов $\mathbf{e}_1, \ldots, \mathbf{e}_n$, обладающих свойствами

$$\|\mathbf{e}_1\|_V = \|\mathbf{e}_2\|_V = \cdots = \|\mathbf{e}_n\|_V = 1$$

и

$$(\mathbf{e}_i, \mathbf{e}_j)_V = 0, \quad i \neq j.$$

Любой вектор $\mathbf{v}$ в пространстве V может быть записан в форме

$$\mathbf{v} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n, \quad x_i = (\mathbf{v}, \mathbf{e}_i).$$

Соответственно определяется отображение $L : V \to \mathbb{R}^n$

$$L(\mathbf{v}) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Тогда

$$(\mathbf{v}, \mathbf{w})_V = (L\mathbf{v}, L\mathbf{w}).$$

Итак, если у нас есть n независимых векторов $\mathbf{v}_1, \ldots \mathbf{v}_n$, то векторы $\mathbf{e}_i$ можно последовательно получить с помощью следующего алгоритма:

$$\mathbf{e}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1,$$

$$\mathbf{u}_2 = \mathbf{v}_2 - (\mathbf{v}_2, \mathbf{e}_1)\mathbf{e}_1,$$

$$\mathbf{e}_2 = \frac{1}{\|\mathbf{u}_2\|}\mathbf{u}_2,$$

$$\mathbf{u}_3 = \mathbf{v}_3 - (\mathbf{v}_3, \mathbf{e}_1)\mathbf{e}_1 - (\mathbf{v}_3, \mathbf{e}_2)\mathbf{e}_2,$$

$$\mathbf{e}_3 = \frac{1}{\|\mathbf{u}_3\|}\mathbf{u}_3,$$

и так далее. Этот алгоритм называется *ортонормировочной процедурой Грама–Шмидта.*

Покажем начало процедуры Грама–Шмидта в применении к векторам $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, $\mathbf{v}_4$ пространства $\mathbb{R}^4$, где

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 3 \\ 3 \\ 5 \\ 3 \end{pmatrix}.$$

(Мы сделаем только два первых шага, поэтому векторы $\mathbf{v}_3$, $\mathbf{v}_4$ нам не потребуются.) Скалярное произведение в пространстве

$\mathbb{R}^4$ определяется, как обычно. Иными словами, мы предполагаем, что векторы

$$\begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

образуют ортонормированный базис.

Первый шаг состоит в преобразовании вектора $\mathbf{v}_1$ к единичному вектору:

$$(\mathbf{v}_1, \mathbf{v}_1) = 1^2 + 1^2 + 1^2 + (-1)^2 = 4,$$

поэтому

$$\sqrt{(\mathbf{v}_1, \mathbf{v}_1)} = 2,$$

следовательно,

$$\mathbf{e}_1 = \frac{1}{2}\mathbf{v}_1 = \begin{pmatrix}1/2\\1/2\\1/2\\-1/2\end{pmatrix}.$$

Далее, из вектора $\mathbf{v}_2$ вычитаем его проекцию на вектор $\mathbf{e}_1$:

$$\begin{aligned}\mathbf{w}_2 &= \begin{pmatrix}3\\3\\5\\3\end{pmatrix} - \left(\begin{pmatrix}3\\3\\5\\3\end{pmatrix}, \begin{pmatrix}1/2\\1/2\\1/2\\-1/2\end{pmatrix}\right) \cdot \begin{pmatrix}1/2\\1/2\\1/2\\-1/2\end{pmatrix} \\ &= \begin{pmatrix}3\\3\\5\\3\end{pmatrix} - (3/2 + 3/2 + 5/2 - 3/2)\begin{pmatrix}1/2\\1/2\\1/2\\-1/2\end{pmatrix} = \begin{pmatrix}1\\1\\3\\5\end{pmatrix}.\end{aligned}$$

И наконец, преобразуем вектор $\mathbf{w}_2$ в единичный вектор:

$$(\mathbf{w}_2, \mathbf{w}_2) = \left(\begin{pmatrix}1\\1\\3\\5\end{pmatrix}, \begin{pmatrix}1\\1\\3\\5\end{pmatrix}\right) = 1^2 + 1^2 + 3^2 + 5^2 = 36,$$

поэтому

$$\mathbf{e}_2 = \frac{1}{6}\mathbf{w}_2 = \frac{1}{6}\begin{pmatrix}1\\1\\3\\5\end{pmatrix}.$$

В качестве проверки заметим, что $\left(\begin{pmatrix}1\\1\\1\\-1\end{pmatrix}, \begin{pmatrix}1\\1\\3\\5\end{pmatrix}\right) = 0.$

Заметим, что любой вектор $\mathbf{v}$ в пространстве $\mathbb{R}^4$ можно записать как сумму вектора $\pi\mathbf{v}$, который является линейной комбинацией векторов $\mathbf{e}_1$, $\mathbf{e}_2$ (и, следовательно, векторов $\mathbf{v}_1$ и $\mathbf{v}_2$), и вектора, перпендикулярного $\mathbf{e}_1$ и $\mathbf{e}_2$. Например, возьмем вектор

$$\mathbf{v} = \begin{pmatrix}4\\0\\-1\\7\end{pmatrix}.$$

Определим вектор $\pi\mathbf{v}$ следующим образом:

$$\pi\mathbf{v} = (\mathbf{v}, \mathbf{e}_1)\mathbf{e}_1 + (\mathbf{v}, \mathbf{e}_2)\mathbf{e}_2,$$

$$\begin{aligned}\pi\mathbf{v} &= \frac{1}{4}\left(\begin{pmatrix}4\\0\\-1\\7\end{pmatrix}, \begin{pmatrix}1\\1\\1\\-1\end{pmatrix}\right)\cdot\begin{pmatrix}1\\1\\1\\-1\end{pmatrix} + \frac{1}{36}\left(\begin{pmatrix}4\\0\\-1\\7\end{pmatrix}, \begin{pmatrix}1\\1\\3\\5\end{pmatrix}\right)\cdot\begin{pmatrix}1\\1\\3\\5\end{pmatrix}\\ &= \frac{1}{4}(-4)\begin{pmatrix}1\\1\\1\\-1\end{pmatrix} + \frac{1}{36}\cdot 36\begin{pmatrix}1\\1\\3\\5\end{pmatrix} = \begin{pmatrix}-1\\-1\\-1\\1\end{pmatrix} + \begin{pmatrix}1\\1\\3\\5\end{pmatrix} = \begin{pmatrix}0\\0\\2\\6\end{pmatrix}.\end{aligned}$$

Можно проверить, что вектор

$$\mathbf{v} - \pi\mathbf{v} = \begin{pmatrix}4\\0\\-1\\7\end{pmatrix} - \begin{pmatrix}0\\0\\2\\6\end{pmatrix} = \begin{pmatrix}4\\0\\-3\\1\end{pmatrix}$$

перпендикулярен векторам $\mathbf{e}_1$ и $\mathbf{e}_2$. (Это можно сделать прямой проверкой ортогональности этого вектора исходным базисным векторам $\mathbf{v}_1$ и $\mathbf{v}_2$, либо ортонормированной системе векторов $\mathbf{e}_1$ и $\mathbf{e}_2$.) Мы будем говорить, что преобразование π, переводящее вектор $\mathbf{v}$ в $\pi\mathbf{v}$, есть *ортогональная проекция* на подпространство W, натянутое на векторы $\mathbf{v}_1$ и $\mathbf{v}_2$.

В качестве второго примера применения процедуры Грама–Шмидта рассмотрим четырехмерное пространство полиномов степени ≤ 3, где скалярное произведение определяется формулой

$$(\mathbf{f}, \mathbf{g}) = \int_{-1}^{1} f(t)g(t)\,dt.$$

(Проверьте, что эта формула действительно определяет скалярное произведение!) Начнем со степенного базиса

$$\mathbf{v}_1 = 1, \quad \mathbf{v}_2 = t, \quad \mathbf{v}_3 = t^2, \quad \mathbf{v}_4 = t^3.$$

(Если бы начали с других базисных элементов, либо с теми же элементами, но в другом порядке, то могли бы получить другой ортонормированный базис.) Сначала вычислим

$$(\mathbf{v}_1, \mathbf{v}_1) = \int_{-1}^{1} dt = 2$$

и преобразуем вектор $\mathbf{v}_1$ к единичному вектору:

$$\mathbf{e}_1 = \mathbf{v}_1 \sqrt{(\mathbf{v}_1, \mathbf{v}_1)} = 1/\sqrt{2}.$$

Затем вычислим

$$(\mathbf{e}_1, \mathbf{v}_2) = \int_{-1}^{1} (t/\sqrt{2})\,dt = 0,$$

откуда следует, что вектор $\mathbf{v}_2$ уже ортогонален вектору $\mathbf{e}_1$. Поскольку

$$(\mathbf{v}_2, \mathbf{v}_2) = \int_{-1}^{1} t^2\,dt = 2/3,$$

то получаем $\mathbf{e}_2 = t/\sqrt{(2/3)} = \sqrt{(3/2)}t$.

Теперь можно вычислить вектор $\mathbf{w}_3$:

$$\begin{aligned}\mathbf{w}_3 &= \mathbf{v}_3 - \mathbf{e}_1(\mathbf{e}_1, \mathbf{v}_3) - \mathbf{e}_2(\mathbf{e}_2, \mathbf{v}_3)\\ &= t^2 - \frac{1}{2}\int_{-1}^{1} t^2\,dt - \frac{3}{2}t\int_{-1}^{1} t^3\,dt\\ &= t^2 - \frac{1}{3}.\end{aligned}$$

Вычисляем скалярное произведение

$$(\mathbf{w}_3, \mathbf{w}_3) = \int_{-1}^{1}\left(t^2 - \frac{1}{3}\right)dt = \frac{8}{45}$$

и получаем третий нормированный базисный вектор

$$\mathbf{e}_3 = \frac{t^2 - \frac{1}{3}}{\sqrt{\frac{8}{45}}} = \sqrt{\frac{45}{8}}\left(t^2 - \frac{1}{3}\right) = \sqrt{\frac{5}{8}}(3t^2 - 1).$$

Наконец, вычисляем вектор $\mathbf{w}_4$:

$$\begin{aligned}\mathbf{w}_4 &= \mathbf{v}_4 - \mathbf{e}_1(\mathbf{e}_1, \mathbf{v}_4) - \mathbf{e}_2(\mathbf{e}_2, \mathbf{v}_4) - \mathbf{e}_3(\mathbf{e}_3, \mathbf{v}_4)\\ &= t^3 - \frac{1}{2}\int_{-1}^{1} t^3\,dt - \frac{3}{2}t\int_{-1}^{1} t^4\,dt - \frac{5}{8}(3t^2 - 1)\int_{-1}^{1}(3t^5 - t^3)\,dt\\ &= t^2 - 0 - \frac{3}{2}t\cdot\frac{2}{5} - 0 = t^3 - \frac{3}{5}t.\end{aligned}$$

Разделив его на $\sqrt{(\mathbf{w}_4, \mathbf{w}_4)}$, получаем четвертый нормированный вектор

$$\mathbf{e}_4 = \frac{\mathbf{w}_4}{\sqrt{(\mathbf{w}_4, \mathbf{w}_4)}} = \sqrt{\frac{7}{8}}(5t^3 - 3t).$$

Очевидно, что действуя таким образом, мы можем построить последовательность ортогональных полиномов более высоких степеней. Эти полиномы, известные как полиномы Лежандра, появляются при решении задач электростатики в сферических координатах. На самом деле, обычно при решении физических задач векторные пространства функций, получающихся при решении дифференциальных уравнений, имеют ортогональный базис, возникающий естественным образом из физических соображений. Поэтому на практике не так уж часто приходится прибегать к громоздкой процедуре Грама–Шмидта.

4.3. Квадратичные формы и симметричные матрицы

В параграфах 4.1 и 4.2 мы изучали *евклидовы* скалярные произведения, обладающие тремя свойствами: они билинейны, симметричны и положительно определены. Сейчас мы займемся изучением «скалярных проиизведений» более общего вида, когда положительная определенность не обязательна. Эти скалярные произведения играют центральную роль в теории относительности.

Возвращаемся в пространство $\mathbb{R}^2$. Допустим, что в этом пространстве нам дано скалярное произведение $\langle\cdot,\cdot\rangle$, которое не обязательно будет положительно определенным. При этом мы полагаем, что оно

$$\text{билинейно:}\quad \langle\mathbf{v}, a\mathbf{u}+b\mathbf{w}\rangle = a\langle\mathbf{v},\mathbf{u}\rangle + b\langle\mathbf{v},\mathbf{w}\rangle$$

$$\text{и симметрично:}\quad \langle\mathbf{u},\mathbf{v}\rangle = \langle\mathbf{v},\mathbf{u}\rangle$$

для всех векторов $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ и всех вещественных чисел a и b. Интересно сравнить скалярное произведение $\langle\cdot,\cdot\rangle$ с евклидовым скалярным произведением $(\cdot,\cdot)$. Начнем с элементарной леммы.

Пусть $l : \mathbb{R}^2 \to \mathbb{R}$ есть линейное отображение. Тогда существует единственный вектор $\mathbf{w}$, для которого

$$l(\mathbf{v}) = (\mathbf{v},\mathbf{w}) \quad \text{для всех} \quad \mathbf{v} \ \text{ из } \ \mathbb{R}^2.$$

Действительно, l определяется матрицей 1×2: $(a\ b)$, т. е.

$$l\begin{pmatrix} x \\ y \end{pmatrix} = ax+by \quad \text{для любого } \mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2.$$

Выберем вектор

$$\mathbf{w} = \begin{pmatrix} a \\ b \end{pmatrix}.$$

Тогда

$$(\mathbf{v}, \mathbf{w}) = \left(\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix}\right) = ax + by,$$

что мы и хотели доказать. Очевидно, что $\mathbf{w}$ — единственный вектор в пространстве $\mathbb{R}^2$, обладающий таким свойством. Теперь рассмотрим $\langle \mathbf{u}, \mathbf{v} \rangle$ как функцию $\mathbf{v}$ при фиксированном векторе $\mathbf{u}$. Это скалярное произведение линейно по $\mathbf{v}$, следовательно, существует такой вектор $\mathbf{w}$, для которого

$$\langle \mathbf{u}, \mathbf{v} \rangle = (\mathbf{v}, \mathbf{w}) \quad \text{для всех} \quad \mathbf{v} \in V.$$

Вектор $\mathbf{w}$ зависит от $\mathbf{u}$, поэтому в этом уравнении следует писать $\mathbf{w}(\mathbf{u})$. Итак, $\mathbf{w}(\mathbf{u})$ — это такой вектор, для которого *евклидово* скалярное произведение с любым вектором $\mathbf{v}$ равно $\langle \mathbf{u}, \mathbf{v} \rangle$. Пусть $\mathbf{u}_1$ и $\mathbf{u}_2$ — два вектора, а им соответствуют $\mathbf{w}(\mathbf{u}_1)$ и $\mathbf{w}(\mathbf{u}_2)$. Тогда

$$\begin{aligned}
\langle a\mathbf{u}_1 + b\mathbf{u}_2, \mathbf{v} \rangle &\overset{(a)}{=} \langle \mathbf{v}, a\mathbf{u}_1 + b\mathbf{u}_2 \rangle \\
&\overset{(b)}{=} a\langle \mathbf{v}, \mathbf{u}_1 \rangle + b\langle \mathbf{v}, \mathbf{u}_2 \rangle \\
&\overset{(c)}{=} a\langle \mathbf{u}_1, \mathbf{v} \rangle + b\langle \mathbf{u}_2, \mathbf{v} \rangle \\
&= a(\mathbf{v}, \mathbf{w}(\mathbf{u}_1)) + b(\mathbf{v}, \mathbf{w}(\mathbf{u}_2)) \\
&= (\mathbf{v}, a\mathbf{w}(\mathbf{u}_1) + b\mathbf{w}(\mathbf{u}_2)),
\end{aligned}$$

где (a) — следствие симметрии, (b) — следствие билинейности, (c) — следствие симметрии. Таким образом,

$$\mathbf{w}(a\mathbf{u}_1 + b\mathbf{u}_2) = a\mathbf{w}(\mathbf{u}_1) + b\mathbf{w}(\mathbf{u}_2).$$

Другими словами, $\mathbf{w}$ линейно зависит от $\mathbf{u}$. Но тогда можно написать $\mathbf{w}(\mathbf{u}) = A\mathbf{u}$, где A — линейное преобразование. Вводя прежние обозначения $\mathbf{w}(\mathbf{u}) = A\mathbf{u}$, получаем:

$$\boxed{\langle \mathbf{u}, \mathbf{v} \rangle = (\mathbf{v}, A\mathbf{u})}$$

для всех $\mathbf{u}$ и $\mathbf{v}$ в пространстве $\mathbb{R}^2$. До этого момента мы использовали только билинейность $\langle \mathbf{u}, \mathbf{v} \rangle$, т. е. линейность по $\mathbf{u}$ при фиксированном $\mathbf{v}$ и линейность по $\mathbf{v}$ при фиксированном $\mathbf{u}$. (Здесь, по существу, использована также симметрия $\langle\cdot,\cdot\rangle$.) Симметрия $\langle\cdot,\cdot\rangle$ позволяет записать

$$\langle\mathbf{u},\mathbf{v}\rangle = \langle\mathbf{v},\mathbf{u}\rangle,$$

следовательно,

$$(\mathbf{v}, A\mathbf{u}) = (\mathbf{u}, A\mathbf{v}).$$

Однако, $(\mathbf{u},\mathbf{v}) = (\mathbf{v},\mathbf{u})$, поэтому

$$(\mathbf{v}, A\mathbf{u}) = (A\mathbf{v}, \mathbf{u})$$

для всех векторов $\mathbf{u}$ и $\mathbf{v}$ в пространстве V. Посмотрим, что из этого можно получить для матрицы A.

Для любой матрицы B выражение $(B\mathbf{v}, \mathbf{u})$ линейно в отдельности по $\mathbf{u}$ и по $\mathbf{v}$. Воспользовавшись предыдущими аргументами, можно показать, что существует единственное линейное преобразование (обозначим его B^T), которое удовлетворяет условию

$$(B\mathbf{v}, \mathbf{u}) = (\mathbf{v}, B^T\mathbf{u})$$

для всех векторов $\mathbf{u}$ и $\mathbf{v}$ в пространстве V. Это преобразование называется *транспонированным* по отношению к B. Вычислим B^T, приняв обозначения

$$\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} x' \\ y' \end{pmatrix}, \quad B = \begin{pmatrix} e & f \\ g & h \end{pmatrix}.$$

Тогда

$$\begin{aligned}(B\mathbf{v}, \mathbf{u}) &= (ex + fy)x' + (gx + hy)y' \\ &= exx' + fyx' + gxy' + hyy' \\ &= x(ex' + gy') + y(fx' + hy'),\end{aligned}$$

следовательно,

$$B^T \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} ex' + gy' \\ fx' + hy' \end{pmatrix},$$

или

$$B^T = \begin{pmatrix} e & g \\ f & h \end{pmatrix}.$$

Другими словами, транспонированная матрица B^T получается отражением матрицы B относительно главной диагонали.

Отсюда для матрицы A, определяющей наше «скалярное произведение», следует, что

$$\boxed{A = A^T,}$$

т. е. матрица A *симметрична*. Значит, A может быть записана в виде

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

Если мы обозначим

$$Q(\mathbf{v}) = \langle \mathbf{v}, \mathbf{v} \rangle,$$

тогда, как и в параграфе 4.1, можем получить тождество

$$\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{2}(Q(\mathbf{u}) + Q(\mathbf{v}) - Q(\mathbf{u} - \mathbf{v})),$$

и если $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$, то

$$Q(\mathbf{v}) = (A\mathbf{v}, \mathbf{v}) = ax^2 + 2bxy + cy^2.$$

Такая функция Q называется *квадратичной формой*. Таким образом, предыдущие формулы показывают, что квадратичная форма Q определяет $\langle \cdot, \cdot \rangle$, и обратно, каждое скалярное произведение определяет квадратичную форму.

Коэффициенты a, $2b$, c в квадратичном полиноме $Q(\mathbf{v})$ определяют матрицу A. Другими словами, Q определяет A и, следовательно, $\langle \cdot, \cdot \rangle$.

Характеристический полином для матрицы A имеет вид

$$x^2 - (a + c)x + ac - b^2,$$

при этом выполнено условие

$$(a + c)^2 - 4(ac - b^2) = (a - c)^2 + 4b^2 \geq 0.$$

Выражение $(a - c)^2 + 4b^2$ называется *дискриминантом* квадратичной формы Q. Дискриминант равен нулю в том и только том случае, если $a = c$ и $b = 0$. Тогда

$$A = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} = a\mathbb{I}$$

и

$$\langle \mathbf{u}, \mathbf{v} \rangle = a(\mathbf{u}, \mathbf{v}).$$

В этом случае $\langle \cdot, \cdot \rangle$ пропорционально $(\cdot, \cdot)$.

Предположим, что A имеет различные собственные значения $\lambda_1 \neq \lambda_2$, соответствующие собственным векторам $\mathbf{v}_1$ и $\mathbf{v}_2$. Докажем, что вектор $\mathbf{v}_1$ перпендикулярен вектору $\mathbf{v}_2$, т. е. $(\mathbf{v}_1, \mathbf{v}_2) = 0$. Это очень просто: $(A\mathbf{v}_1, \mathbf{v}_2) = (\mathbf{v}_1, A\mathbf{v}_2)$, т. к. A симметрична; $(\lambda_1\mathbf{v}_1, \mathbf{v}_2) = (\mathbf{v}_1, \lambda_2\mathbf{v}_2)$, т. к. $\mathbf{v}_1$, $\mathbf{v}_2$ — собственные векторы; $\lambda_1(\mathbf{v}_1, \mathbf{v}_2) = \lambda_2(\mathbf{v}_1, \mathbf{v}_2)$, т. к. скалярное произведение линейно; $(\mathbf{v}_1, \mathbf{v}_2) = 0$, потому что $\lambda_1 \neq \lambda_2$.

Докажем обратное утверждение: пусть собственный вектор $\mathbf{v}_1$ матрицы A соответствует собственному значению λ_1. Пусть вектор $\mathbf{v}_2$ не равен нулю и перпендикулярен $\mathbf{v}_1$, т. е. $(\mathbf{v}_1, \mathbf{v}_2) = 0$. Тогда

$$(\mathbf{v}_1, A\mathbf{v}_2) = (A\mathbf{v}_1, \mathbf{v}_2) = \lambda_1(\mathbf{v}_1, \mathbf{v}_2) = 0,$$

следовательно, $A\mathbf{v}_2$ перпендикулярен $\mathbf{v}_1$. Но существует только одна прямая, перпендикулярная вектору $\mathbf{v}_1$, и вектор $\mathbf{v}_2 \neq 0$ лежит на ней. Это значит, что $A\mathbf{v}_2$ пропорционален $\mathbf{v}_2$, т. е. $A\mathbf{v}_2 = \lambda_2\mathbf{v}_2$ для некоторого собственного значения λ_2.

Итак, мы доказали, что любая симметричная матрица A имеет два *ортогональных* собственных вектора $\mathbf{v}_1$ и $\mathbf{v}_2$. Умножив векторы $\mathbf{v}_1$ и $\mathbf{v}_2$ на соответствующие скаляры, можно сделать их длину равной единице. Тогда матрица $\begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}$, где $\mathbf{v}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ и $\mathbf{v}_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$, является матрицей вращения.

Таким образом, можно записать $A = R \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} R^{-1}$, где R — матрица вращения. Для ортогональной матрицы M выполняются соотношения $(M\mathbf{v}, M\mathbf{v}) = (\mathbf{v}, M^T M\mathbf{v}) = (\mathbf{v}, \mathbf{v})$ для всех векторов $\mathbf{v}$. Отсюда следует, что $M^T = M^{-1}$, и мы можем написать равносильное равенство:

$$A = R \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} R^T.$$

Предположим, что мы выбрали собственные векторы так, что $\lambda_1 > \lambda_2$. Тогда собственный вектор $\mathbf{v}_1$, уже преобразованный к единичной длине, отличается от всех других векторов единичной длины $\mathbf{v}$ тем, что $Q(\mathbf{v})$ принимает максимальное значение. Вектору $\mathbf{v}_2$, тоже имеющему единичную длину, соответствует минимальное значение $Q(\mathbf{v})$, т.е.

$$Q(\mathbf{v}_1) \geq Q(\mathbf{v}) \geq Q(\mathbf{v}_2)$$

для любого вектора $\mathbf{v}$ единичной длины ($(\mathbf{v},\mathbf{v}) = 1$). Чтобы доказать это утверждение, запишем $\mathbf{v} = \mathbf{v}_1 \cos\theta + \mathbf{v}_2 \sin\theta$ (рис. 4.7). Векторы $\mathbf{v}_1$ и $\mathbf{v}_2$ ортогональны и имеют единичную длину, следовательно,

$$(\mathbf{v},\mathbf{v}) = (\mathbf{v}_1,\mathbf{v}_1)\cos^2\theta + (\mathbf{v}_2,\mathbf{v}_2)\sin^2\theta = 1.$$

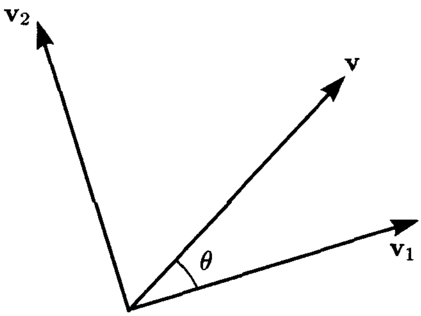


Рис. 4.7

Тогда

$$\begin{aligned}
Q(\mathbf{v}) &= (A\mathbf{v},\mathbf{v}) \\
&= (A\mathbf{v}_1\cos\theta + A\mathbf{v}_2\sin\theta,\ \mathbf{v}_1\cos\theta + \mathbf{v}_2\sin\theta) \\
&= (\lambda_1\mathbf{v}_1\cos\theta + \lambda_2\mathbf{v}_2\sin\theta,\ \mathbf{v}_1\cos\theta + \mathbf{v}_2\sin\theta) \\
&= \lambda_1(\mathbf{v}_1,\mathbf{v}_1)\cos^2\theta + \lambda_2(\mathbf{v}_2,\mathbf{v}_2)\sin^2\theta \quad (\text{т. к. } (\mathbf{v}_1,\mathbf{v}_2) = 0) \\
&= \lambda_1\cos^2\theta + \lambda_2\sin^2\theta \\
&= \lambda_1 - (\lambda_1 - \lambda_2)\sin^2\theta.
\end{aligned}$$

Очевидно, что $Q(\mathbf{v})$ имеет максимум при $\sin^2\theta = 0$, т.е когда $\mathbf{v} = \pm\mathbf{v}_1$, и принимает минимальное значение при $\sin^2\theta = 1$, т.е. когда $\mathbf{v} = \pm\mathbf{v}_2$.

Теперь понятно, как нарисовать линию $Q(\mathbf{v}) = (A\mathbf{v}, \mathbf{v}) = \mathrm{const}$. С помощью матрицы вращения R приведем A к диагональному виду:

$$A = R\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} R^{-1}.$$

Тогда

$$Q(\mathbf{v}) = \left(R\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} R^{-1}\mathbf{v}, \mathbf{v}\right).$$

Матрица R ортогональна, поэтому $R^T = R^{-1}$, и мы получаем

$$Q(\mathbf{v}) = \left(\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} R^{-1}\mathbf{v}, R^{-1}\mathbf{v}\right).$$

Если записать $\begin{pmatrix} x' \\ y' \end{pmatrix} = R^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = R^{-1}\mathbf{v}$, то

$$Q(\mathbf{v}) = \left(\begin{pmatrix} \lambda_1 & x' \\ \lambda_2 & y' \end{pmatrix}, \begin{pmatrix} x' \\ y' \end{pmatrix}\right) = \lambda_1 x'^2 + \lambda_2 y'^2.$$

Если оба собственные значения λ_1 и λ_2 положительны, то линия $Q(\mathbf{v}) = k$ — это эллипс при $k > 0$, начало координат при $k = 0$ и пустое множество при $k < 0$. Если λ_1 и λ_2 отрицательны, то линия будет эллипсом при $k < 0$. Если λ_1 и λ_2 имеют противоположные знаки, то линия $Q(\mathbf{v}) = k$ будет гиперболой, вырождающейся в две прямых при $k = 0$. Вершины эллипса или гиперболы, в которых расстояние от точки на кривой до начала координат имеет локальный экстремум, лежат на прямых, задаваемых собственными векторами матрицы A.

Рассмотрим пример, когда $A = \begin{pmatrix} 9 & 2 \\ 2 & 6 \end{pmatrix}$. Тогда

$$Q(\mathbf{v}) = 9x^2 + 4xy + 6y^2.$$

Собственные значения матрицы A таковы: $\lambda_1 = 10$, $\lambda_2 = 5$. Соответствующие им собственные векторы: $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ и $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$. Мы можем написать $A = R\begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix} R^{-1}$, где матрица вращения R равна

$$R = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}.$$

Введем новые координаты

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = R^{-1} \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

т. е.

$$x' = \frac{1}{\sqrt{5}}(2x + y),$$
$$y' = \frac{1}{\sqrt{5}}(-x + 2y).$$

В новых координатах

$$Q(\mathbf{v}) = 10x'^2 + 5y'^2.$$

График множества $Q(\mathbf{v}) = 1$, т. е.

$$10x'^2 + 5y'^2 = 1,$$

является эллипсом, у которого малая ось — $\sqrt{1/10}$, а большая равна $\sqrt{1/5}$. Оси расположены вдоль собственных векторов матрицы A : $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ и $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$, см. рис. 4.8.

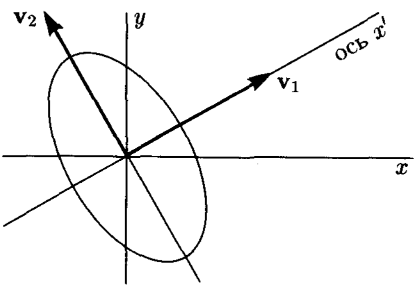


Рис. 4.8

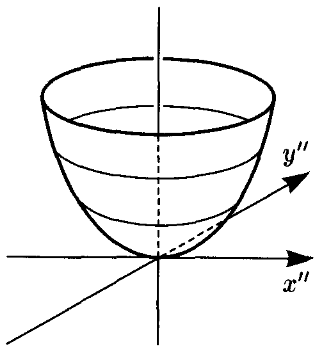


Рис. 4.9

Предположим, что помимо вращения, изменяющего координаты, возможно неортогональное преобразование вида: $x'' = \alpha x'$ и $y'' = \beta y'$. Тогда, введя новые координаты x'' и y'', получим

$$Q(\mathbf{v}) = \frac{\lambda_1}{\alpha^2} x''^2 + \frac{\lambda_2}{\beta^2} y''^2.$$

Если $\lambda_1 \neq 0$, то можно выбрать $\alpha^2 = |\lambda_1|$. Тогда $\lambda_1/\alpha^2 = \pm 1$. Аналогичную процедуру проделаем и для λ_2. Мы приходим к следующему результату.

Пусть Q — любая квадратичная форма в пространстве $\mathbb{R}^2$. Мы можем найти такие координаты x'' и y'', для которых Q определяется одной из следующих формул:

$$Q(\mathbf{v}) = \begin{cases} x''^2 + y''^2, \\ x''^2, \\ 0, \\ -x''^2 \\ x''^2 - y''^2, \\ -x''^2 - y''^2. \end{cases}$$

Если обе координаты входят в формулу со знаком плюс, то $Q(\mathbf{v})$ имеет минимум при $\mathbf{v} = 0$ (рис. 4.9). Если они входят со знаком минус, то $Q(\mathbf{v})$ имеет в этой точке максимум. Если одна координата входит со знаком плюс, а другая со знаком минус, то $Q(\mathbf{v})$ не имеет ни максимума, ни минимума, а *седловую точку* в начале координат, как это показано на рис. 4.10.

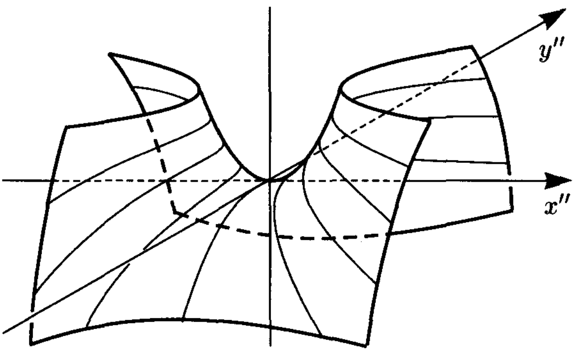


Рис. 4.10

4.4. Нормальные колебания

Теория связанных осцилляторов является одним из самых важных приложений результатов предыдущего параграфа. Чтобы это продемонстрировать, рассмотрим механическую систему, состоящую из двух незатухающих осцилляторов (рис. 4.11), которые мы соединим пружиной с константой упругости k (рис. 4.12). Уравнения движения, определяемые законами Ньютона, имеют вид

$$\begin{aligned} m_1\ddot{x}_1 &= -k_1x_1 - k(x_1 - x_2), \\ m_2\ddot{x}_2 &= -k_2x_2 - k(x_2 - x_1), \end{aligned}$$

или

$$T\begin{pmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{pmatrix} = -H\begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

где T и H — симметричные матрицы:

$$T = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}, \quad H = \begin{pmatrix} k_1 + k & -k \\ -k & k_2 + k \end{pmatrix}.$$

Рис. 4.11. Несвязанные осцилляторы

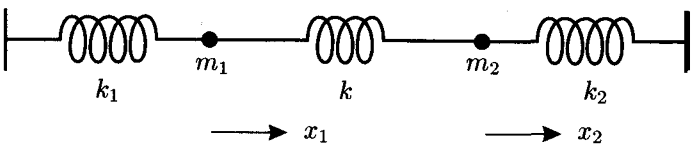


Рис. 4.12. Связанные осцилляторы

Наша задача — попытаться *одновременно* диагонализировать матрицы T и H, чтобы «развязать» уравнения.

Давайте обсудим общий случай. Мы рассматриваем две симметричные матрицы T и H, причем T положительно определена. Сначала покажем, что можно найти положительно определенную матрицу B, удовлетворяющую уравнению

$$T = B^2.$$

Действительно, если T диагональна, как в нашем примере, то можно взять

$$B = \begin{pmatrix} m_1^{1/2} & 0 \\ 0 & m_2^{1/2} \end{pmatrix}.$$

С другой стороны, можно найти такую матрицу вращения R_θ, что

$$T = R_\theta \Delta R_\theta{}^T, \quad \text{где} \quad \Delta \text{ — диагональная матрица.}$$

Пусть $\Delta = C^2$, где C положительно определена. Тогда матрица $B = R_\theta C R_\theta^T$ симметрична, положительно определена и удовлетворяет условию $B^2 = T$.

Теперь определим

$$\mathbf{w} = B\mathbf{v},$$

откуда

$$\mathbf{v} = B^{-1}\mathbf{w}.$$

Тогда $\ddot{\mathbf{v}} = B^{-1}\ddot{\mathbf{w}}$, и наше уравнение $T\ddot{\mathbf{v}} = -H\mathbf{v}$ принимает вид

$$TB^{-1}\ddot{\mathbf{w}} = -HB^{-1}\mathbf{w},$$

или

$$B\ddot{\mathbf{w}} = -HB^{-1}\mathbf{w},$$

так как $T = B^2$. Теперь получаем

$$\ddot{\mathbf{w}} = -A\mathbf{w}, \quad \text{где} \quad A = B^{-1}HB^{-1}.$$

Заметим, что A — тоже симметричная матрица. Таким образом, мы свели задачу к случаю, когда $T = \mathbb{I}$. (Проницательный читатель, должно быть, заметил, что с геометрической точки зрения мы просто перешли в систему координат, в которой *квадратичная функция*, связанная с T, приведена к нормальному виду $x^2 + y^2$.)

Чтобы решить уравнение $\ddot{\mathbf{w}} = -A\mathbf{w}$, найдем собственные значения и собственные векторы матрицы A. Предположим, что $\mathbf{v}_1$ — собственный вектор матрицы A с положительным собственным значением ω_1^2. Тогда очевидно, что функция

$$\mathbf{w}(t) = \rho \cos(\omega_1 t + \alpha)\mathbf{v}_1$$

является решением при произвольной амплитуде ρ и фазе α. Аналогично, второй собственный вектор и собственное значение дадут $\rho\cos(\omega_2 t + \alpha)$. Эти решения называются *нормальными колебаниями* осциллирующей системы.

Запишем матрицу A в виде

$$A = RDR^{-1},$$

где D — диагональная матрица. Тогда, определив вектор $\mathbf{u}$ равенством

$$\mathbf{w} = R\mathbf{u},$$

получаем уравнение

$$\ddot{\mathbf{w}} = R\ddot{\mathbf{u}} = -RDR^{-1}R\mathbf{u},$$

или

$$\ddot{\mathbf{u}} = -D\mathbf{u}.$$

Так как матрица D диагональна, для каждой компоненты вектора $\mathbf{u}$ получается независимое дифференциальное уравнение. Предположим, что оба собственных значения матрицы A положительны. Пусть они равны ω_1^2 и ω_2^2. Тогда общее решение уравнения

$$\ddot{\mathbf{u}} = -D\mathbf{u}$$

имеет вид

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \rho_1\cos(\omega_1 t + \alpha_1) \\ \rho_2\cos(\omega_2 t + \alpha_2) \end{pmatrix}.$$

Если собственные векторы матрицы A равны

$$\mathbf{v}_1 = R\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = R\begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

то общее решение уравнения $\ddot{\mathbf{w}} = -A\mathbf{w}$ можно записать в форме

$$\boxed{\mathbf{w}(t) = \rho_1\cos(\omega_1 t + \alpha_1)\mathbf{v}_1 + \rho_2\cos(\omega_2 t + \alpha_2)\mathbf{v}_2.}$$

Таким образом, общее решение является «суперпозицией» нормальных колебаний.

В качестве примера рассмотрим две *одинаковые* связанные между собой пружины, т.е. предположим, что $m_1 = m_2 = m$, $k_1 = k_2$. В отсутствие связи каждая пружина описывается уравнением

$$\ddot{x} = -\omega_0^2 x, \quad \omega_0 2 = k_1/m_1 = k_2/m_2.$$

При наличии связи получаем систему уравнений

$$\ddot{x}_1 = -(\omega_0^2 + s)x_1 + sx_2, \quad \ddot{x}_2 = sx_1 - (\omega_0^2 + s)x_2, \quad s = k/m$$

или

$$\ddot{\mathbf{v}} = -A\mathbf{v}, \quad A = \begin{pmatrix} \omega_0^2 + s & -s \\ -s & \omega_0^2 + s \end{pmatrix}.$$

Из соображений симметрии получаем, что собственные векторы матрицы A равны

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{для собственного значения } \omega_0^2$$

и

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{для собственного значения } \omega_0^2 + 2s.$$

Эти векторы задают два вида нормальных колебаний нашей системы. В первом случае шарики движутся в одну сторону, а во втором — в противоположных направлениях (рис. 4.13).

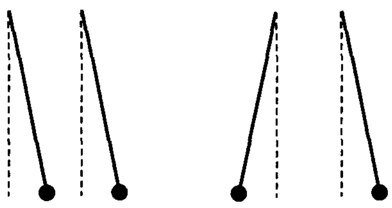


Рис. 4.13

$$(\omega_0^2 + 2s)^{1/2} \sim \omega_0 + s/\omega_0, \quad \text{если } s/\omega_0 \quad \text{мало}.$$

Общее решение нашей системы дифференциальных уравнений имеет вид

$$x_1 = \rho_1 \cos(\omega_0 t + \alpha_1) + \rho_2 \cos(\omega' t + \alpha_2),$$
$$x_2 = \rho_1 \cos(\omega_0 t + \alpha_1) - \rho_2 \cos(\omega' t + \alpha_2).$$

Изучим частное решение нашей системы, когда мы натянули одну пружину и отпустили ее в $t = 0$. Иными словами, зададим начальные условия

$$\begin{aligned} x_1(0) = C, &\quad \dot{x}_1(0) = 0, \\ x_2(0) = 0, &\quad \dot{x}_2(0) = 0. \end{aligned}$$

Из общего решения для данных начальных условий получаем $\rho_1 = \rho_2 = C/2$ и $\alpha_1 = \alpha_2 = 0$. Таким образом, частное решение имеет вид

$$\begin{aligned} x_1 &= \frac{1}{2}C(\cos\omega_0 t + \cos\omega' t), \\ x_2 &= \frac{1}{2}C(\cos\omega_0 t - \cos\omega' t). \end{aligned}$$

Вспомним, что $\cos\alpha = \frac{1}{2}(e^{i\alpha} + e^{-i\alpha})$. Отсюда можно получить известные формулы:

$$\cos\alpha + \cos\beta = 2\cos\left(\frac{\alpha-\beta}{2}\right)\cos\left(\frac{\alpha+\beta}{2}\right)$$

и

$$\cos\alpha - \cos\beta = -2\sin\left(\frac{\alpha-\beta}{2}\right)\sin\left(\frac{\alpha+\beta}{2}\right).$$

Подставив $\alpha = \omega_0 t$ и $\beta = \omega' t$, получаем наше частное решение в форме

$$\begin{aligned} x_1 &= C\cos(\omega' - \omega_0)t\cos\omega_0 t, \\ x_2 &= -C\sin(\omega' - \omega_0)t\sin\omega_0 t. \end{aligned}$$

В случае слабой связи, т. е. при малой константе s, величина $\omega' - \omega_0$ мала. На рис. 4.14 изображены графики движения обеих пружин как функции от времени. Колебания каждой пружины (с собственной частотой ω_0) модулируются. *Биения* определяются модулирующими множителями $\cos(\omega' - \omega_0)t$, $\sin(\omega' - \omega_0)t$. Энергия переходит от одной пружины к другой. Когда одна пружина колеблется с максимальной амплитудой, вторая отдыхает, и наоборот. Это явление называется *резонансом.*

Если пружины не одинаковы, но отличие невелико («неточная настройка»), то поведение системы будет почти таким же.

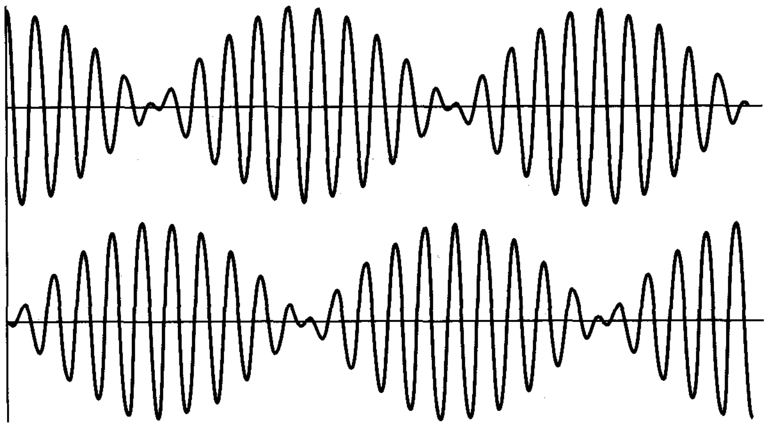


Рис. 4.14

По-прежнему колебания каждой пружины модулируются. Вторая пружина отдыхает через определенные промежутки времени, а первая осциллирует даже в то время, когда вторая осциллирует с максимальной амплитудой. Неточность «настройки» приводит к неполной передаче энергии от первой пружины ко второй. Мы не будем говорить об этом более подробно. Детали исследования здесь просты, хотя и несколько утомительны. Вычисление собственных векторов и собственных значений матрицы A предоставим читателю в качестве упражнения.

4.5. Нормальные моды в многомерном пространстве

Пусть V — n-мерное векторное пространство с положительно определенным скалярным произведением. Кроме того, имеется другое, не обязательно положительно определенное скалярное произведение $\langle\cdot,\cdot\rangle$. Аргументы, изложенные в параграфе 4.3, показывают, что существует линейное преобразование $A : V \to V$, при котором

$$\langle\mathbf{u},\mathbf{v}\rangle = (A\mathbf{u},\mathbf{v}) \quad \text{для всех } \mathbf{u},\mathbf{v} \quad \text{в пространстве } V,$$

где A симметрично, т. е.

$$(A\mathbf{u}, \mathbf{v}) = (\mathbf{u}, A\mathbf{v}).$$

Действительно, из параграфа 4.2 мы знаем, как найти изоморфизм V пространства $\mathbb{R}^n$, чтобы $(\cdot,\cdot)$ преобразовалось в евклидово скалярное произведение. Тогда без изменения можно воспользоваться аргументами параграфа 4.3 и показать, что A — симметричная матрица.

С помощью аргументов параграфа 4.3 можно показать, что матрица A имеет n взаимно перпендикулярных собственных векторов. Действительно, рассмотрим квадратичную форму

$$Q(\mathbf{v}) = \langle \mathbf{v}, \mathbf{v} \rangle = (A\mathbf{v}, \mathbf{v})$$

на *единичной сфере*

$$\{\mathbf{v} | \, \|\mathbf{v}\| = 1\}.$$

Эта функция непрерывна и ограничена. Если для всех матричных элементов A_{ij} выполняется неравенство $|A_{ij}| \le M$, где M — некоторое число, то для

$$\mathbf{v} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

мы имеем $\|\mathbf{v}\|^2 = \sum x_i^2 = 1$, следовательно, $|x_i| \le 1$ при всех значениях i и

$$(A\mathbf{v}, \mathbf{v}) = \sum A_{ij} x_i x_j.$$

Откуда следует

$$|(A\mathbf{v}, \mathbf{v})| \le \sum |A_{ij}| \le nM.$$

Пусть $\mathbf{v}$ определяет точку на поверхности единичной сферы, где $Q(\mathbf{v})$ принимает максимальное значение. (В этом месте мы реально используем свойства системы вещественных чисел, которые гарантируют, что на самом деле существует точка на сфере, где Q принимает максимальное значение.) Утверждается, что

$\mathbf{v}$ является собственным вектором матрицы A. Действительно, определим вектор $\mathbf{w}$ формулой

$$\mathbf{w} = A\mathbf{v} - (A\mathbf{v}, \mathbf{v})\mathbf{v}.$$

Покажем, что если в точке $\mathbf{v}$ квадратичная форма Q принимает максимальное значение, то $\mathbf{w} = \mathbf{0}$. Так как $(\mathbf{v}, \mathbf{v}) = 1$, то вектор $\mathbf{w}$ перпендикулярен вектору $\mathbf{v}$, т. е.

$$(\mathbf{w}, \mathbf{v}) = 0,$$

и, следовательно,

$$(A\mathbf{v}, \mathbf{w}) = \|\mathbf{w}\|^2.$$

Тогда для любого вещественного числа s

$$\|\mathbf{v} + s\mathbf{w}\|^2 = (\mathbf{v} + s\mathbf{w}, \mathbf{v} + s\mathbf{w}) = \|\mathbf{v}\|^2 + s^2\|\mathbf{w}\|^2 = 1 + s^2\|\mathbf{w}\|^2$$

и

$$(A(\mathbf{v} + s\mathbf{w}), \mathbf{v} + s\mathbf{w}) = (A\mathbf{v}, \mathbf{v}) + s(A\mathbf{w}, \mathbf{v}) + s(A\mathbf{v}, \mathbf{w}) + s^2(A\mathbf{w}, \mathbf{w}).$$

Однако, $(A\mathbf{v}, \mathbf{w}) = (\mathbf{w}, A\mathbf{v})$, следовательно,

$$\begin{aligned}(A(\mathbf{v} + s\mathbf{w}), \mathbf{v} + s\mathbf{w}) &= (A\mathbf{v}, \mathbf{v}) + 2s(A\mathbf{v}, \mathbf{w}) + s^2(A\mathbf{w}, \mathbf{w}) \\ &= (A\mathbf{v}, \mathbf{v}) + 2s\|\mathbf{w}\|^2 + s^2(A\mathbf{w}, \mathbf{w}).\end{aligned}$$

Домножим вектор $\mathbf{v} + s\mathbf{w}$ на число, чтобы получить вектор единичной длины:

$$\mathbf{u} = \frac{1}{\|\mathbf{v} + s\mathbf{w}\|}(\mathbf{v} + s\mathbf{w}).$$

Возьмем функцию

$$\begin{aligned}f(s) \stackrel{\text{def}}{=} (A\mathbf{u}, \mathbf{u}) &= \frac{1}{\|\mathbf{v} + s\mathbf{w}\|^2}(A(\mathbf{v} + s\mathbf{w}), \mathbf{v} + s\mathbf{w}) \\ &= \frac{1}{1 + s^2\|\mathbf{w}\|^2}((A\mathbf{v}, \mathbf{v}) + 2s\|\mathbf{w}\|^2 + s^2(A\mathbf{w}, \mathbf{w})).\end{aligned}$$

Это выражение представляет дифференцируемую функцию переменной s. По нашему предположению она имеет максимум в точке $s = 0$, поэтому первая производная $f'(0)$ при $s = 0$ должна обращаться в нуль: $f'(0) = 2\|\mathbf{w}\|^2$. Отсюда следует, что $\mathbf{w} = \mathbf{0}$. Таким образом, мы доказали, что

$$\boxed{A\mathbf{v} = (A\mathbf{v}, \mathbf{v})\mathbf{v}.}$$

Другими словами, $\mathbf{v}$ является собственным вектором A для собственного значения $(A\mathbf{v}, \mathbf{v})$. Обозначим этот вектор $\mathbf{v}_1$, а соответствующее ему собственное значение $(A\mathbf{v}_1, \mathbf{v}_1) = \lambda_1$.

Теперь рассмотрим пространство всех векторов $\mathbf{z}$ в V, которые перпендикулярны вектору $\mathbf{v}_1$ (рис. 4.15), т. е. для всех векторов $\mathbf{z}$ должно выполняться условие

$$(\mathbf{z}, \mathbf{v}_1) = 0.$$

Для таких $\mathbf{z}$ мы имеем

$$(A\mathbf{z}, \mathbf{v}_1) = (\mathbf{z}, A\mathbf{v}_1) = \lambda_1(\mathbf{z}, \mathbf{v}_1) = 0, \quad \lambda_1 = (A\mathbf{v}_1, \mathbf{v}_1).$$

Рассмотрим множество всех $\mathbf{z}$ единичной длины, т. е. таких $\mathbf{z}$, для которых справедливы равенства

$$\|\mathbf{z}\| = 1, \quad (\mathbf{z}, \mathbf{v}_1) = 0.$$

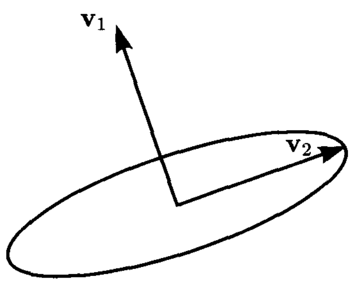


Рис. 4.15

Пусть $\mathbf{v}_2$ определяет точку, где Q принимает максимальное значение. Запишем

$$A\mathbf{v}_2 = (A\mathbf{v}_2, \mathbf{v}_2)\mathbf{v}_2 + \mathbf{w}_2.$$

Из того, что $(\mathbf{v}_2, \mathbf{v}_1) = 0$ и $(A\mathbf{v}_2, \mathbf{v}_1) = 0$, получаем $(\mathbf{w}_2, \mathbf{v}_1) = 0$. Как и раньше, мы делаем вывод, что $(\mathbf{w}_2, \mathbf{v}_2) = 0$. Тогда

$$\begin{aligned} &\frac{1}{\|\mathbf{v}_2 + s\mathbf{w}_2\|^2}(A(\mathbf{v}_2 + s\mathbf{w}_2), (\mathbf{v}_2 + s\mathbf{w}_2)) \\ &\qquad = \frac{1}{1 + s^2 + \|\mathbf{w}_2\|^2}((A\mathbf{v}_2, \mathbf{v}_2) + 2s\|\mathbf{w}_2\|^2 + s^2(A\mathbf{w}_2, \mathbf{w}_2)) \end{aligned}$$

имеет максимум при $s = 0$ и, следовательно, $\mathbf{w}_2 = \mathbf{0}$, а вектор $\mathbf{v}_2$ является собственным вектором матрицы A.

Мы можем продолжить наши рассуждения: рассмотрим векторы $\mathbf{z}$, удовлетворяющие условию $(\mathbf{z}, \mathbf{v}_1) = (\mathbf{z}, \mathbf{v}_2) = 0$ и $\|\mathbf{z}\| = 1$, и так далее. Каждый раз мы будем находить новый собственный вектор матрицы A, перпендикулярный ко всем предыдущим. Когда этот процесс закончится? Когда мы выберем все ненулевые векторы, перпендикулярные $\mathbf{v}_1, \ldots, \mathbf{v}_k$. Это произойдет только при условии $k = n$. Действительно, k не может быть больше n, так как тогда векторы $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}$ будут взаимно перпендикулярны и, следовательно, линейно независимы. Но это противоречит исходному предположению, что в пространстве V нет $n+1$ линейно независимых векторов (мы предположили, что пространство V n-мерно). С другой стороны, если $k < n$, то уравнения

$$\begin{gathered} (\mathbf{v}_1, \mathbf{w}) = 0, \\ \cdots \\ (\mathbf{v}_k, \mathbf{w}) = 0 \end{gathered}$$

в пространстве $\mathbb{R}^n$ образуют систему k однородных линейных уравнений с n неизвестными. Такие системы всегда имеют решение. В общем случае мы докажем это в главе 10. Здесь мы докажем существование вектора $\mathbf{w} \neq \mathbf{0}$. Если n-компонентный вектор $\mathbf{v}_k$ не равен нулю, т. е.

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \text{где} \quad x_n \neq 0,$$

то последнее уравнение системы имеет вид

$$x_1\mathbf{w}_1 + \cdots + x_n\mathbf{w}_n = \mathbf{0}.$$

Из этого уравнения вектор $\mathbf{w}_n$ выразим через $\mathbf{w}_1, \ldots, \mathbf{w}_{n-1}$:

$$\mathbf{w}_n = -\frac{1}{x_n}(x_1\mathbf{w}_1 + \cdots + x_{n-1}\mathbf{w}_{n-1}).$$

Подстановка этого вектора в предыдущее уравнение дает систему $k-1$ уравнений с $n-1$ неизвестными. Процедура может быть продолжена. Если n-я компонента *любого* из векторов $\mathbf{v}_1, \ldots, \mathbf{v}_k$ не равна нулю, мы можем проделать то же самое — взять вектор $\mathbf{v}_j$ с ненулевой n-й компонентой и получить $\mathbf{w}_n$. Если n-я компонента *всех* векторов $\mathbf{v}_1, \ldots, \mathbf{v}_n$ равна нулю, то решением является вектор

$$\mathbf{w} = \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix}$$

(все первые $n-1$ компонент обращаются в 0).

Таким образом, продолжая этот алгоритм, мы получим n попарно ортогональных собственных векторов.

Нормальные колебания как волны

Поработаем с интересным примером в пространстве n измерений. Представим себе последовательность равных точечных масс, где каждая соединена с ближайшими соседками пружинками, причем все пружинки одинаковые. Тогда сила, действующая на i-ю точечную массу, равна

$$-k(x_i - x_{i+1}) + k(x_i - x_{i-1}).$$

Запишем уравнения Ньютона

$$m\ddot{x}_i = -k(2x_i - x_{i-1} - x_{i+1}).$$

Предположим, что первая и последняя точки также соединены пружинкой. Система изображена на рис. 4.16.

Введем частоту $\omega^2 = k/m$ и перепишем уравнения движения в форме

$$\ddot{x} = -\omega^2 Ax,$$

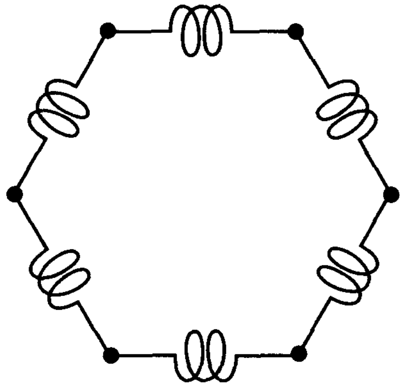


Рис. 4.16

где

$$A = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & -1 \\ -1 & 2 & -1 & \cdots & & \\ 0 & -1 & 2 & \cdots & & \\ \vdots & \vdots & \vdots & & & -1 \\ -1 & 0 & & & -1 & 2 \end{pmatrix}.$$

Наша задача — найти собственные значения и собственные векторы матрицы A. Прежде чем рассматривать общий случай, начнем с рассмотрения случая $n = 3$. Из равенства

$$\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

следует, что $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ — собственный вектор этой матрицы, соответствующий собственному значению 0. Мы знаем, что другие собственные векторы должны быть ортогональны $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. Попробуем вектор $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$. Тогда

$$\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ -3 \\ 0 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}.$$

Значит, вектор $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ и, аналогично, третий вектор $\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$ являются собственными векторами с собственным значенем 3. Таким образом, собственные значения равны 0 и 3, причем значение 3 реализуется дважды.

Перейдем к случаю $n = 4$:

$$A = \begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{pmatrix}.$$

Тогда

$$A \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

как и раньше. Кроме того,

$$A \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} = 4 \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}.$$

Остающиеся собственные векторы должны быть ортогональны этим двум. Попробуем вектор

$$\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}.$$

Получаем

$$\begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}.$$

Находим последний собственный вектор

$$\begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} = 2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}.$$

Итак, мы получили четыре собственных значения: 0, 4 и два раза появилось собственное значение 2.

Чтобы рассмотреть n-мерный случай, заметим, что наша система инвариантна относительно поворота, когда первая точка становится на место n-й, вторая на место первой, и т. д. Такой поворот осуществляется матрицей

$$S = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & \\ \vdots & & & & & \\ 0 & \cdots & \cdots & \cdots & \cdots & 1 \\ 1 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}.$$

Легко проверить, что $SA = AS$. Найдем собственные векторы матрицы S. Если $S\mathbf{w} = \lambda\mathbf{w}$, то $SA\mathbf{w} = AS\mathbf{w} = A(\lambda\mathbf{w}) = \lambda A\mathbf{w}$. Поэтому, если $\mathbf{w}$ является собственным вектором матрицы S для собственного значения λ, то $A\mathbf{w}$ — тоже собственный вектор с тем же собственным значением. Ниже мы получим n различных собственных значений матрицы S. Тогда если $S\mathbf{w} = \lambda\mathbf{w}$, то вектор $A\mathbf{w}$ должен быть пропорционален $\mathbf{w}$, т. е. быть собственным вектором матрицы A.

Окажется, что «собственные значения» матрицы S и координаты «собственных векторов» будут комплексными числами. Вещественные и мнимые части этих векторов будут собственными векторами матрицы A. Приведем это доказательство.

Пусть $\tau = e^{2\pi \mathrm{i}/n}$, откуда $\tau^n = 1$. Тогда

$$S\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} = \begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix},$$

$$S\begin{pmatrix}1\\ \tau\\ \tau^2\\ \vdots\\ \tau^{n-1}\end{pmatrix} = \begin{pmatrix}\tau\\ \tau^2\\ \\ \vdots\\ 1\end{pmatrix} = \begin{pmatrix}\tau\\ \tau^2\\ \\ \vdots\\ \tau^n\end{pmatrix} = \tau\begin{pmatrix}1\\ \tau\\ \tau^2\\ \vdots\\ \tau^{n-1}\end{pmatrix},$$

$$S\begin{pmatrix}1\\ \tau^2\\ \tau^4\\ \tau^6\\ \vdots\\ \tau^{2(n-1)}\end{pmatrix} = \tau^2\begin{pmatrix}1\\ \tau^2\\ \tau^4\\ \tau^6\\ \vdots\\ \tau^{2(n-1)}\end{pmatrix}$$

и так далее. Все собственные значения $1,\ \tau,\ \tau^2,\ \ldots,\ \tau^{n-1}$ различны. Таким образом, каждый собственный вектор матрицы S должен быть собственным вектором матрицы A. Обозначим эти «собственные векторы» $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Мы знаем, что

$$A\mathbf{e}_k = \lambda_k \mathbf{e}_k$$

для некоторых λ_k, которые нам предстоит вычислить. Вторая координата вектора $A\mathbf{e}_k$ равна

$$\begin{pmatrix}-1 & 2 & -1 & 0 & \ldots & 0\end{pmatrix}\begin{pmatrix}1\\ \tau^k\\ \tau^{2k}\\ \vdots\end{pmatrix}$$

$$= -1 + 2\tau^k - \tau^{2k} = (-\tau^{-k} + 2 - \tau^k)\tau^k = 2(1 - \cos(2\pi k/n))\tau^k.$$

Отсюда следует, что k-е собственное значение равно

$$\lambda_k = 2(1 - \cos(2\pi k/n)).$$

Мы получаем одинаковые собственные значения для k и $n-k$. Тогда, складывая и вычитая собственные векторы для k и $n-k$, мы можем получить вещественные собственные векторы. Итак, векторы

$$\begin{pmatrix} 1 \\ \cos(2\pi k/n) \\ \cos(4\pi k/n) \\ \cos(6\pi k/n) \\ \vdots \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} 0 \\ \sin(2\pi k/n) \\ \sin(4\pi k/n) \\ \sin(6\pi k/n) \\ \vdots \end{pmatrix}$$

являются собственными векторами для собственного значения

$$2(1+\cos(2\pi k/n)).$$

Если $n = 2m$ — четное число, то второй столбец равен нулю при $k = m$. Во всех остальных случаях собственные векторы не равны нулю. Таким образом, каждую нормальную моду системы можно рассматривать как волну сжатия (синус или косинус).

4.6. Специальная теория относительности

В этом параграфе мы подробно рассмотрим геометрию двумерного векторного пространства с квадратичной формой $Q(\mathbf{v}) = \langle \mathbf{v}, \mathbf{v} \rangle$, которая может принимать как положительные, так и отрицательные значения. Как мы уже знаем, это пространство можно идентифицировать с $\mathbb{R}^2$, а квадратичную форму Q определить соотношением

$$Q\begin{pmatrix} t \\ x \end{pmatrix} = t^2 - x^2.$$

Мы увидим, что геометрия этого пространства (рис. 4.17) обеспечивает хорошую модель для понимания специальной теории относительности. Слово *модель* мы применяем в следующем смысле. Наше обычное пространство трехмерно. Поэтому, если добавить *время* в качестве *дополнительного измерения*, мы получим четырехмерное *пространство-время*. В нашей модели мы будем считать пространство одномерным, поэтому наше пространство-время вместо четырехмерного будет двумерным.

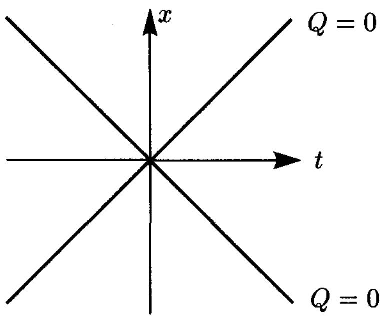


Рис. 4.17

В этом случае мы сможем нарисовать все геометрические конструкции. На самом деле, большая часть того, что будет сказано, работает в обычном четырехмерном мире, с некоторой модификацией нашей двумерной модели.

Первый постулат специальной теории относительности подтверждает закон Ньютона: частицы, на которые не действуют никакие силы, движутся вдоль прямых линий. Таким образом, геометрия нашего пространства-времени среди всех возможных кривых выделяет прямые линии. Наше пространство-время является аффинной плоскостью с некоторыми дополнительными геометрическими структурами.

Второй постулат — конечная абсолютная константа. Таким образом, в каждой точке пространства-времени существуют две определенные линии, изображающие распространение света влево или вправо. Пространственная и временна́я инвариантность скорости света означает, что при смещении точки P в точку Q два световых луча, проходящие через точку P, преобразуются в два световых луча, проходящих через точку Q (рис. 4.18).

Мы будем исследовать такие аффинные преобразования, которые световые лучи переводят в световые лучи. Трансляции попадают в это категорию. Поэтому задача сводится к изучению линейных преобразований, которые сохраняют световые лучи, проходящие через начало координат. Итак, у нас есть две прямые $x = \pm ct$. Требуется найти такие линейные преобразования, которые сохраняют эти прямые. Для дальнейших вычислений удобно ввести *естественные единицы* длины и времени, для которых

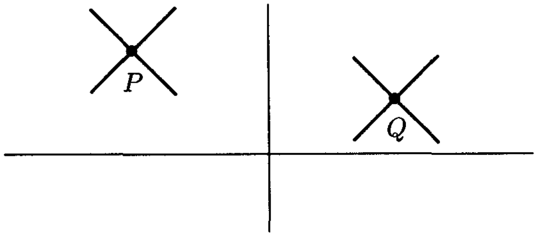


Рис. 4.18

скорость света равна единице. Например, мы могли бы измерять t в годах, а x в световых годах. Или если мы выберем наносекунду (10^{-9} секунд) в качестве единицы времени, то соответствующая единица длины, с великолепной точностью, будет один фут. Поэтому в качестве естественных единиц t можно измерять в наносекундах, а x в футах.

Итак, мы будем изучать такие линейные преобразования, которые сохраняют изображенную на рис. 4.19 фигуру, определяемую линиями $x = t$ и $x = -t$.

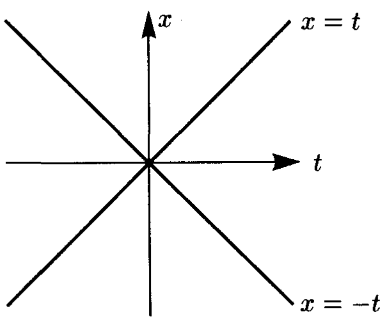


Рис. 4.19

Получив эти линейные преобразования, мы *определим все преобразования пространства-времени, которые сохраняют прямые линии и скорость света.*

На самом деле, мы хотели бы, по крайней мере временно, исключить из рассмотрения определенные типы преобразований.

Например, преобразования отражения

$$\begin{pmatrix} t \\ x \end{pmatrix} \to \begin{pmatrix} -t \\ x \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} t \\ x \end{pmatrix} \to \begin{pmatrix} t \\ -x \end{pmatrix}$$

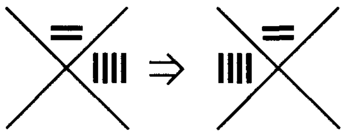


Рис. 4.20

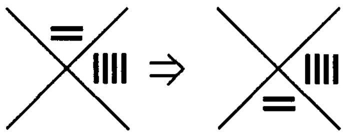


Рис. 4.21

и преобразование инверсии (рис. 4.22).

$$\begin{pmatrix} t \\ x \end{pmatrix} \to \begin{pmatrix} -t \\ -x \end{pmatrix}$$

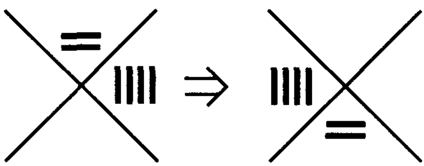


Рис. 4.22

Эти преобразования все прямые $x = \pm t$ переводят в $x = \pm t$, при этом возможна перестановка линий. На рисунках показано, как эти преобразования переставляют четыре области на плоскости. Поэтому, умножая на одно из них, можно сделать так, чтобы необходимые нам преобразования сохраняли бы каждую из этих четырех областей. Таким образом, мы будем искать преобразования F, которые сохраняют прямые $x = t$, $x = -t$ и *переднюю область* (область абсолютного будущего): $t^2 > x^2$, $t > 0$.

Чтобы получить такие преобразования, перейдем к новым координатам, для которых эти прямые становятся осями координат (рис. 4.23):

$$\begin{pmatrix} p \\ q \end{pmatrix} = R^{-1} \begin{pmatrix} t \\ x \end{pmatrix}, \quad R = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad R^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

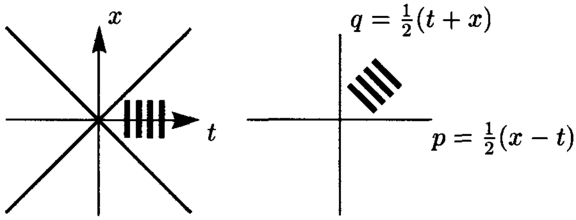


Рис. 4.23

Таким образом, $R^{-1}FR$ должно сохранять координатные оси и положительный квадрант, и преобразование $R^{-1}FR$ является диагональной матрицей

$$R^{-1}FR = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}, \quad a > 0, \quad d > 0.$$

Введем обозначения: $ad = s^2$ и $a/d = r^2$ выбрав $s > 0$, $r > 0$. Тогда

$$R^{-1}FR = \begin{pmatrix} s & 0 \\ 0 & s \end{pmatrix} \begin{pmatrix} r & 0 \\ 0 & r^{-1} \end{pmatrix}.$$

Значит, мы доказали, что

$$F = SL_r,$$

где

$$S = \begin{pmatrix} s & 0 \\ 0 & s \end{pmatrix}$$

является преобразованием смены масштаба и

$$L_r = R \begin{pmatrix} r & 0 \\ 0 & r^{-1} \end{pmatrix} R^{-1} = \frac{1}{2} \begin{pmatrix} r + r^{-1} & r - r^{-1} \\ r - r^{-1} & r + r^{-1} \end{pmatrix}. \tag{4.2}$$

Преобразование L_r называется *собственным преобразованием Лоренца* с параметром r. Мы утверждаем, что для любой пары векторов $\mathbf{v}_1$, $\mathbf{v}_2$

$$\boxed{\langle L_r\mathbf{v}_1, L_r\mathbf{v}_2\rangle = \langle \mathbf{v}_1, \mathbf{v}_2\rangle.} \tag{4.3}$$

Действительно, по аналогии с формулой (4.1) для скалярного произведения $\langle\cdot,\cdot\rangle$, достаточно доказать, что

$$Q(\mathbf{v}) = Q(L_r\mathbf{v}).$$

В нашем случае

$$Q(\mathbf{v}) = t^2 - x^2 = -4pq$$

и, если $\mathbf{v}' = L_r\mathbf{v} = \begin{pmatrix} t' \\ x' \end{pmatrix}$, то

$$\begin{pmatrix} p' \\ q' \end{pmatrix} = \begin{pmatrix} rp \\ r^{-1}q \end{pmatrix} \quad \text{и} \quad Q(\mathbf{v}') = -4p'q' = -4pq = Q(\mathbf{v}),$$

что и требовалось доказать[2].

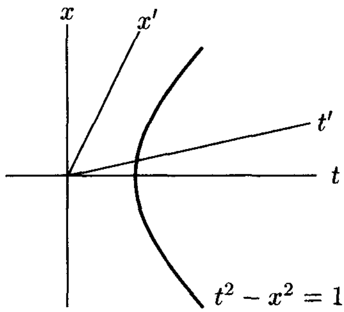


Рис. 4.24. $\begin{pmatrix} t' \\ x' \end{pmatrix} = L_r \begin{pmatrix} t \\ x \end{pmatrix}$

Результат преобразования масштаба S состоит в том, все длины и время умножаются на множитель s. Существование атомных часов вместе с определенными спектральными линиями показывает, что природные явления не инвариантны относительно преобразования S при $s \neq 1$.

Линейное преобразование A, которое сохраняет квадратичную форму Q в том смысле, что

$$Q(A\mathbf{v}) = Q(\mathbf{v}) \quad \text{для всех } \mathbf{v} \quad \text{в пространстве } V,$$

называется *преобразованием Лоренца*. Такое преобразование A должно переводить *световой конус* (называемый также *нулевым конусом*)

$$\{\mathbf{v} \mid Q(\mathbf{v}) = 0\}$$

[2]Инвариантность $Q(\mathbf{v})$ при преобразовании Лоренца проиллюстрирована на рис. 4.24. — *Прим. ред.*

в себя, т. е. множество $\{x = \pm t\}$ должно сохраняться. Если, кроме того, A преобразует область абсолютного будущего в себя же, то оно является собственным преобразованием Лоренца $A = L_r$ для некоторого числа r.

Производя ряд преобразований Лоренца, собственные преобразования Лоренца могут быть непрерывно сведены к тождественному. В этом состоит их отличие от других преобразований Лоренца. Действительно, пусть $A(t)$ есть семейство преобразований Лоренца, удовлетворяющее условиям $A(0) = \mathbb{I}$, $A(1) = A$. Пусть $\mathbf{v}$ указывет точку в области абсолютного будущего. Тогда $A(t)\mathbf{v}$ не может пересекать световой конус, потому что $Q(A(t)\mathbf{v}) = Q(\mathbf{v}) > 0$. Для любого преобразования Лоренца $\operatorname{Det} A = \pm 1$, т. к. A, умножаемое на матрицу вида $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ или $\begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix}$, есть собственное преобразование Лоренца и $\operatorname{Det} L = 1$ для собственного преобразования Лоренца L. Таким образом, поскольку $\operatorname{Det} A(t)$ непрерывно изменяется со временем t и $\operatorname{Det} A(0) = 1$, мы должны иметь $\operatorname{Det} A(t) \equiv 1$, следовательно, $\operatorname{Det} A = 1$. Но если A может быть непрерывным образом сведено к тождественному преобразованию, значит, A является собственным преобразованием Лоренца. С другой стороны, если $A = L_r$, то положим $A(t) = L_{tr}$, так что $A(0) = \mathbb{I}$ и $A(1) = A$.[3]

[3]Собственные преобразования среди всех преобразований Лоренца можно охарактеризовать возможностью перевести их в тождественное отображение непрерывным изменением в классе всех преобразований Лоренца. Точнее, для того, чтобы A было собственным преобразованием Лоренца, необходимо и достаточно существование семейства $A(\tau)$ преобразований Лоренца, непрерывно зависящего от τ при $0 \le \tau \le 1$, такого, что $A(0) = \mathbb{I}$, $A(1) = A$.

Докажем это. Необходимость доказывается совсем просто: если $A = L_r$, то в качестве $A(\tau)$ можно взять $A(\tau) = L_{r^\tau}$. Докажем достаточность. Пусть $\mathbf{v}$ — произвольная точка из передней области. Поскольку при всех τ $Q(A(\tau)\mathbf{v}) = Q(\mathbf{v}) > 0$, то точка $A(\tau)\mathbf{v}$ не может пересечь световой конус, т. е. и при $\tau = 1$ находится внутри передней области. Далее, детерминант любого преобразования Лоренца равен $+1$ или -1, поскольку такое преобразование может быть получено из собственного домножением на матрицу $\begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix}$. Но поскольку $\operatorname{Det} A(0) = 1$, то из соображений непрерывности $\operatorname{Det} A(\tau) = 1$, откуда $\operatorname{Det} A = 1$. Из этих фактов следует, что A — собственное преобразование Лоренца. — *Прим. ред.*

Произведение двух собственных преобразований Лоренца тоже является собственным преобразованием Лоренца. Действительно, если

$$L_r = R\begin{pmatrix} r & 0 \\ 0 & r^{-1} \end{pmatrix} R^{-1} \quad \text{и} \quad L_{r'} = R\begin{pmatrix} r' & 0 \\ 0 & r'^{-1} \end{pmatrix} R^{-1},$$

то

$$L_r L_r' = R\begin{pmatrix} r & 0 \\ 0 & r^{-1} \end{pmatrix} R^{-1} R\begin{pmatrix} r' & 0 \\ 0 & r'^{-1} \end{pmatrix} R^{-1},$$

так что

$$\boxed{L_r L_r' = L_{rr'}.} \tag{4.4}$$

Удобно обозначить $r = e^\alpha$ и положить

$$L^\alpha = L_{e^\alpha} = \frac{1}{2}\begin{pmatrix} e^\alpha + e^{-\alpha} & e^\alpha - e^{-\alpha} \\ e^\alpha - e^{-\alpha} & e^\alpha + e^{-\alpha} \end{pmatrix}.$$

Тогда

$$\boxed{L^\alpha \cdot L^{\alpha'} = L^{\alpha+\alpha'}.} \tag{4.5}$$

Если ввести в рассмотрение гиперболические функции $\cosh\alpha = \dfrac{1}{2}(e^\alpha + e^{-\alpha})$, $\sinh\alpha = \dfrac{1}{2}(e^\alpha - e^{-\alpha})$, то можно написать

$$L^\alpha = \begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix}.$$

Теперь по своей записи преобразование Лоренца L^α очень похоже на преобразование вращения R_θ:

$$L^\alpha = \begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix}, \quad R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Похожие формулы получаются для произведений преобразований:

$$L^{\alpha_1} \cdot L^{\alpha_2} = L^{\alpha_1+\alpha_2}, \quad \text{в то время как} \quad R_{\theta_1} \cdot R_{\theta_2} = R_{\theta_1+\theta_2}.$$

При изменении параметра α точка $L^\alpha \mathbf{v}$ движется вдоль гиперболы, за исключением пограничного случая, когда $\mathbf{v}$ лежит на световом конусе. В этом случае точка $L^\alpha \mathbf{v}$ движется к вершине светового конуса или от нее (кроме $\mathbf{v} = \mathbf{0}$, когда $L^\alpha \mathbf{v} \equiv \mathbf{0}$ для всех α). Именно поэтому функции $\sinh, \cosh$ называются *гиперболическим синусом* и *гиперболическим косинусом*. Если изменяется параметр θ, то точка $R_\theta \mathbf{v}$ движется по окружности, кроме случая $\mathbf{v} = \mathbf{0}$, когда она остается неподвижной. Поэтому функции cos, sin называются *круговыми*.

Евклидово движение плоскости — это преобразование вида $\begin{pmatrix} x \\ y \end{pmatrix} \to R\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}$, где R — *ортогональное преобразование*; другими словами, евклидово преобразование состоит из трансляции и ортогонального преобразования. Евклидова геометрия изучает свойства подмножеств плоскости, которые инвариантны при всех евклидовых преобразованиях. *Преобразование Пуанкаре* на плоскости есть преобразование вида $\begin{pmatrix} x \\ y \end{pmatrix} \to L\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}$, где L — преобразование Лоренца. Геометрия специальной теории относительности изучает свойства, инвариантные относительно всех преобразований Пуанкаре. Свойство прямой l быть параллельной оси y не является евклидовым свойством прямой l. Действительно, если l параллельна оси y, то Rl не будет параллельна y для всех вращений R, кроме поворотов на $0°$ и $180°$. Аналогично, свойство прямой быть параллельной оси x не является свойством, которое определяет специальная теория относительности. Если l параллельна x, то Ll не будет параллельной оси x для всех преобразований L, кроме тождественного. Это утверждение обычно формулируется так: «в рамках специальной теории относительности понятие одновременности не имеет смысла».

Точно так же не имеет смысла утверждение, что «частица покоится». Нам хотелось бы сказать, что линия $x = 0$ (ось t) представляют частицу, покоящуюся в начале координат. Но преобразование Лоренца L_r переводит эту линию в другую, проходящую через начало координат, и

$$L_r \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} r + r^{-1} \\ r - r^{-1} \end{pmatrix}.$$

Таким образом, преобразование L_r переводит ось t в прямую

$$x = vt, \quad \text{где} \quad v = \frac{r - r^{-1}}{r + r^{-1}} = \frac{r^2 - 1}{r^2 + 1}.$$

Результат похож на траекторию частицы, движущейся с постоянной скоростью v. Мы можем решить уравнение

$$v = (r^2 - 1)/(r^2 + 1),$$

выразив r через v:

$$r = \sqrt{(1 + v)/(1 - v)},$$

что легко проверяется. При желании мы можем использовать v в качестве параметра для описания L: определим

$$L(v) = L_r = L_{e^\alpha},$$

где

$$r = \sqrt{(1 + v)/(1 - v)} = e^\alpha,$$

$$v = \frac{r - r^{-1}}{r + r^{-1}} = \frac{e^\alpha - e^{-\alpha}}{e^\alpha + e^{-\alpha}} = \frac{\sinh \alpha}{\cosh \alpha} = \tanh \alpha.$$

Заметим, что

$$L(v)L(v') = L_{rr'}, \quad r = \sqrt{(1 + v)/(1 - v)}, \quad r' = \sqrt{(1 + v')/(1 - v')}.$$

При этом произведение rr' можно записать в виде

$$rr' = \left[\frac{1 + \dfrac{v + v'}{1 + vv'}}{1 - \dfrac{v + v'}{1 + vv'}} \right]^{1/2},$$

так что

$$\boxed{L(v)L(v') = L\left(\frac{v + v'}{1 - vv'} \right).} \tag{4.6}$$

Эта формула представляет *закон сложения скоростей* в специальной теории относительности.

Итак, мы пользуемся тремя способами параметризации одного и того же преобразования Лоренца:

$$L_r = L^\alpha = L(v),$$

где

$$r = e^\alpha = \sqrt{(1+v)/(1-v)}.$$

В зависимости от выбора параметризации умножение преобразований определяется формулами (4.4), (4.5) и (4.6).

Мы показали, что линейные преобразования специальной теории относительности сохраняют квадратичную форму $Q(\mathbf{v})$. Но мы еще не дали физической интерпретации $Q(\mathbf{v})$. Приведем такой пример: у нас есть световые лучи и часы. Рассмотрим (рис. 4.25) точки t_1 и t_2 на оси t, которые соединены с точкой $\begin{pmatrix} t \\ x \end{pmatrix}$ световыми лучами (прямыми, параллельными $t = x$ и $t = -x$).

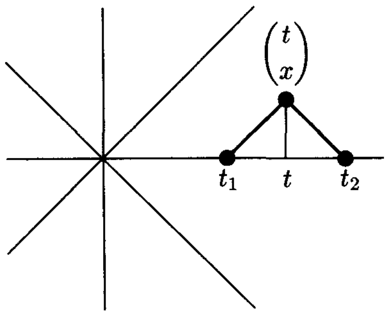


Рис. 4.25

Тогда

$$t - t_1 = x \quad \text{или} \quad t_1 = t - x,$$
$$t_2 - t = x \quad \text{или} \quad t_2 = t + x,$$

так что

$$t_1 t_2 = Q(\mathbf{v}) \quad \text{для} \quad \mathbf{v} = \begin{pmatrix} t \\ x \end{pmatrix}.$$

Наблюдатель находится в точке $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Он покоится или движется с постоянной скоростью и хочет связаться с точкой $\mathbf{v}$. Он излучает световой сигнал в момент времени t_1 и регистрирует время

t_2 возвращения к нему отраженного сигнала. Произведение $t_1 t_2$ называется *расстоянием Минковского* — интервалом между двумя событиями. Отметим, что если $\mathbf{v}$ лежит на линии $x = 0$, то $t_1 = t_2 = t$, потому что передача сигнала происходит мгновенно. Если $\mathbf{v}$ лежит на световом луче, проходящем через точку $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, то $t_1 = 0$. Если $Q(\mathbf{v}) < 0$, то $t_1 < 0$ и $t_2 > 0$ (рис. 4.26).

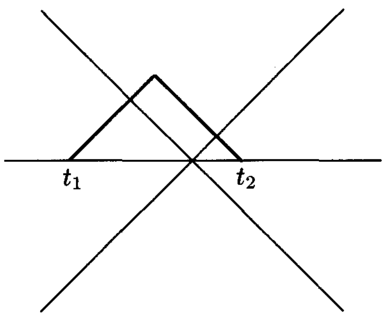


Рис. 4.26

Еще одно важное свойство геометрии пространства Минковского. Вспомним, что в геометрии Евклида у нас есть *неравенство треугольника*

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|,$$

причем равенство будет только в случае, когда $\mathbf{u}$ и $\mathbf{v}$ лежат на одной прямой в одном направлении. Это проиллюстрировано на рис. 4.27.

$$c = a + b, \quad a = \|\mathbf{u}\|, \quad b = \|\mathbf{v}\|.$$

Очевидно, что ломаная линия длиннее прямой. Это значит, что в геометрии Евклида прямая линия есть кратчайшее расстояние между двумя точками.

А теперь рассмотрим аналогичную диаграмму для пространственно-временной геометрии, когда окружности

$$\|A - P\|^2 = a^2 \quad \text{и} \quad \|B - R\|^2 = b^2$$

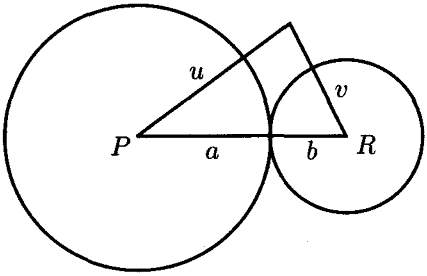


Рис. 4.27

заменяются гиперболами $Q(A-P)=a^2$ и $Q(B-R)=b^2$ (рис. 4.28). Но в этом случае для любых отрезков l и m, образующих ломаную линию между точками P и R, мы имеем

$$Q(l) < a^2 \quad \text{и} \quad Q(m) < b^2.$$

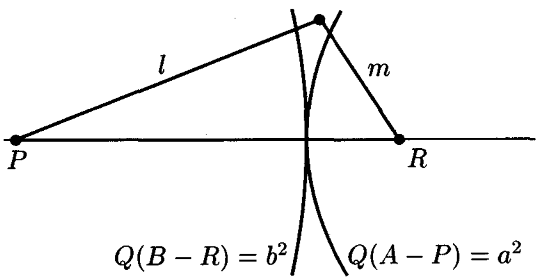


Рис. 4.28

Здесь $Q(l)$ равняется квадрату временнóго отрезка, который показывают часы, движущиеся с постоянной скоростью вдоль линии l. Итак, в этом случае мы имеем *обратное неравенство треугольника.*

Время, измеренное по часам, движущимся с постоянной скоростью из точки P в точку R будет *больше*, чем время, измеренное по часам, движущимся вдоль любой ломаной линии, соединяющей точки P и R.

Это явление называют эффектом близнецов: близнец, двигающийся вдоль ломаной линии, (если он выживет на ухабах) будет моложе другого близнеца, который двигается с постоянной скоростью из точки P в точку R. Этот эффект называют также *парадоксом близнецов*. Конечно, здесь нет никакого парадокса, мы имеем просто очевидное следствие обратного неравенства треугольника.

4.7. Группа Пуанкаре и группа Галилея

До сих пор мы описывали преобразования геометрии Евклида и специальной теории относительности в терминах естественных единиц. Точки в пространстве-времени называют *событиями*. Они отмечают, где и когда что-то произошло. Если мы зарегистрируем все события, случившиеся с одним человеком (скажем, за 70 лет) в секундах и возьмем несколько сотен или тысяч метров, тоже измеренных в секундах[4], то мы получим множество событий, протяженность которого вдоль временно́й оси примерно в 10^{18} раз больше, чем в направлении пространственных осей (рис. 4.29)[5].

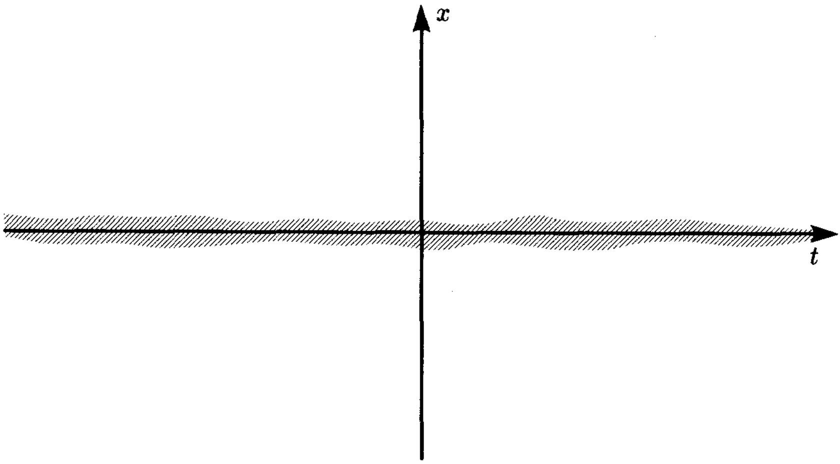


Рис. 4.29

[4]Авторы имеют в виду световую секунду, т. е. расстояние, проходимое светом за 1 секунду. — *Прим. ред.*

[5]Расчет авторов ошибочный: 70 лет/1000 метров $\approx 2 \cdot 10^{14}$. — *Прим. ред.*

Получив такую вытянутую вдоль временно́й оси область, возникает желание измерять расстояния в более мелких единицах (например в метрах), чем те единицы (секунды), которые мы применили для измерения времени. Конечно, если единица измерения будет меньше, то значение измеряемой велинины станет больше. Поэтому в «обычной» системе единиц пространственные расстояния будут сильно увеличены по сравнению с временны́ми интервалами. Конкретно это значит, что мы будем рассматривать новые переменные T и X, связанные с естественными единицами t и x соотношениями $T = t$ и $X = cx$, т. е.

$$\begin{pmatrix} T \\ X \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix},$$

где c — большое число. Световой конус $|x| = |t|$ в новой системе единиц преобразуется в $c^{-1}|X| = |T|$ или $|X| = c|T|$.

Мы будем говорить, что «в обычных единицах скорость света равна c». Аналогично, гипербола $t^2 - x^2 = k$ преобразуется в кривую $T^2 - c^2X^2 = k$ (рис. 4.30). «Времениподобные гиперболы», соответствующие $k > 0$, становятся менее кривыми, при малых значениях X они становятся почти вертикальными прямыми линиями.

Выразим преобразования Лоренца через обычные единицы. Для этого проделаем следующее: возьмем точку $\begin{pmatrix} T \\ X \end{pmatrix}$ и найдем точку

$$\begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & c^{-1} \end{pmatrix} \begin{pmatrix} T \\ X \end{pmatrix}.$$

Подействуем на нее преобразованием Лоренца L и получим

$$L \begin{pmatrix} 1 & 0 \\ 0 & c^{-1} \end{pmatrix} \begin{pmatrix} T \\ X \end{pmatrix}.$$

Выразим этот новый вектор через обычные единицы, умножив его на матрицу $\begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix}$. В результате получим:

$$\begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix} L \begin{pmatrix} 1 & 0 \\ 0 & c^{-1} \end{pmatrix} \begin{pmatrix} T \\ X \end{pmatrix}.$$

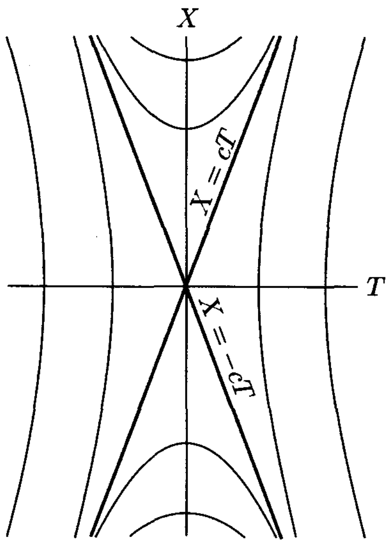


Рис. 4.30

Итак, в терминах обычных единиц преобразование Лоренца переводит вектор $\begin{pmatrix} T \\ X \end{pmatrix}$ в $M\begin{pmatrix} T \\ X \end{pmatrix}$, где матрица M равна

$$M = \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix} L \begin{pmatrix} 1 & 0 \\ 0 & c^{-1} \end{pmatrix}.$$

Пусть $L = L^\alpha$. Перемножив матрицы, получаем

$$M = \begin{pmatrix} \cosh\alpha & c^{-1}\sinh\alpha \\ c\sinh\alpha & \cosh\alpha \end{pmatrix}.$$

Эта формула справедлива для любого значения α. Вернемся к преобразованию Лоренца L, для которого параметр α мал. Введем $\tanh\alpha = \sinh\alpha/\cosh\alpha = v/c$, где, по нашему предположению, c — большая скорость, а v — обычная скорость, так что α — очень маленький параметр. Тогда

$$M = \begin{pmatrix} \cosh\alpha & c^{-1}\tanh\alpha\cosh\alpha \\ c\tanh\alpha\cosh\alpha & \cosh\alpha \end{pmatrix}$$

или

$$M = \cosh\alpha \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix},$$

где
$$\cosh\alpha = \frac{1}{(1-\tanh^2\alpha)^{1/2}} = (1 - v^2/c^2)^{-1/2}.$$

Разложение в ряд по малому параметру v^2/c^2 дает
$$\cosh\alpha = 1 + \frac{1}{2}v^2/c^2 + \cdots .$$

Подставив это разложение в формулу для M, получаем
$$M = \left(1 + \frac{1}{2}v^2/c^2 + \cdots\right)\begin{pmatrix}1 & v/c^2\\ v & 1\end{pmatrix},$$
или
$$M = \begin{pmatrix}1 & 0\\ v & 1\end{pmatrix} + E,$$
где элементы матрицы E имеют порядок величины c^{-2}. Матрица
$$G_v = \begin{pmatrix}1 & 0\\ v & 1\end{pmatrix}$$
называется *преобразованием скорости*, соответствующим значению скорости, равному v. Это преобразование сохраняет прямые $T = \text{const}$, т. к. $G_v\begin{pmatrix}T\\ X\end{pmatrix} = \begin{pmatrix}T\\ X+vT\end{pmatrix}$. Таким образом, мы видим, что преобразования скорости можно рассматривать как «предельный случай» преобразований Лоренца при малом значении α. Иными словами, предполагая обычную скорость малой по сравнению со скоростью света, мы получаем, что времениподобные гиперболы превращаются в вертикальные прямые линии, а преобразования Лоренца становятся преобразованиями скорости.

Множество преобразований скорости образует группу, элементы которой обладают свойством $G_{v_1}G_{v_2} = G_{v_1+v_2}$, что можно легко проверить. Для этой группы сохраняется понятие одновременности.

Преобразования типа $\begin{pmatrix}T\\ X\end{pmatrix} \to G_v\begin{pmatrix}T\\ X\end{pmatrix} + \begin{pmatrix}a\\ b\end{pmatrix}$ называются *преобразованиями Галилея.* Итак, преобразования Галилея — это композиции трансляций и преобразований скорости. Механика Ньютона основана на принципе относительности Галилея; ее основные понятия инвариантны относительно преобразований Галилея. Гениальность открытий Лоренца, Пуанкаре и Эйнштейна состояла в том, что они поняли приближенный характер понятия одновременности: это понятие справедливо только на малых расстояниях и для малых скоростей. А преобразование скорости $G_v = \begin{pmatrix} 1 & 0 \\ v & 1 \end{pmatrix}$ следует рассматривать как аппроксимацию преобразований Лоренца (в обычной системе единиц):

$$\begin{pmatrix} 1/\sqrt{1-v^2/c^2} & v/c^2\sqrt{1-v^2/c^2} \\ v/\sqrt{1-v^2/c^2} & 1/\sqrt{1-v^2/c^2} \end{pmatrix}.$$

4.8. Импульс, энергия и масса

Переход от группы Галилея к группе Пуанкаре потребовал переопределения базисных понятий механики. Основные контуры такой теории Пуанкаре изложил в обращении к Всемирной Выставке, проходившей в 1904 году в Сент-Луисе. После этого он и работавший независимо А. Эйнштейн опубликовали в 1905 году серию фундаментальных работ. Опишем некоторые идеи из этих работ.

В классической механике существуют два фундаментальных закона, используемых при изучении движения частиц, — закон сохранения импульса и закон сохранения энергии. Например, мы изучаем столкновение двух частиц A и B. Импульс частицы A до столкновения обозначим $\mathbf{p}_A$, а после столкновения $\mathbf{p}'_A$, аналогично для частицы B. Закон сохранения импульса говорит, что

$$\mathbf{p}_A + \mathbf{p}_B = \mathbf{p}'_A + \mathbf{p}'_B.$$

Столкновение называется упругим, если полная кинетическая энергия сохраняется. Пример неупругого столкновения: в результате столкновения двух частиц получается одна. Обратная ситуация: пусть в начальный момент две покоящиеся частицы были связаны, потом произошел взрыв, в результате которого две частицы разлетаются. Это можно рассматривать как процесс, «обратный» столкновению. Если бы мы записали его на пленку и прокрутили в обратном направлении, то увидели бы столкновение двух частиц, в результате которого получается одна частица. Полная кинетическая энергия не сохраняется — до взрыва она была равна нулю, а стала больше нуля. Энергия, освободившаяся при взрыве, преобразовалась в кинетическую энергию. В обратном процессе мы полагаем, что когда из двух частиц получается одна, кинетическая энергия преобразуется либо в потенциальную энергию, либо в тепло. Для неупругих столкновений выполняется закон сохранения импульса. А для упругих столкновений, когда нет преобразования кинетической энергии в другие формы энергии, полная кинетическая энергия тоже сохраняется:

$$E_A + E_B = E'_A + E'_B,$$

где E_A — кинетическая энергия частицы A до столкновения, E'_A — кинетическая энергия той же частицы после столкновения и т. д.

Оказывается, что в специальной теории относительности законы сохранения энергии и импульса выполняются, как и в механике Ньютона. Что необходимо изменить в определениях энергии и импульса?

Импульс частицы в механике Ньютона равен

$$\mathbf{p} = m\mathbf{v},$$

где $\mathbf{v}$ — скорость движущейся частицы и m — ее *масса*. Скорость (и, следовательно, импульс) является вектором в трехмерном пространстве. В нашей модели мы будем рассматривать его как одномерный. (Другими словами, мы ограничимся рассмотрением частиц, движущихся вдоль одной прямой. Массу можно, в принципе, измерить, проделав следующие эксперименты.

Рис. 4.31

Допустим, у нас есть несколько объектов — например, шарики, сделанные из разных материалов. Пусть два удерживаются вместе и покоятся. Далее происходит взрыв или отпускается пружина, и шарики разлетаются в разные стороны. Один полетит вправо, другой влево (рис. 4.31). Если шарики сделаны из одного материала и имеют одинаковый размер, то их движение должно быть симметричным. Например, если на равных расстояниях от точки, где покоились шарики, поместить две отражающие стенки, следует ожидать, что после отражения шарики столкнутся в той же точке. Этот эксперимент можно проделать и убедиться, что все будет происходить именно так. А теперь возьмем шарики, сделанные из одного материала, но имеющие разные размеры. Пусть правый шарик будет больше. Мы можем убедиться, что после отражения шарики столкнутся правее центра, т. е. меньший шарик пройдет большее расстояние. Тот же эксперимент можно проделать с шариками, сделанными из разных материалов. Например, пусть шарики имеют одинаковый размер, но правый сделан из свинца, а левый из алюминия. В этом случае столкновение произойдет справа от центра. Однако, если свинцовый шарик будет очень маленький, а алюминиевый того же размера, то столкновение произойдет левее центра. Если у нас много свинцовых шариков разных размеров, то среди них можно найти такой, который будет точно соответствовать алюминиевому шарику (в том смысле, что столкновение будет происходить в центре).

А теперь давайте сравнивать свинцовые шарики с медными. Допустим, мы нашли свинцовый шарик, соответствующий алюминиевому, и медный шарик, соответствующий свинцовому. Теперь мы можем сравнить алюминиевый шарик с медным. Эксперимент покажет, что они соответствуют друг другу. И это не логический вывод, а *закон природы*[6]. Теперь мы можем *определить* понятие массы: два объекта имеют *одинаковую массу*, если в нашем эксперименте они соответствуют друг другу. Закон природы, который мы обсуждаем, позволяет корректно определить понятие массы. Если A и B имеют одинаковую массу, а масса B

[6]Повернем наш прибор на 180°, т. е. поменяем местами правую и левую стороны. Этот поворот не отразится на свойствах шариков соответствовать друг другу. И это тоже закон природы.

равна массе C, то A и C имеют одинаковые массы. В эксперименте можно проверить закон природы: если шарик A_1 соответствует шарику B_1 и шарик A_2 соответствует B_2, то, экспериментируя с парами шариков: A_1 и A_2 против B_1 и B_2, можно убедиться, что A_1 и A_2 соответствуют B_1 и B_2 (рис. 4.32).

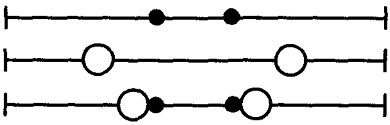


Рис. 4.32

Экспериментально можно также убедиться, что масса шарика пропорциональна его объему: если свинцовый шарик с радиусом r_A соответствует медному шарику с радиусом r_B, то свинцовый шарик с радиусом $3r_A$ будет соответствовать 27 медным шарикам с радиусом r_B.

Это позволяет нам ввести единицу массы: выберем один объект, скажем свинцовый шарик объемом $1\,\text{см}^3$. Тогда мы можем сравнивать с ним другие объекты, масса которых будет кратна m (массе выбранного свинцового шарика). Таким образом, мы получим численное значение любой массы. Первоначально в метрической системе единиц за единицу массы — 1 грамм — была взята масса 1 кубического сантиметра воды при 4°C. Поскольку нам сложно проводить эксперименты по столкновениям с водой при 4°C, мы могли бы определить 1 грамм как массу медного шарика объемом $0.11\ldots\,\text{см}^3$. Интересно отметить, что во всех описанных экспериментах не было потребности в каких-либо инструментах измерения времени.

Вернемся к закону сохранения импульса. В механике Ньютона, если мы определим импульс частицы

$$\mathbf{p} = m\mathbf{v},$$

то полный импульс сохраняется. (Нетрудно показать, что такая формулировка закона сохранения импульса является следствием нашего определения массы и предположения, что законы природы инвариантны относительно группы преобразований *Галилея*.

См. *Лекции по физике* Фейнмана, главу 10, где дается четкое изложение этого утверждения.) В специальной теории относительности такое определение импульса не имеет смысла, потому что *не имеет смысла скорость*! Ведь по определению скорость равна

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

при этом предполагается, что мы выбрали координатные оси x и t и решили параметризовать движение частицы временем t. Именно поэтому мы записываем траекторию частицы с помощью функции $x(t)$. Если мы сделаем преобразование Лоренца, то получим новые координатные оси t', x' и, следовательно, новую скорость v'. Поставим задачу немного по-другому. Допустим, что траектория частицы описывается третьим нейтральным параметром s. Например, в роли s могут быть показания неких внутренних часов, которые частица несет вместе с собой. Тогда траектория частицы на пространственно-временной плоскости дается вектором

$$\mathbf{u}(s) = \begin{pmatrix} t(s) \\ x(s) \end{pmatrix}.$$

В некоторый момент s_0 мы можем вычислить касательную к этой кривой (рис. 4.33)

$$\frac{d\mathbf{u}}{ds} = \begin{pmatrix} dt/ds \\ dx/ds \end{pmatrix} = \mathbf{w} = \begin{pmatrix} a \\ b \end{pmatrix}.$$

Тогда в системе координат t, x скорость $v = dx/dt$ равна

$$v = b/a.$$

Очевидно, что для вектора

$$\mathbf{w} = \begin{pmatrix} a \\ b \end{pmatrix}$$

отношение $v = b/a$ не имеет физического смысла. Если мы заменим $\mathbf{w}$ на

$$\mathbf{w}' = L\mathbf{w}$$

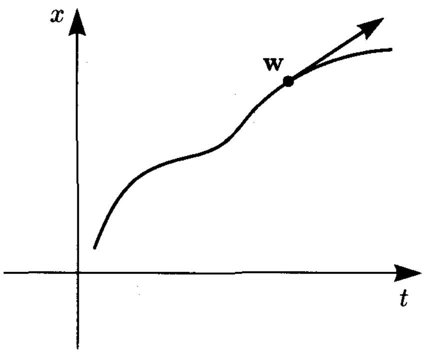


Рис. 4.33

и напишем

$$\mathbf{w}' = \begin{pmatrix} a' \\ b' \end{pmatrix}, \quad v' = b'/a',$$

то v' не будет равно v, за исключением случаев, когда $v = \pm 1$ или $L = \mathbb{I}$. При преобразованиях Лоренца сохраняется одно свойство вектора $\mathbf{w}$ — величина

$$Q(\mathbf{w}) = a^2 - b^2.$$

Условие $Q(\mathbf{w}) > 0$ эквивалентно условию $|v| < 1$. Поскольку в нашей системе единиц скорость света равна 1, это значит, что скорость v меньше скорости света.

Опыт показывает, что все частицы с положительной массой покоя (будет определена ниже) движутся со скоростями, меньшими скорости света, т. е. для них всегда $Q(\mathbf{w}) > 0$. Обозначим

$$Q(\mathbf{w}) = \mu^2 > 0,$$

так что

$$a^2 - b^2 = \mu^2,$$

и пусть

$$b/a = v$$

в определенном секторе пространства-времени. Заметим, что соотношение $b/a = v$ не определяет сами величины a, b. (Это —

отражение того факта, что мы не конкретизировали наш мистический параметр s для кривой $\mathbf{u}(s)$.) Теперь можно решить два уравнения $a^2 - b^2 = \mu^2$, $b/a = v$ и получить

$$a = \frac{\mu}{\sqrt{1-v^2}}, \quad b = \frac{\mu v}{\sqrt{1-v^2}}$$

для данного сектора пространства-времени. Для малых значений v можно воспользоваться разложением в ряд Тейлора

$$\frac{1}{\sqrt{1-v^2}} = 1 + \frac{1}{2}v^2 - \cdots ,$$

так что

$$a \doteq \mu + \frac{1}{2}\mu v^2 + \cdots ,$$
$$b \doteq \mu v + \frac{1}{2}\mu v^3 + \cdots .$$

Отметим, что разложение величины b очень похоже на $p = mv$, *если мы идентифицируем* μ *с* m и отбросим члены более высокого порядка по v. Точно так же второй член в разложении a похож на выражение для кинетической энергии в механике Ньютона. Итак, мы подошли к возможности переопределить понятия энергии и импульса. Каждому объекту соответствует определенное значение величины μ. Во избежание путаницы обозначим эту величину m_0 и назовем ее *массой покоя* данного объекта. Если оказывается, что объект покоится в некотором секторе пространства-времени, то масса покоя совпадает (с точностью до выбранной системы единиц) с массой покоя, определенной выше экспериментально. Допустим, что $m_0 > 0$ (мы это все время подразумевали). Тогда для движущегося объекта вводится *вектор энергии-импульса* (*4-импульс*)

$$\mathbf{w} = \begin{pmatrix} E \\ \mathbf{p} \end{pmatrix}$$

такой, что

$$Q(\mathbf{w}) = E^2 - \mathbf{p}^2 = m_0^2$$

и $\mathbf{w}$ есть произведение $\dot{\mathbf{u}} = d\mathbf{u}/ds$ на скалярный множитель, где $\mathbf{u}(s)$ — траектория объекта, движущегося в пространстве-времени.

В данном секторе пространства-времени, где траектория объекта задается формулой

$$\mathbf{u}(s) = \begin{pmatrix} t(s) \\ x(s) \end{pmatrix},$$

соответственно

$$\dot{\mathbf{u}}(s) = \begin{pmatrix} \dot{t}(s) \\ \dot{x}(s) \end{pmatrix}, \quad v = \frac{dx}{dt} = \frac{\dot{x}(s)}{\dot{t}(s)},$$

мы получаем

$$p = \frac{m_0 v}{\sqrt{1-v^2}}, \quad E = \frac{m_0}{\sqrt{1-v^2}}.$$

В частности, если объект покоится, так что $v = 0$, то

$$p = 0, \quad E = m_0$$

в данной системе координат.

Закон сохранения энергии-импульса в новой формулировке выглядит как закон сохранения для *векторов* в пространстве-времени при любом столкновении:

$$\begin{pmatrix} E_A \\ \mathbf{p}_A \end{pmatrix} + \begin{pmatrix} E_B \\ \mathbf{p}_B \end{pmatrix} = \begin{pmatrix} E'_A \\ \mathbf{p}'_A \end{pmatrix} + \begin{pmatrix} E'_B \\ \mathbf{p}'_B \end{pmatrix}$$

Эти уравнения записаны в естественной системе единиц, когда скорость света равна 1, а v — число. Поэтому выражение $\sqrt{1-v^2}$ имеет смысл. Если же мы будем работать с «психологическими единицами», то v не будет числом, а будет скоростью, выраженной, например, в см/сек. Тогда выражение $\sqrt{1-v^2}$ уже не имеет смысла. В этом случае $\sqrt{1-v^2}$ следует заменить на $\sqrt{1-v^2/c^2}$. Соответственно изменяется выражение для p:

$$p = \frac{m_0 \mathbf{v}}{\sqrt{1-v^2/c^2}}.$$

При малых значения v мы получим ньютоновское выражение для импульса. Аналогично изменяется формула для энергии E:

$$E = \frac{m_0c^2}{\sqrt{1 - v^2/c^2}}.$$

Для покоящейся частицы мы получаем известную формулу Эйнштейна

$$E = m_0c^2.$$

4.9. Антисимметричные формы

В предыдущих параграфах мы рассмотрели два типа скалярных произведений на плоскости. Первое — евклидово скалярное произведение, определяемое формулой

$$(\mathbf{w}, \mathbf{w}') = xx' + yy', \quad \text{где} \quad \mathbf{w} = \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{и} \quad \mathbf{w}' = \begin{pmatrix} x' \\ y' \end{pmatrix};$$

второе — лоренцево скалярное произведение

$$(\mathbf{v}, \mathbf{v}') = tt' - xx', \quad \text{где} \quad \mathbf{v} = \begin{pmatrix} t \\ x \end{pmatrix} \quad \text{и} \quad \mathbf{v}' = \begin{pmatrix} t' \\ x' \end{pmatrix}.$$

Оба типа скалярных произведений билинейны, т. е. если один вектор зафиксировать, то произведение будет линейной функцией от другого вектора:

$$(a\mathbf{w}_1 + b\mathbf{w}_2, \mathbf{w}') = a(\mathbf{w}_1, \mathbf{w}') + b(\mathbf{w}_2, \mathbf{w}').$$

Оба скалярных произведения симметричны:

$$(\mathbf{w}, \mathbf{w}') = (\mathbf{w}', \mathbf{w}) \quad \text{и} \quad (\mathbf{v}, \mathbf{v}') = (\mathbf{v}', \mathbf{v}).$$

Введем третий тип произведения между двумя векторами на плоскости, который будет по-прежнему билинейным, но вместо симметричности обладает свойством антисимметрии. Именно, определим следующую функцию от двух векторов:

$$\omega(\mathbf{v}, \mathbf{v}') = qp' - q'p = \operatorname{Det}\begin{pmatrix} q & q' \\ p & p' \end{pmatrix}, \quad \text{где} \quad \mathbf{v} = \begin{pmatrix} q \\ p \end{pmatrix} \quad \text{и} \quad \mathbf{v}' = \begin{pmatrix} q' \\ p' \end{pmatrix}.$$

Здесь выполняется тождество

$$\omega(\mathbf{v}, \mathbf{v}') = -\omega(\mathbf{v}', \mathbf{v}),$$

которое мы и называем *антисимметричностью.*

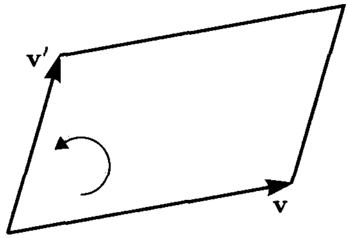


Рис. 4.34

Геометрический смысл $\omega(\mathbf{v}, \mathbf{v}')$ понятен: это ориентированная площадь параллелограмма, построенного на векторах $\mathbf{v}$ и $\mathbf{v}'$ (рис. 4.30). Очевидно, что функция $\omega(\mathbf{v}, \mathbf{v}')$ билинейна. Величина ω называется *симплектическим* скалярным произведением.

Линейное преобразование A называется *симплектическим*, если оно сохраняет симплектическое скалярное произведение ω. Таким образом, A симплектично тогда и только тогда, когда для любых $\mathbf{v}$, $\mathbf{v}'$

$$\omega(A\mathbf{v}, A\mathbf{v}') = \omega(\mathbf{v}, \mathbf{v}').$$

Матрица, столбцы которой $A\mathbf{v}$ и $A\mathbf{v}'$, получается умножением матрицы A на матрицу $\begin{pmatrix} q & q' \\ p & p' \end{pmatrix}$. Отсюда

$$\omega(A\mathbf{v}, A\mathbf{v}') = \operatorname{Det}\left[A\begin{pmatrix} q & q' \\ p & p' \end{pmatrix}\right]$$
$$= \operatorname{Det} A \operatorname{Det}\begin{pmatrix} q & q' \\ p & p' \end{pmatrix} = \operatorname{Det} A\, \omega(\mathbf{v}, \mathbf{v}').$$

Следовательно, преобразование A является симплектическим тогда и только тогда, когда $\operatorname{Det} A = 1$. Очевидно, что любая симплектическая матрица имеет обратную, которая также симплектична, и произведение двух симплектических матриц тоже всегда симплектично. Это означает, что совокупность всех симплектических матриц образует группу, называемую (двумерной) *симплектической группой*. Симплектическая группа играет очень важную роль в оптике, мы увидим это в главе 9.

Резюме

A. Евклидово скалярное произведение

Изучив эту главу, вы должны усвоить понятие евклидова скалярного произведения и научиться пользоваться его свойствами.

Вы должны уметь транспонировать матрицу и применять операцию транспонирования в связи со скалярным произведением и евклидовыми преобразованиями.

Вы должны знать, как в векторном пространстве двух и более измерений с евклидовым скалярным произведением применить процесс Грама–Шмидта для построения ортонормированного базиса и для получения ортогональной проекции вектора на подпространство.

B. Квадратичные формы

Вы должны уметь выразить квадратичную форму $Q(\mathbf{v})$ через симметричную матрицу A и связать максимальные и минимальные значения Q с собственными векторами и собственными значениями A.

Если на плоскости задана квадратичная форма Q, то вы должны уметь ввести такие координаты x' и y', в которых она имеет вид

$$Q = \lambda_1 {x'}^2 + \lambda_2 {y'}^2,$$

и воспользоваться этими координатами для построения графика $Q = \text{const}$.

C. Связанные колебания

Вы должны уметь свести задачу о двух связанных осцилляторах к уравнению $\ddot{\mathbf{w}} = -A\mathbf{w}$ и выразить нормальные колебания через собственные векторы и собственные значения A.

D. Лоренцево скалярное произведение

Вы должны уметь вычислять лоренцево скалярное произведение двух векторов, определять преобразования Лоренца, которые сохраняют это скалярное произведение. Применять эти понятия в специальной теории относительности.

Задачи

4.1. (a) Используя три свойства скалярного произведения (симметричность, линейность, положительная определенность), докажите *неравенство Коши–Шварца*

$$(\mathbf{v}, \mathbf{w}) \leq \sqrt{(\mathbf{v}, \mathbf{v})(\mathbf{w}, \mathbf{w})}$$

для любой пары векторов $\mathbf{v}$ и $\mathbf{w}$. (Указание: рассмотрите скалярное произведение $(\mathbf{v} - \alpha\mathbf{w}, \mathbf{v} - \alpha\mathbf{w})$. Этот квадратичный полином по α не может иметь вещественных корней, кроме случая $\mathbf{v} = \alpha\mathbf{w}$.)

(b) Докажите *неравенство треугольника*

$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|,$$

где $\|\mathbf{v}\|^2 = (\mathbf{v}, \mathbf{v})$. (Указание: возведите неравенство в квадрат и примените неравенство Коши–Шварца из пункта (a).)

4.2. (a) На плоскости заданы два вектора $\mathbf{v}$ и $\mathbf{v}'$. Покажите, что поворот R_θ на угол θ, для которого

$$\cos\theta = \frac{(\mathbf{v}, \mathbf{v}')}{\sqrt{(\mathbf{v}, \mathbf{v})(\mathbf{v}', \mathbf{v}')}},$$

сделает вектор $\mathbf{v}$ пропорциональным вектору $\mathbf{v}'$. Найдите угол между векторами $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ и $\begin{pmatrix} 4 \\ -2 \end{pmatrix}$.

(b) В двумерном пространстве-времени заданы два вектора $\mathbf{v}$ и $\mathbf{v}'$. Оба вектора либо пространственноподобны, либо времениподобны вперед или времениподобны назад. Покажите, что собственное преобразование Лоренца L_α, для которого

$$\cosh\alpha = \frac{\{\mathbf{v}, \mathbf{v}'\}}{\sqrt{\{\mathbf{v}, \mathbf{v}\}\{\mathbf{v}', \mathbf{v}'\}}},$$

переводит вектор $\mathbf{v}$ в вектор, пропорциональный вектору $\mathbf{v}'$.

Воспользуйтесь этим результатом и найдите преобразование Лоренца, переводящее вектор $\begin{pmatrix}5\\4\end{pmatrix}$ в вектор, пропорциональный вектору $\begin{pmatrix}5\\3\end{pmatrix}$.

Какая проблема возникнет, если вектор $\mathbf{v}$ — пространственноподобный, а вектор $\mathbf{v}'$ — времениподобный? Если вектор $\mathbf{v}$ — времениподобный вперед, а $\mathbf{v}'$ — времениподобный назад? Если векторы $\mathbf{v}$ и $\mathbf{v}'$ находятся на световом конусе?

4.3. Чтобы научиться вычислять лоренцевы скалярные произведения, поработайте с векторами в двумерном пространстве-времени. (Первая координата соответствует времени t, а вторая оси x.)

$$\mathbf{v}_1 = \begin{pmatrix}-2\\2\end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix}-1\\3\end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix}1\\3\end{pmatrix},$$
$$\mathbf{v}_4 = \begin{pmatrix}1\\1\end{pmatrix}, \quad \mathbf{v}_5 = \begin{pmatrix}3\\-1\end{pmatrix}, \quad \mathbf{v}_6 = \begin{pmatrix}3\\1\end{pmatrix}.$$

(a) Вычислите лоренцево скалярное произведение каждого вектора на самого себя $\{\mathbf{v}, \mathbf{v}\}$. В системе координат x, t изобразите каждый вектор и разберитесь, к какому типу он относится: пространственноподобный, времениподобный вперед или назад, лежит на световом конусе вперед или назад.

(b) Вычислите лоренцевы скалярные произведения $\{\mathbf{v}_2, \mathbf{v}_3\}$, $\{\mathbf{v}_5, \mathbf{v}_6\}$ и $\{\mathbf{v}_3, \mathbf{v}_6\}$.

(c) Найдите векторы $\mathbf{w}_1, \ldots, \mathbf{w}_6$, в которые переходят векторы $\mathbf{v}_1, \ldots, \mathbf{v}_6$ под действием преобразования Лоренца

$$L_2 = \begin{pmatrix}5/4 & 3/4\\3/4 & 5/4\end{pmatrix}.$$

Изобразите полученные векторы в системе координат x, t.

(d) Вычислите скалярные произведения $\{\mathbf{w}_2, \mathbf{w}_2\}$, $\{\mathbf{w}_6, \mathbf{w}_6\}$, $\{\mathbf{w}_2, \mathbf{w}_3\}$, $\{\mathbf{w}_5, \mathbf{w}_6\}$ и $\{\mathbf{w}_3, \mathbf{w}_6\}$. Они должны совпадать с соответствующими скалярными произведениями векторов $\mathbf{v}$.

4.4. Пусть матрица S симметрична и имеет положительные собственные значения. Введем новое скалярное произведение $[\mathbf{v}, \mathbf{w}]_S$, определяемое уравнением $[\mathbf{v}, \mathbf{w}]_S = (S\mathbf{v}, \mathbf{w})$.

(a) Покажите, что это скалярное произведение симметрично, билинейно и положительно определено.

(b) Покажите, что матрица C сохраняет это скалярное произведение (т. е. $[C\mathbf{v}, C\mathbf{w}]_S = [\mathbf{v}, \mathbf{w}]_S$) тогда и только тогда, когда $C^T SC = S$).

(c) Опишите процедуру построения матрицы B, обладающей следующим свойством: если $\mathbf{v}' = B^{-1}\mathbf{v}$, $\mathbf{w}' = B^{-1}\mathbf{w}$, то $[\mathbf{v}, \mathbf{w}]_S = (\mathbf{v}', \mathbf{w}')$. Объясните как, имея одну матрицу B, обладающую этим свойством, можно построить много других.

4.5. Пусть в предыдущем упражнении $S = \begin{pmatrix} 3.7 & 0.9 \\ 0.9 & 1.3 \end{pmatrix}$.

(a) Получите вектор $\mathbf{v}$, ортогональный вектору $\mathbf{w} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, для скалярного произведения, определенного матрицей S так, что $[\mathbf{v}, \mathbf{w}]_S = 0$.

(b) Постройте матрицу B, обладающую свойствами, описанными в задаче 4.4(c). Проверьте, что для векторов $\mathbf{v}$ и $\mathbf{w}$ из пункта (a) выполняется равенство $(B^{-1}\mathbf{v}, B^{-1}\mathbf{w}) = 0$.

(c) Постройте матрицу *ортогонального проектирования* P, удовлетворяющую уравнению $P^2 = P$, чей образ состоит из векторов, пропорциональных вектору $\mathbf{w} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Проверьте, что для всех векторов $\mathbf{v}$ выполняется равенство $[\mathbf{w}, P\mathbf{v}]_S = [\mathbf{w}, \mathbf{v}]_S$.

(d) Постройте матрицу C, которая сохраняет скалярное произведение, определяемое матрицей S, и удовлетворяет уравнению $C^2 = -\mathbb{I}$. (Указание: матрица $R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ удовлетворяет уравнению $R^2 = -\mathbb{I}$ и сохраняет обычное скалярное произведение.

4.6. Для квадратичной формы

$$Q(\mathbf{v}) = 8x^2 + 12xy + 17y^2$$

проделайте следующие операции.

(a) Запишите Q в виде $(A\mathbf{v}, \mathbf{v})$, где A — симметричная матрица.

(b) Найдите собственные значения матрицы A.

(c) Представьте матрицу A в форме

$$A = R_\theta \begin{pmatrix} \lambda_1 & 0 \\ & \lambda_2 \end{pmatrix} R_\theta^{-1}.$$

(d) Найдите такие координаты x' и y', для которых квадратичная форма Q принимает вид

$$Q(\mathbf{v}) = 20x'^2 + 5y'^2.$$

(e) Нарисуйте линию $Q(\mathbf{v}) = 20$. Отметьте на ней координатные оси x, y и x', y'.

4.7. (a) Найдите собственные значения λ_1 и λ_2 матрицы $S = \begin{pmatrix} 9 & 2 \\ 2 & 6 \end{pmatrix}$, определите собственные векторы $\mathbf{v}_1$ и $\mathbf{v}_2$, соответствующие этим собственным значениям.

(b) Постройте такую матрицу вращения R, чтобы выполнялось равенство $S = R\Lambda R^{-1}$, где Λ — диагональная матрица. Проверьте, что R на самом деле осуществляет поворот!

(c) Найдите новые координаты x', y', являющиеся линейными функциями x и y, для которых

$$9x^2 + 4xy + 6y^2 = \lambda_1 x'^2 + \lambda_2 y'^2.$$

4.8. (a) Найдите собственные значения и собственные векторы матрицы

$$A = \begin{pmatrix} 10 & 6 \\ 6 & 10 \end{pmatrix}.$$

(b) Постройте матрицу вращения R и диагональную матрицу Λ, для которых выполняется соотношение $A = R\Lambda R^{-1}$.

(c) Нарисуйте график уравнения $10x^2 + 12xy + 10y^2 = 24$.

4.9. (a) Найдите такие координаты x', y', чтобы квадратичная форма

$$Q(\mathbf{v}) = -x^2 + 6xy + 7y^2$$

приняла вид

$$Q(\mathbf{v}) = \lambda_1 x'^2 + \lambda_2 y'^2.$$

Нарисуйте линию $Q(\mathbf{v}) = 40$.

(b) Пусть x и y лежат на окружности единичного радиуса, так что $x = \cos\theta$, $y = \sin\theta$. Найдите значения углов θ, при которых Q принимает максимальное и минимальное значение. Вычислите эти значения. Какова связь между этими результатами и результатами, полученными в предыдущем пункте?

4.10. Предположим, что M и K — *симметричные* матрицы 2×2.

(a) Приведите пример, показывающий, что $M^{-1}K$ не обязательно симметрична.

(b) Опишите, как построить такую *симметричную* матрицу B, чтобы выполнялось равенство $B^2 = M^{-1}$. Покажите, что матрица $S = BKB$ симметрична, и, следовательно, ее можно записать в форме $S = R\Lambda R^{-1}$, где R — матрица вращения, а $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$.

(c) Покажите, что если $A = BR$, то $M^{-1}K = A\Lambda A^{-1}$. Это доказывает, что $M^{-1}K$ имеет вещественные собственные значения.

(d) Найдите новые координаты x' и y' из условия $\begin{pmatrix} x' \\ y' \end{pmatrix} = A^{-1}\begin{pmatrix} x \\ y \end{pmatrix}$. Покажите, что если $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$, то

$$(M\mathbf{v}, \mathbf{v}) = x'^2 + y'^2, \quad \text{а} \quad (K\mathbf{v}, \mathbf{v}) = \lambda_1 x'^2 + \lambda_2 y'^2.$$

(Указание: B — симметричная матрица, поэтому $(B\mathbf{v}, \mathbf{w}) = (\mathbf{v}, B\mathbf{w})$. R — ортогональная матрица, следовательно,

$$(R\mathbf{v}, \mathbf{w}) = (\mathbf{v}, R^{-1}\mathbf{w}).)$$

4.11. (a) Покажите, что если $A = \begin{pmatrix} 0 & \alpha \\ \alpha & 0 \end{pmatrix}$, то $\exp(tA)$ является преобразованием Лоренца.

(b) В релятивистской механике полная энергия E и импульс p частицы, имеющей массу m и движущейся вдоль прямой, образуют вектор $\mathbf{v} = \begin{pmatrix} E \\ p \end{pmatrix}$, при этом $\langle \mathbf{v}, \mathbf{v} \rangle = E^2 - p^2 = m^2$. Если частица движется так, что ее ускорение всегда равно α с точки зрения наблюдателя, который видит частицу покоящейся в данный момент времени, то E и p связаны между собой уравнениями

$$\frac{dE}{d\tau} = \alpha p, \quad \frac{dp}{d\tau} = \alpha E,$$

где τ — время, измеренное по часам, движущимися вместе с частицей. Решите эту систему уравнений с начальными условиями $\begin{pmatrix} E_0 \\ p_0 \end{pmatrix}$ при $\tau = 0$ и получите вектор $\begin{pmatrix} E(\tau) \\ p(\tau) \end{pmatrix}$.

4.12. Допустим, что расстояния вдоль двух взаимно перпендикулярных осей на плоскости измеряются в единицах, которые отличаются большим множителем c. Например, у нас есть сверхскоростная прямая дорога. Ее проекция на ось x равна 1000 километрам, а проекция на ось y только 1000 сантиметрам. Мы можем ввести новые координаты $X = x$ и $Y = cy$, где $c = 10^5$. Теперь X измеряется в километрах, а Y в сантиметрах. Постройте матрицу поворота на угол θ в системе координат X, Y. Покажите, что для прямых, чей наклон X/Y порядка единицы, матрица вращения превращается в матрицу сдвига в пределе $c \to \infty$. Объясните это явление геометрически, рассмотрев, что происходит с окружностями $x^2 + y^2 = k$. (Замечание: выполнив это упражнение, еще раз прочитайте в параграфе 4.3 обсуждение предела $c \to \infty$ для преобразований Лоренца.)

4.13. Вычислите симплектическое скалярное произведение $\omega(\mathbf{v}_1, \mathbf{v}_2)$ для векторов

$$\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

Проверьте непосредственно, что это скалярное произведение сохраняется при действии на него симплектической матрицы

$$A = \begin{pmatrix} 5 & 3 \\ -2 & -1 \end{pmatrix}.$$

4.14. Рассмотрите конструкцию, состоящую из масс и пружин, изображенную на рис. 4.35.

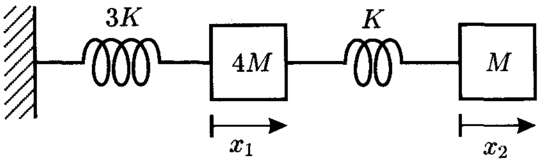


Рис. 4.35

(a) Покажите, что если x_1 и x_2 задают смещения масс вправо от положения равновесия, то уравнение движения этой системы имеет вид

$$T \begin{pmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{pmatrix} = -H \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

где

$$T = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{и} \quad H = \begin{pmatrix} 4 & -1 \\ -1 & 1 \end{pmatrix}.$$

(b) Пусть B — диагональная матрица с положительными матричными элементами, удовлетворяющая уравнению $B^2 = T$. Постройте матрицу $A = B^{-1}HB^{-1}$. Найдите ее собственные значения и собственные векторы, используйте их для нахождения общего решения уравнения $\ddot{\mathbf{w}} = -A\mathbf{w}$.

(c) Опишите нормальные колебания системы, определив их частоту через $\omega_0 = \sqrt{(K/M)}$ и отношение x_2/x_1.

4.15. Рассмотрите систему масс и пружин, изображенную на рис. 4.36. Пусть x_1 и x_2 обозначают смещение масс вправо от положения равновесия.

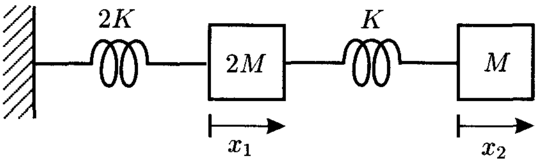


Рис. 4.36

(a) Определите частоты нормальных колебаний ω_a и ω_b. Для каждого колебания получите отношение x_2/x_1.

(b) Пусть в начальный момент смещения масс равны $x_1 = A$, $x_2 = 0$, и они покоятся. Потом их отпускают. Найти выражения $x_1(t)$ и $x_2(t)$, определяющие дальнейшее движение системы.

4.16. Частица с 4-импульсом $\begin{pmatrix} E \\ p \end{pmatrix}$ подвергается действию преобразования Лоренца L с матрицей

$$\frac{1}{2}\begin{pmatrix} r + r^{-1} & r - r^{-1} \\ r - r^{-1} & r + r^{-1} \end{pmatrix}.$$

Покажите, что при этом сумма энергии и импульса умножается на r, а их разность делится на r. Объясните этот результат в терминах собственных значений и собственных векторов преобразования L.

4.17. Масса частицы равна 15 (в произвольных единицах), а ее скорость равна $u = 12/13$ ($c = 1$). Эта частица сталкивается с покоящейся частицей, масса которой равна 6, и они образуют одну частицу.

(a) Получите 4-импульсы $\begin{pmatrix} E \\ p \end{pmatrix}$ всех трех частиц, участвующих в реакции. Определите массу и скорость образовавшейся частицы.

(b) Возьмите преобразование Лоренца $L = \begin{pmatrix} 5/3 & -4/3 \\ -4/3 & 5/3 \end{pmatrix}$, соответствующее скорости $\frac{4}{5}c$. Определите 4-импульсы каждой частицы $\begin{pmatrix} E' \\ p' \end{pmatrix}$ в системе координат, движущейся со скоростью $\frac{4}{5}c$.

4.18. Допустим, что две частицы имеют 4-импульсы $\mathbf{w}_1 = \begin{pmatrix} E_1 \\ p_1 \end{pmatrix}$ и $\mathbf{w}_2 = \begin{pmatrix} E_2 \\ p_2 \end{pmatrix}$ соответственно, где

$$m_1^2 = E_1^2 - p_1^2, \quad m_2^2 = E_2^2 - p_2^2.$$

(a) Напишите лоренцево скалярное произведение этих векторов $\{\mathbf{w}_1, \mathbf{w}_2\} = m_1 m_2 \cosh\alpha$. Покажите, что $v = \tanh\alpha = \sqrt{\cosh^2\alpha - 1}/\cosh\alpha$ представляет скорость одной из частиц *в системе покоя другой частицы.*

(b) Определите скорость v для двух случаев:

$$\mathbf{w}_1 = \begin{pmatrix} 13 \\ -5 \end{pmatrix}, \quad \mathbf{w}_2 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$$

и

$$\mathbf{w}_1 = \begin{pmatrix} 13 \\ 5 \end{pmatrix}, \quad \mathbf{w}_2 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}.$$

4.19. В системе единиц, где c не равняется 1, матрица преобразования Лоренца, действующая на 4-импульс $\begin{pmatrix} t \\ x \end{pmatrix}$, равна

$$L = \begin{pmatrix} \cosh\alpha & (1/c)\sinh\alpha \\ c\sinh\alpha & \cosh\alpha \end{pmatrix}.$$

(a) Покажите, что на 4-импульс $\begin{pmatrix} E \\ p \end{pmatrix}$ действует матрица, *транспонированная* этой.

(b) Покажите, что та же матрица L действует на 4-импульс, если мы представим его в виде *строки*, т. е.

$$(E', p') = (E, p)L.$$

4.20. В системе единиц, где $c = 1$, энергия фотона равна его импульсу, т. е. 4-импульс фотона равен $E\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ или $E\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ в зависимости от направления его движения.

(a) Допустим, что покоящаяся частица с массой $2m$ распадается на частицу с массой m и фотон. Используя закон сохранения 4-импульса, получите скорость частицы с массой m и энергию фотона.

(b) Используйте преобразование Лоренца для описания этого распада в системе координат, где частица с массой $2m$ движется со скоростью $\frac{3}{5}$.

4.21. Фотон имеет энергию E_γ; соответственно, в системе координат, где $c = 1$, 4-импульс фотона равен $\begin{pmatrix} E_\gamma \\ E_\gamma \end{pmatrix}$. Фотон сталкивается с неподвижной частицей (масса m_1), и образуется одна частица с массой m_2. Покажите, что

$$E_\gamma = \frac{m_2^2 - m_1^2}{2m_1}.$$

4.22. Используя скалярное произведение $(f, g) = \int_0^\infty f(t)g(t)\,dt$, постройте ортонормированный базис для пространства функций, удовлетворяющих дифференциальному уравнению

$$\ddot{x} + 3\dot{x} + 2x = 0.$$

4.23. Постройте ортонормированный базис для подпространства $\mathbb{R}^4$, определяемого векторами

$$\mathbf{v}_1 = \begin{pmatrix} 2 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 4 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 0 \\ 4 \\ 1 \\ 2 \end{pmatrix}.$$

4.24. В пространстве $\mathbb{R}^2$ определите скалярное произведение согласно формуле $(\mathbf{v}_1, \mathbf{v}_2) = 4x_1x_2 + y_1y_2$. Постройте матрицу P (2×2), которая осуществляет ортогональную проекцию любого вектора $\mathbf{v}$ (по отношению к скалярному произведению, определенному выше) на прямую, порожденную вектором $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Покажите, что $(\mathbb{I} - P)_{\mathbf{v}}$ ортогональна $P_{\mathbf{v}}$.

Глава 5

Дифференциальное исчисление на плоскости

В главах 5 и 6 излагаются основы дифференциального исчисления. В главе 5 мы определим дифференциал отображения одного векторного пространства на другое, обсудим его основные свойства, в частности, цепное правило. Дадим примеры физических приложений, в частности, движение Кеплера и приближение Борна. Дадим определение частной производной, производной по направлению и линейных дифференциальных форм.

Введение

Наша первая задача — разработать дифференциальное исчисление для четырех типов отображений:

(i) $\mathbb{R}^1 \to \mathbb{R}^2$,

(ii) $\mathbb{R}^2 \to \mathbb{R}^1$,

(iii) $\mathbb{R}^1 \to \mathbb{R}^1$,

(iv) $\mathbb{R}^2 \to \mathbb{R}^2$.

Отображения первого типа $\mathbb{R}^1 \to \mathbb{R}^2$ можно изобразить, как кривые на плоскости (рис. 5.1), функции $\mathbb{R}^2 \to \mathbb{R}^1$ — как поверхности в трехмерном пространстве (рис. 5.2). Функции $\mathbb{R}^1 \to \mathbb{R}^1$ знакомы читателю с первого года изучения математического анализа. В главе 1 рассмотрены *линейные* отображения $\mathbb{R}^2 \to \mathbb{R}^2$.

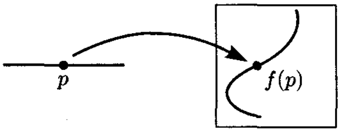


Рис. 5.1. Функция из $\mathbb{R}^1$ в $\mathbb{R}^2$

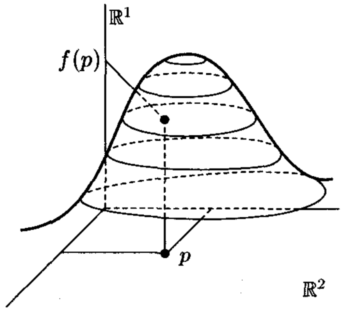


Рис. 5.2. Функция из $\mathbb{R}^2$ в $\mathbb{R}^1$

Сейчас мы хотим расширить изучение и включить в рассмотрение нелинейные отображения одной плоскости в другую (рис. 5.3). Чтобы не рассматривать каждый случай отдельно, введем некоторые единообразные обозначения. В дальнейшем буквы V, W, Z и т. д. будут обозначать либо $\mathbb{R}^1$, либо $\mathbb{R}^2$. Поэтому, когда мы пишем

$$f : V \to W$$

(читайте «f отображает V в W», или, другими словами, «f есть отображение из V в W»), то это значит, что рассматривается любой из четырех случаев: V — это $\mathbb{R}^1$ или $\mathbb{R}^2$, и W — это $\mathbb{R}^1$ или $\mathbb{R}^2$.

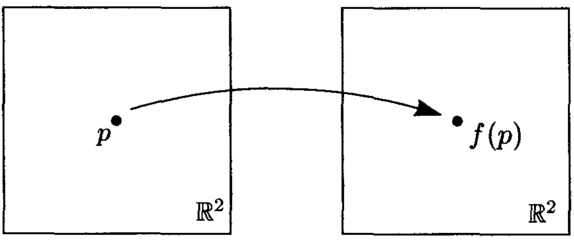


Рис. 5.3

На самом деле наши обозначения и доказательства будут таковы, что позволят множествам V, W и т. д. быть пространствами $\mathbb{R}^n$, т. е. они могут быть любыми конечномерными вещественными векторными пространствами или аффинными пространствами (эти пространства мы будем изучать в главе 10). В этой главе мы проделаем некоторые общие вычисления, даже не имея в распоряжении всех формальных определений.

Сначала отметим факт, вероятно, уже известный читателю, который следует из простого обобщения нашего обсуждения в главе 1: линейное отображение пространства $\mathbb{R}^p$ в $\mathbb{R}^q$ определяется матрицей, имеющей q строк и p столбцов. Итак, матрица

$$A = \begin{pmatrix} 5 & 2 & 1 & 0 & 1 \\ 4 & 1 & -1 & 0 & 3 \\ 3 & 0 & 1 & 1 & 0 \end{pmatrix}$$

задает линейное отображение $\mathbb{R}^5$ в $\mathbb{R}^3$:

$$A\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \\ 3 \end{pmatrix}, \quad A\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \quad \text{и т. д.}$$

Если

$$B = \begin{pmatrix} 1 & 5 & 9 \\ 2 & 4 & 6 \\ 3 & 7 & 11 \\ 4 & 8 & 12 \end{pmatrix},$$

т. е. B отображает $\mathbb{R}^3 \to \mathbb{R}^4$, то BA отображает $\mathbb{R}^5 \to \mathbb{R}^4$ и, следовательно, является матрицей, имеющей четыре строки и пять столбцов, причем матричные элементы вычисляются по обычным правилам умножения матриц.

$$BA = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & x & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}.$$

В нашем примере

$$x = (2)(1) + (6)(-1) + (10)(1).$$

В частности, линейное отображение $\mathbb{R}^p \to \mathbb{R}^1 = \mathbb{R}$ (обычно называемое линейной функцией) определяется матрицей, у которой одна строка и p столбцов. Такую матрицу называют вектором-строкой. Итак, $\mathbf{I} = (1, 2, 3, 4)$ является линейным отображением $\mathbb{R}^4 \to \mathbb{R}$, так что

$$\mathbf{I}\begin{pmatrix}1\\0\\0\\0\end{pmatrix} = 1, \quad \mathbf{I}\begin{pmatrix}0\\1\\0\\0\end{pmatrix} = 2, \quad \text{и т. д.}$$

Результат действия этого отображения на произвольный вектор задается формулой

$$\mathbf{I}\begin{pmatrix}x\\y\\z\\w\end{pmatrix} = x + 2y + 3z + 4w.$$

Еще пример: пусть вектор-строка $\mathbf{I} = (a, b, c, d)$ действует на вектор-столбец

$$\mathbf{v} = \begin{pmatrix}x\\y\\z\\w\end{pmatrix}.$$

Согласно обычным правилам умножения матриц, мы получаем один матричный элемент

$$\mathbf{I}(\mathbf{v}) = ax + by + cz + dw.$$

Если $A : \mathbb{R}^p \to \mathbb{R}^q$ и $\mathbf{I} : \mathbb{R}^q \to \mathbb{R}$, то $\mathbf{I} \circ A : \mathbb{R}^p \to \mathbb{R}$ определяется матричным умножением: матрица $1 \times q$ умножается на матрицу $q \times p$. Например, если $q = 3$, $p = 5$, $\mathbf{I} = (1, 2, 3)$,

$$A = \begin{pmatrix}5 & 2 & 1 & 0 & 1\\4 & 1 & -1 & 0 & 3\\3 & 0 & 1 & 1 & 0\end{pmatrix},$$

то

$$\mathbf{I}A = (1, 2, 3)\begin{pmatrix}5 & 2 & 1 & 0 & 1\\4 & 1 & -1 & 0 & 3\\3 & 0 & 1 & 1 & 0\end{pmatrix} = (22, 4, 2, 3, 7).$$

Еще небольшое напоминание об обозначениях из параграфа 4.1. В пространстве $\mathbb{R}^k$ мы имеем евклидово скалярное произведение $(\cdot,\cdot)$ и соответствующую норму $\|\cdot\| \geq 0$, определяемую формулой

$$\|\mathbf{v}\|^2 = (\mathbf{v},\mathbf{v}) = x_1^2 + x_2^2 + \cdots + x_k^2$$

для

$$\mathbf{v} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}.$$

Неравенство треугольника утверждает, что при любых $\mathbf{u}$ и $\mathbf{v}$

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

5.1. «О» большие и «о» малые

В одномерном дифференциальном исчислении говорится, что функция f, зависящая от одной переменной, имеет *производную* A в точке x, если f определена в некоторой окрестности точки x и разностное отношение

$$\frac{f(x+v)-f(x)}{v},$$

определенное для всех достаточно малых $v \neq 0$, имеет предел A при $v \to 0$. Мы хотели бы обобщить это определение на отображения $f : V \to W$. Первое препятствие состоит в том, что деление на вектор не имеет смысла, и мы не можем использовать понятие разностного отношения. Поэтому в многомерном случае обойдемся без дробей:

$$f(\mathbf{x}+\mathbf{v}) - f(\mathbf{x}) = A\mathbf{v} + \phi(\mathbf{v}). \tag{5.1}$$

По определению, A будет производной f в точке x, если поправочный член $\phi(\mathbf{v})$ стремится к нулю «быстрее, чем $\mathbf{v}$». Утверждению в кавычках можно придать точный смысл, потребовав, чтобы

$$\lim \frac{\|\phi(\mathbf{v})\|}{\|\mathbf{v}\|} = 0 \quad \text{при} \quad \|\mathbf{v}\| \to 0. \tag{5.2}$$

Подробнее: для любого значения $\varepsilon > 0$ существует такая величина $\delta > 0$, что

$$\|\phi(\mathbf{v})\| \le \varepsilon\|\mathbf{v}\| \tag{5.3}$$

для всех $\mathbf{v}$, удовлетворяющих условию $\|\mathbf{v}\| \le \delta$.

В формулах (5.2) и (5.3) выражение $\|\mathbf{v}\|$ обозначает длину вектора $\mathbf{v}$ в пространстве V, что следует отразить в обозначении, записав $\|\mathbf{v}\|_V$. Аналогично, $\|\phi(\mathbf{v})\|$ обозначает длину вектора $\phi(\mathbf{v})$ в пространстве W, что следует обозначить $\|\phi(\mathbf{v})\|_W$. Тогда неравенство (5.3) приобретет вид

$$\|\phi(\mathbf{v})\|_W \le \varepsilon\|\mathbf{v}\|_V.$$

Нижние индексы делают обозначения слишком громоздкими, поэтому мы не будем их использовать, а будем писать, как в формулах (5.2) и (5.3).

Например, пусть f и, следовательно, ϕ — отображения $\mathbb{R}^2$ на пространство $\mathbb{R}^1$. Наиболее общее выражение для вектора $\mathbf{v}$ в пространстве $\mathbb{R}^2$ имеет вид $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$. Для типографского упрощения запишем $\phi(\mathbf{v})$ как $\phi(x, y)$. Тогда $\|\mathbf{v}\| = (x^2 + y^2)^{1/2}$ и $\|\phi(\mathbf{v})\| = |\phi(x, y)|$. В этом случае условие (5.3) можно прочитать следующим образом:

> Для любого заданного значения $\varepsilon > 0$ существует такая величина $\delta > 0$, что неравенство
>
> $$|\phi(x, y)| < \varepsilon(x^2 + y^2)^{1/2}$$
>
> выполняется для всех x и y, удовлетворяющих условию
>
> $$(x^2 + y^2)^{1/2} < \delta.$$

Функцию $\phi : V \to W$, определенную в малой окрестности вокруг начала координат, удовлетворяющую условию (5.3), называют «"о" малое от $\mathbf{v}$». Символически мы говорим «ϕ есть $o(\mathbf{v})$» или, несколько злоупотребляя обозначениями, пишем $\phi = o(\mathbf{v})$. Теперь условие того, что A является производной функции f в точке $\mathbf{x}$, можно записать в виде

$$f(\mathbf{x} + \mathbf{v}) - f(\mathbf{x}) = A\mathbf{v} + \phi(\mathbf{v}),$$

где $\phi(\mathbf{v})$ есть $o(\mathbf{v})$, или $\phi(\mathbf{v}) = o(\mathbf{v})$, или более кратко:

$$f(\mathbf{x}+\mathbf{v}) - f(\mathbf{x}) = A\mathbf{v} + o(\mathbf{v}). \tag{5.4}$$

С точки зрения логики этот вариант записи слегка хромает, но из соображений удобства мы будем его часто использовать. Член $o(\mathbf{v})$ в формуле (5.4) на самом деле обозначает «некоторую функцию $\phi(\mathbf{v})$, являющуюся $o(\mathbf{v})$». Очень часто остаточный член ϕ нас не будет интересовать, нам достаточно будет знать, что он удовлетворяет условию (5.3). Именно поэтому удобно не вводить отдельный символ для каждой подобной функции ϕ.

Чтобы прочувствовать понятие $o(\mathbf{v})$, давайте докажем лемму: *Предположим, что $\phi : V \to W$ — линейное преобразование и что $\phi(\mathbf{v}) = o(\mathbf{v})$. Тогда*

$$\phi \equiv \mathbf{0}. \tag{5.5}$$

Доказательство. Ввиду линейности ϕ для любого вещественного числа r справедливо равенство $\phi(r\mathbf{v}) = r\phi(\mathbf{v})$. Для любого $\varepsilon > 0$ выберем такое δ, чтобы выполнялось неравенство $\|\phi(\mathbf{v})\| \leq \varepsilon\|\mathbf{v}\|$, когда $\|\mathbf{v}\| \leq \delta$. Теперь для произвольного вектора $\mathbf{w} \neq \mathbf{0}$ выберем $r = \|\mathbf{w}\|/\delta$ и запишем

$$\mathbf{w} = r\mathbf{w}'.$$

Тогда $\|\mathbf{w}'\| = \delta$, следовательно,

$$\|\phi(\mathbf{w})\| = r\|\phi(\mathbf{w}')\| \leq r\varepsilon\|\mathbf{w}'\| = \varepsilon r\delta = \varepsilon\|\mathbf{w}\|.$$

Итак, для каждого $\mathbf{w} \neq \mathbf{0}$

$$\|\phi(\mathbf{w})\| \leq \varepsilon\|\mathbf{w}\|$$

(очевидно, что ввиду линейности ϕ это справедливо и для $\mathbf{w} = \mathbf{0}$). Но это неравенство должно выполняться для всех ε. Следовательно, $\phi \equiv \mathbf{0}$.

Из леммы следует, что если справедливо равенство (5.4), то производная A, стоящая в нем, определена однозначно. Действительно, предположим, что

$$f(\mathbf{x}+\mathbf{v}) - f(\mathbf{x}) = A\mathbf{v} + \phi(\mathbf{v})$$

и

$$f(\mathbf{x}+\mathbf{v}) - f(\mathbf{x}) = A'\mathbf{v} + \phi'(\mathbf{v}),$$

где и ϕ и ϕ' являются $o(\mathbf{v})$. Тогда

$$(A' - A)\mathbf{v} = \phi(\mathbf{v}) - \phi'(\mathbf{v}).$$

Но мы утверждаем, что сумма или разность двух функций, каждая из которых есть $o(\mathbf{v})$, тоже является функцией $o(\mathbf{v})$. Действительно, для любого $\varepsilon > 0$ можно найти такие $\delta_1 > 0$ и $\delta_2 > 0$, чтобы выполнялись неравенства

$$\|\phi(\mathbf{v})\| \leq \frac{1}{2}\varepsilon\|\mathbf{v}\| \quad \text{для} \quad \|\mathbf{v}\| \leq \delta_1$$

и

$$\|\phi'(\mathbf{v})\| \leq \frac{1}{2}\varepsilon\|\mathbf{v}\| \quad \text{для} \quad \|\mathbf{v}\| \leq \delta_2.$$

Теперь, выбрав δ, равное наименьшему из двух чисел δ_1 и δ_2, из неравенства треугольника мы получаем при $\|\mathbf{v}\| < \delta$

$$\|\phi(\mathbf{v}) \pm \phi'(\mathbf{v})\| \leq \|\phi(\mathbf{v})\| + \|\phi'(\mathbf{v})\| \leq \varepsilon\|\mathbf{v}\|.$$

Мы получили, что линейное преобразование $A' - A$ есть $o(\mathbf{v})$ и, следовательно, является тождественно нулевым, т.е. $A = A'$.

Повторим, что отображение f, удовлетворяющее условию (5.4) для некоторого (единственного) A называется *дифференцируемым* в точке $\mathbf{x}$. В этом случае линейное преобразование A называется *дифференциалом* f в точке $\mathbf{x}$ и обозначается $df_{\mathbf{x}}$. Повторяем еще раз: дифференциал f в точке $\mathbf{x}$ есть единственное линейное отображение V на W, которое аппроксимирует истинное изменение отображения f в точке $\mathbf{x}$ для малых $\mathbf{v}$, что может быть записано как

$$f(\mathbf{x}+\mathbf{v}) - f(\mathbf{x}) = df_{\mathbf{x}}[\mathbf{v}] + o(\mathbf{v}).$$

Для того, чтобы доказать основные теоремы дифференциального исчисления, нам необходимо собрать сведения о функциях, которые попадают в категорию $o(\mathbf{v})$. Для этого удобно ввести следующие обозначения.

Подпространство S пространства V называется *окрестностью* нуля, если оно содержит некоторый шарик с центром в начале координат, т. е. если для некоторого $\delta > 0$ оно содержит множество всех $\mathbf{v}$, удовлетворяющих условию $\|\mathbf{v}\| \leq \delta$. Очевидно, что пересечение двух окрестностей тоже является окрестностью (просто возьмем меньший из двух шариков, он находится в пересечении). Точно так же можно говорить об окрестности любой точки $\mathbf{x}$. Это будет множество, которое содержит некоторый шарик вокруг точки $\mathbf{x}$, т. е. содержит множество $\{\mathbf{y} \mid \|\mathbf{y} - \mathbf{x}\| \leq \delta\}$.

Если A — обратимое линейное преобразование из V в W (V и W имеют одинаковую размерность), тогда можно найти константы k_1 и $k_2 > 0$, для которых выполняются неравенства

$$\|A\mathbf{v}\| \leq k_1\|\mathbf{v}\| \quad \text{и} \quad \|A^{-1}\mathbf{w}\| \leq k_2\|\mathbf{w}\|,$$

или, положив $\mathbf{w} = A\mathbf{v}$, неравенство

$$k_2^{-1}\|\mathbf{v}\| \leq \|A\mathbf{v}\|.$$

Таким образом, образ шарика с радиусом r содержится в шарике с радиусом $k_1 r$, внутри которого содержится шарик с радиусом $k_2^{-1}r$ (рис. 5.4). В частности, A преобразует окрестности в окрестности, этим же свойством обладает A^{-1}.

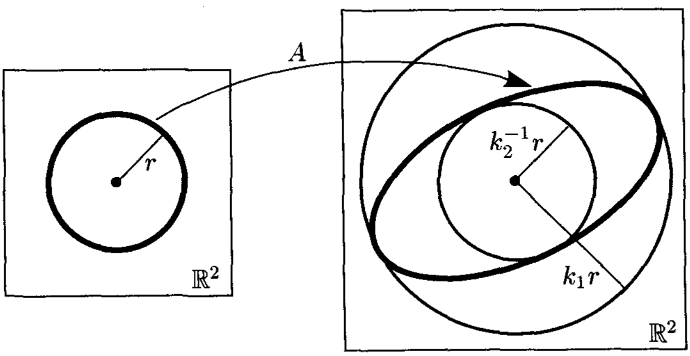


Рис. 5.4

Давайте рассмотрим общий случай, когда размерности V и W не совпадают. Введем обозначение $o(V, W)$ для пространства функций, попадающих в категорию $o(\mathbf{v})$, т. е. функция ϕ принадлежит $o(V, W)$, если $\phi = o(\mathbf{v})$. Более подробно: $\phi \in o(V, W)$, если ϕ определена в некоторой окрестности вокруг начала координат и удовлетворяет условию (5.3).

Функция ψ принадлежит категории $O(\mathbf{v})$ (следует читать «ψ есть "O" большое от $\mathbf{v}$»), если она определена в некоторой окрестности $\mathbf{0}$ и существует такая постоянная $k > 0$, для которой неравенство

$$\|\psi(\mathbf{v})\| \leq k\|\mathbf{v}\|$$

справедливо для любого $\mathbf{v}$ в этой окрестности. Например, линейное отображение автоматически является $O(\mathbf{v})$. Очевидно, что любая функция, попадающая в категорию $o(\mathbf{v})$, заведомо является $O(\mathbf{v})$. Пусть $O(V, W)$ обозначает пространство всех функций, попадающих в категорию $O(\mathbf{v})$, и пусть $I(V, W)$ обозначает пространство функций, определенных вблизи $\mathbf{0}$ и стремящихся к 0 при $\mathbf{v} \to \mathbf{0}$. Иначе говоря,

> $\chi \in I(V, W)$, если χ определена в некоторой окрестности вокруг начала координат и для каждого значения $\varepsilon > 0$ существует такое значение $\delta > 0$, что

$$\|\chi(\mathbf{v})\| < \varepsilon, \quad \text{когда} \quad \|\mathbf{v}\| < \delta.$$

Очевидно, что

$$o(V, W) \subset O(V, W) \subset I(V, W).$$

Если, например, мы возьмем $V = W = \mathbb{R}^1$ и определим $\phi(x) = x^2$, $\psi(x) = x$ и $\chi(x) = |x|^{1/2}$, то

$$\phi \in o(V, W),$$

$$\psi \in O(V, W), \quad \text{но} \quad \psi \notin o(V, W)$$

и

$$\chi \in I(V, W), \quad \text{но} \quad \chi \notin O(V, W).$$

Следовательно, выписанные выше включения — строгие.

Выше мы доказали, что сумма двух функций, являющихся $o(V, W)$, по-прежнему попадает в категорию $o(V, W)$. Аналогично доказывается, что сумма двух функций, являющихся $O(V, W)$, также есть $O(V, W)$. То же — для $I(V, W)$.

Рассмотрим поведение этих пространств при композиции. Пусть X обозначает третье пространство. Докажем следующие утверждения.

$$\text{Если } \psi_1 \in O(V, W) \text{ и } \psi_2 \in O(W, X), \text{ то } \psi_2 \circ \psi_1 \in O(V, X). \quad (5.6)$$

$$\text{Если } \psi_1 \in O(V, W) \text{ и } \psi_2 \in o(W, X), \text{ то } \psi_2 \circ \psi_1 \in o(V, X). \quad (5.7)$$

$$\text{Если } \psi_1 \in o(V, W) \text{ и } \psi_2 \in O(W, X), \text{ то } \psi_2 \circ \psi_1 \in o(V, X). \quad (5.8)$$

Доказательство. Если $\|\psi_1(\mathbf{v})\| \leq k_1\|\mathbf{v}\|$ для $\|\mathbf{v}\| \leq \delta_1$ и $\|\psi_2(\mathbf{w})\| \leq k_2\|\mathbf{w}\|$ для $\|\mathbf{w}\| \leq \delta_2$, то композиция $\psi_2 \circ \psi_1$ будет определена для $\|\mathbf{v}\| \leq \delta$, где δ есть меньший из двух параметров δ_1 и δ_2/k_1. Для таких $\mathbf{v}$ мы имеем

$$\|\psi_2 \circ \psi_1(\mathbf{v})\| = \|\psi_2(\psi_1(\mathbf{v}))\| \leq k_2\|\psi_1(\mathbf{v})\| \leq k_2k_1\|\mathbf{v}\|,$$

что доказывает (5.6). Если $\psi_2 \in o(W, X)$, мы можем сделать k_2 сколь угодно малым, выбирая δ_2 (и, следовательно, δ) малым. Это доказывает утверждение (5.7). Если $\psi_1 \in o(V, W)$, то мы можем сделать сколь угодно малым значение k_1, выбирая δ_1 (и, следовательно, δ) достаточно малым, что доказывает (5.8).

Если ϕ — отображение из V в W и $g : V \to \mathbb{R}$ — функция с вещественными значениями, то произведение $g(\mathbf{v})\phi(\mathbf{v})$ имеет смысл только для тех $\mathbf{v}$, которые лежат в пересечении областей определения ϕ и g. Так получается функция $g\phi$, которая отображает V на W.

$$\text{Если } \psi \in O(V, W) \quad \text{и} \quad g \in I(V, \mathbb{R}), \quad \text{то} \quad g\psi \in o(V, W). \quad (5.9)$$

Доказательство. Мы знаем, что существует такое значение k, для которого $\|\psi(\mathbf{v})\| \leq k\|\mathbf{v}\|$ в некоторой окрестности вокруг начала координат. Задав произвольное значение $\varepsilon > 0$, выберем δ настолько малым, чтобы $\|g(\mathbf{v})\| \leq \varepsilon/k$ для всех $\mathbf{v}$ при условии $\|\mathbf{v}\| \leq \delta$. Тогда для таких $\mathbf{v}$

$$\|g(\mathbf{v})\psi(\mathbf{v})\| \leq (\varepsilon/k)\|\psi(\mathbf{v})\| \leq \varepsilon\|\mathbf{v}\|,$$

что и доказывает утверждение (5.9). Аналогично доказываются следующие утверждения.

$$\text{Если } \psi \in I(V,W) \text{ и } g \in O(V,\mathbb{R}), \text{ то } g\psi \in o(V,W). \tag{5.10}$$

Если $\psi \in o(V,W)$ и g — ограниченная функция из $V \to \mathbb{R}$, то

$$g\psi \in o(V,W). \tag{5.11}$$

Если отображение ψ из V в W, определенное в некоторой окрестности вокруг начала координат, ограничено и $g \in o(V,\mathbb{R})$, то

$$g\psi \in o(V,W). \tag{5.12}$$

(Когда мы говорим, что отображение $\psi : V \to W$ *ограничено*, это значит, что существует некоторое положительное вещественное число k, такое, что $\|\psi(\mathbf{v})\|_W \leq k$ для всех $\mathbf{v}$ из V.)

Итак, мы собрали все необходимые леммы, чтобы приступить к изучению дифференциального исчисления.

5.2. Дифференциальное исчисление

Пусть $f : V \to W$ определена в некоторой окрестности точки $\mathbf{x} \in V$. Определение функции $\nabla_{\mathbf{x}} f$ дается формулой

$$\nabla_{\mathbf{x}} f(\mathbf{h}) = f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}).$$

Она определена для всех значений $\mathbf{h}$ в некоторой окрестности $\mathbf{0}$ и показывает изменение f относительно ее значения в точке $\mathbf{x}$. Функция f *непрерывна* в точке $\mathbf{x}$, если $\nabla_{\mathbf{x}} f \in I(V,W)$. (Это значит, что $\nabla_{\mathbf{x}} f(\mathbf{h})$ стремится к $\mathbf{0}$ при $\mathbf{h} \to \mathbf{0}$, так что $f(\mathbf{x}+\mathbf{h}) \to f(\mathbf{x})$.) Напоминаем, что функция f дифференцируема в точке $\mathbf{x}$, если существует линейное преобразование $df_{\mathbf{x}} : V \to W$ такое, что

$$\nabla_{\mathbf{x}} f(\mathbf{h}) = df_{\mathbf{x}}[\mathbf{h}] + o(\mathbf{h}).$$

Это уравнение единственным образом определяет линейное преобразование $df_{\mathbf{x}}$, называемое *дифференциалом* f в точке $\mathbf{x}$. Любая линейная функция относится к категории $O(V,W)$; сумма функции типа $O(V,W)$ и функции типа $o(V,W)$ тоже относится к $O(V,W)$. Отсюда можно сделать вывод:

$$\text{если } f \text{ дифференцируема в точке } \mathbf{x}, \text{ то } \ \nabla_{\mathbf{x}} f \in O(V,W). \quad (5.13)$$

(В частности, из того, что $O(V,W) \subset I(V,W)$, следует, что если функция f дифференцируема в $\mathbf{x}$, тогда она обязательно непрерывна в точке $\mathbf{x}$.)

Если f — линейная функция, т. е. $f(\mathbf{x}) = A\mathbf{x}$, то $\nabla_{\mathbf{x}} f[\mathbf{h}] = A(\mathbf{x}+\mathbf{h}) - A\mathbf{x} = A\mathbf{h}$. Следовательно, линейная функция $f(\mathbf{x}) = A\mathbf{x}$ дифференцируема во всех точках. Ее дифференциал дается формулой $df_{\mathbf{x}} = A$ и не зависит от $\mathbf{x}$.

Если f — постоянная функция, то $\nabla_{\mathbf{x}} f \equiv \mathbf{0}$ и формула (5.4) справедлива при $A = 0$. Следовательно, постоянная функция дифференцируема везде, и ее производная тождественно равна нулю.

Теперь мы можем сформулировать и доказать правило вычисления дифференциала суммы двух функций.

Если f и g — две функции из V на W, и обе функции дифференцируемы в точке $\mathbf{x}$, то дифференциал суммы равен

$$d(f+g)_{\mathbf{x}} = df_{\mathbf{x}} + dg_{\mathbf{x}}. \quad (5.14)$$

Доказательство. Очевидно, что $\nabla_{\mathbf{x}}(f+g) = \nabla_{\mathbf{x}} f + \nabla_{\mathbf{x}} g$. Из того, что

$$\nabla_{\mathbf{x}} f = df_{\mathbf{x}} + \phi_1 \quad \text{и} \quad \nabla_{\mathbf{x}}\, g = dg_{\mathbf{x}} + \phi_2,$$

где ϕ_1 и ϕ_2 — функции типа $o(V,W)$, следует, что

$$\nabla_{\mathbf{x}}(f+g) = df_{\mathbf{x}} + dg_{\mathbf{x}} + \phi_1 + \phi_2.$$

Так как $(\phi_1 + \phi_2) \in o(V,W)$, формула (5.14) доказана.

Если мы умножим функцию g, значения которой принадлежат $\mathbb{R}$, на функцию f, значения которой принадлежат W, то получим функцию, значения которой принадлежат W. Сформулируем правило вычисления производной произведения функций. Предположим, что $f : V \to W$ и $g : V \to \mathbb{R}$ дифференцируемы в точке $\mathbf{x}$. Тогда произведение gf тоже дифференцируемо в точке $\mathbf{x}$ и

$$d(gf)_{\mathbf{x}}[\mathbf{h}] = g(\mathbf{x})df_{\mathbf{x}}[\mathbf{h}] + (dg_{\mathbf{x}}[\mathbf{h}])f(\mathbf{x}).$$

Доказательство.

$$\begin{aligned}
\nabla_{\mathbf{x}}(gf)[\mathbf{h}] &= g(\mathbf{x}+\mathbf{h})f(\mathbf{x}+\mathbf{h}) - g(\mathbf{x})f(\mathbf{x}) \\
&= g(\mathbf{x}+\mathbf{h})(f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})) + (g(\mathbf{x}+\mathbf{h}) - g(\mathbf{x}))f(\mathbf{x}) \\
&= g(\mathbf{x})(f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})) + (g(\mathbf{x}+\mathbf{h}) - g(\mathbf{x}))f(\mathbf{x}) \\
&\qquad + (g(\mathbf{x}+\mathbf{h}) - g(\mathbf{x}))(f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})) \\
&= g(\mathbf{x})\ \nabla_{\mathbf{x}}\, f[\mathbf{h}] + (\nabla_{\mathbf{x}} g[\mathbf{h}])f(\mathbf{x}) + (\nabla_{\mathbf{x}} g[\mathbf{h}])(\nabla_{\mathbf{x}} f[\mathbf{h}]) \\
&= g(\mathbf{x})(df_{\mathbf{x}}[\mathbf{h}]+o(\mathbf{h})) + (dg_{\mathbf{x}}[\mathbf{h}])+o(\mathbf{h}))f(\mathbf{x})+O(\mathbf{h})o(\mathbf{h}),
\end{aligned}$$

так как f и g дифференцируемы в точке $\mathbf{x}$ и, следовательно, оба дифференциала $\nabla_{\mathbf{x}} f$ и $\nabla_{\mathbf{x}} g$ суть $o(\mathbf{h})$ из (5.13). Из (5.9) следует, что произведение двух функций типа $o(\mathbf{h})$ есть функция типа $o(\mathbf{h})$. Обе функции f и g ограничены вблизи $\mathbf{x}$, т. е. $g(\mathbf{x}+\mathbf{h})-g(\mathbf{x})$ и $f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})$ стремятся к нулю. Произведение ограниченной функции и функции типа $o(\mathbf{h})$ есть тоже функция типа $o(\mathbf{h})$. Используя это в последнем равенстве, получаем

$$\nabla_{\mathbf{x}}(gf)[\mathbf{h}] = g(\mathbf{x})df_{\mathbf{x}}[\mathbf{h}] + (dg_{\mathbf{x}}[\mathbf{h}])f(\mathbf{x}) + o(\mathbf{h}),$$

что и требовалось доказать.

А теперь сформулируем очень важную теорему.

Цепное правило. Предположим, что $f : V \to W$ дифференцируема в точке $\mathbf{x} \in V$ и $g : W \to X$ дифференцируема в точке $\mathbf{y} = f(\mathbf{x}) \in W$. Тогда $g \circ f : V \to X$ дифференцируема в точке $\mathbf{x}$ и ее дифференциал определяется формулой:

$$\boxed{d(g \circ f)_{\mathbf{x}} = (dg_{f(\mathbf{x})}) \cdot (df_{\mathbf{x}}).} \tag{5.15}$$

(В правой части этого выражения стоит композиция двух линейных преобразований: $dg_{f(\mathbf{x})} : W \to X$ и $df_{\mathbf{x}} : V \to W$. В левой части стоит композиция g и f.)

Доказательство.

$$\begin{aligned}
\nabla_{\mathbf{x}}(g \circ f)[\mathbf{h}] &= g(f(\mathbf{x}+\mathbf{h})) - g(f(\mathbf{x})) \\
&= g(f(\mathbf{x}) + \nabla_{\mathbf{x}} f[\mathbf{h}]) - g(f(\mathbf{x})) \\
&= \nabla_{f(\mathbf{x})} g[\nabla_{\mathbf{x}} f[\mathbf{h}]] \\
&= dg_{f(\mathbf{x})}[df_{\mathbf{x}}[\mathbf{h}]] + dg_{f(\mathbf{x})}[o[\mathbf{h}]] + (\phi \circ \psi)(\mathbf{h}),
\end{aligned}$$

где $\phi \in o(V, X)$ (происходит от остаточного члена в $\nabla_{f(\mathbf{x})} g$), а $\psi = \nabla_{\mathbf{x}} f \in O(V, W)$, что следует из (5.13). Согласно (5.8), эта композиция относится к категории $o(V, X)$. Кроме того, $dg_{f(\mathbf{x})}$ линейна и, следовательно, относится к категории $O(W, X)$. Таким образом, второе слагаемое есть композиция элемента типа $O(W, X)$ и элемента типа $o(V, W)$. Согласно (5.7) эта композиция имеет тип $o(V, W)$. Поэтому

$$\nabla_{\mathbf{x}}(g \circ f)[\mathbf{h}] = (dg_{f(\mathbf{x})} \circ df_{\mathbf{x}})[\mathbf{h}] + o(\mathbf{h}),$$

что и требовалось доказать.

Примеры

Приведем несколько примеров на вычисление дифференциалов и на цепное правило. Для функций $\alpha : \mathbb{R}^1 \to \mathbb{R}^1$ дифференциал $d\alpha_x$, вычисленный для $h \in \mathbb{R}$, определяется умножением на производную $\alpha'(x)$, т. е.

$$d\alpha_x[h] = \alpha'(x)h.$$

Это выражение является определением производной $\alpha'(x)$. Например, пусть $\alpha : \mathbb{R}^1 \to \mathbb{R}^1$ и $\beta : \mathbb{R}^1 \to \mathbb{R}^1$ имеют вид

$$\alpha(y) = y^2, \quad \beta(x) = 5x^3 + 1,$$

так что

$$\alpha \circ \beta(x) = (5x^3 + 1)^2.$$

Тогда

$d\alpha_y$ соответствует умножению на $2y$,
$d\beta_x$ соответствует умножению на $15x^2$,
$d(\alpha \circ \beta)_x$ соответствует умножению на $2(5x^3 + 1)(15x^2)$.

В соответствии с цепным правилом $d\alpha_{\beta(x)} \circ d\beta_x = d(\alpha \circ \beta)_x$. Поэтому мы имеем:

$d\alpha_{\beta(x)}$ соответствует умножению на $2(5x^3 + 1)$,
$d\alpha_{\beta(x)} \circ d\beta_x$ соответствует умножению на $15x^2$, а потом на $2(5x^3 + 1)$, или
$d\alpha_{\beta(x)} \circ d\beta_x$ соответствует умножению на $2(5x^3 + 1)(15x^2)$.

Очевидно, что эти обозначения довольно громоздкие. Обозначения Лейбница для функций одной переменной лучше.

Если α — функция от y, то ее производная записывается как

$$\alpha' = \frac{d\alpha}{dy} \quad \text{или} \quad d\alpha = \alpha'\, dy.$$

Последнее равенство показывает, что для каждого значения y

$$d\alpha_y(h) = \alpha'(y)h.$$

Другими словами, dy является фиктивным символом, вместо которого мы подставляем значение h. Тогда

$$d(y^2) = 2y\,dy, \quad d(5x^3+1) = 15x^2\,dx.$$

Согласно цепному правилу, подставляем

$$y = (5x^3+1) \quad \text{и} \quad dy = 15x^2\,dx$$

в формулу для $d(y^2)$ и получаем выражение для $d[(5x^3+1)^2]$. В обозначениях Лейбница цепное правило становится механической подстановкой.

Мы, тем не менее, продолжим вычисление других примеров в наших, более громоздких обозначениях, надеясь прояснить смысл выполняемых операций.

Пусть $f : \mathbb{R}^1 \to \mathbb{R}^2$ и $g : \mathbb{R}^2 \to \mathbb{R}^1$ определены формулами

$$f(x) = \begin{pmatrix} x^2+1 \\ 2x-1 \end{pmatrix}, \quad g\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = x^2 y.$$

Чтобы вычислить df_x, заметим, что

$$\begin{aligned} \nabla_x f[s] &= f(x+s) - f(x) \\ &= \begin{pmatrix} 2xs \\ 2s \end{pmatrix} + \begin{pmatrix} s^2 \\ 0 \end{pmatrix} \\ &= s\begin{pmatrix} 2x \\ 2 \end{pmatrix} + o(s). \end{aligned}$$

Таким образом, df_x представляется матрицей $df_x = \begin{pmatrix} 2x \\ 2 \end{pmatrix}$.

Аналогично,

$$\begin{aligned}
\nabla_{\binom{x}{y}} g\left[\begin{pmatrix} s \\ t \end{pmatrix}\right] &= (x+s)^2(y+t) - x^2y \\
&= x^2t + 2sxy + 2sxt + s^2y + s^2t \\
&= (2xy, x^2)\begin{pmatrix} s \\ t \end{pmatrix} + o\left(\begin{pmatrix} s \\ t \end{pmatrix}\right),
\end{aligned}$$

поэтому матрица для $dg_{\binom{x}{y}}$ имеет вид

$$dg_{\binom{x}{y}} = (2xy, x^2).$$

Композиция $g \circ f : \mathbb{R}^1 \to \mathbb{R}^1$ дается формулой

$$g \circ f(x) = (x^2+1)^2(2x-1),$$

так что

$$d(g \circ f)_x = 2(x^2+1)(2x)(2x-1) + 2(x^2+1)^2.$$

Согласно цепному правилу, это выражение должно быть равно матричному произведению $dg_{f(x)} \circ df_x$. Действительно,

$$\begin{aligned}
dg_{f(x)} \circ df_x &= (2(x^2+1)(2x-1),\ (x^2+1)^2)\begin{pmatrix} 2x \\ 2 \end{pmatrix} \\
&= 2(x^2+1)(2x-1)(2x) + 2(x^2+1)^2,
\end{aligned}$$

а это равняется $d(g \circ f)_x$.

Мы можем также построить композицию этих функций в другом порядке, $f \circ g : \mathbb{R}^2 \to \mathbb{R}^2$, определяемую формулой

$$f \circ g\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} (x^2y)^2 + 1 \\ 2(x^2y) - 1 \end{pmatrix}.$$

Подробно вычислим $d(f \circ g)_{\binom{x}{y}}$:

$$\nabla_{\binom{x}{y}}(f \circ g)\left(\begin{pmatrix} s \\ t \end{pmatrix}\right) = \begin{pmatrix} (x+s)^4(y+t)^2 + 1 \\ 2(x+s)^2(y+t) - 1 \end{pmatrix} - \begin{pmatrix} x^4y^2 + 1 \\ 2x^2y - 1 \end{pmatrix}$$

$$
\begin{aligned}
&= \begin{pmatrix} (x^4+4x^3s+6x^2s^2+4xs^3+s^4)(y^2+2yt+t^2)-x^4y^2 \\ 2(x^2+2xs+s^2)(y+t)-2x^2y \end{pmatrix} \\
&= \begin{pmatrix} 2x^4yt+4x^3y^2s \\ 2x^2t+4xys \end{pmatrix} \\
&\quad + \begin{pmatrix} x^4t^2+4x^3s(2yt+t^2)+(6x^2s^2+4xs^3+s^4)(y+t)^2 \\ 4xst+s^2y+s^2t \end{pmatrix} \\
&= \begin{pmatrix} 4x^3y^2 & 2x^4y \\ 4xy & 2x^2 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} + o\left(\begin{pmatrix} s \\ t \end{pmatrix}\right).
\end{aligned}
$$

Следовательно,

$$
d(f\circ g)_{\binom{x}{y}} = \begin{pmatrix} 4x^3y^2 & 2x^4y \\ 4xy & 2x^2 \end{pmatrix}.
$$

Согласно цепному правилу, это должно быть равно $df_{g(\binom{x}{y})} \circ dg_{\binom{x}{y}}$, определяемому формулой

$$
\begin{aligned}
df_{g(\binom{x}{y})} \circ dg_{\binom{x}{y}} &= \begin{pmatrix} 2x^2y \\ 2 \end{pmatrix} (2xy,\, x^2) \\
&= \begin{pmatrix} 4x^3y^2 & 2x^4y \\ 4xy & 2x^2 \end{pmatrix}.
\end{aligned}
$$

Как и следовало ожидать, мы получили ту же матрицу.

Еще один пример на цепное правило. Пусть $F:\mathbb{R}^2\to\mathbb{R}^2$ и $G:\mathbb{R}^2\to\mathbb{R}^2$ заданы формулой

$$
F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x^2+y \\ xy \end{pmatrix}, \qquad G\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} 3xy^2 \\ x^2 \end{pmatrix}.
$$

Вычислим дифференциалы отображений F и G в отдельности. Для отображения F имеем

$$
\begin{aligned}
\nabla_{\binom{x}{y}} F\left(\begin{pmatrix} s \\ t \end{pmatrix}\right) &= \begin{pmatrix} (x+s)^2+(y+t) \\ (x+s)(y+t) \end{pmatrix} - \begin{pmatrix} x^2+y \\ xy \end{pmatrix} \\
&= \begin{pmatrix} 2xs+t \\ xt+sy \end{pmatrix} + \begin{pmatrix} s^2 \\ st \end{pmatrix} \\
&= \begin{pmatrix} 2x & 1 \\ y & x \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} + o\left(\begin{pmatrix} s \\ t \end{pmatrix}\right),
\end{aligned}
$$

откуда

$$dF_{\binom{x}{y}} = \begin{pmatrix} 2x & 1 \\ y & x \end{pmatrix}.$$

Аналогично, для отображения G имеем

$$\begin{aligned}\nabla_{\binom{x}{y}} G\left(\begin{pmatrix} s \\ t \end{pmatrix}\right) &= \begin{pmatrix} 3(x+s)(y+t)^2 \\ (x+s)^2 \end{pmatrix} - \begin{pmatrix} 3xy^2 \\ x^2 \end{pmatrix} \\ &= \begin{pmatrix} 3sy^2 + 3xy(2t) \\ 2xs \end{pmatrix} + \begin{pmatrix} 3xt^2 + 3s(2yt + t^2) \\ s^2 \end{pmatrix} \\ &= \begin{pmatrix} 3y^2 & 6xy \\ 2x & 0 \end{pmatrix}\begin{pmatrix} s \\ t \end{pmatrix} + o\left(\begin{pmatrix} s \\ t \end{pmatrix}\right),\end{aligned}$$

откуда

$$dG_{\binom{x}{y}} = \begin{pmatrix} 3y^2 & 6xy \\ 2x & 0 \end{pmatrix}.$$

Композиция $F \circ G : \mathbb{R}^2 \to \mathbb{R}^2$ задается формулой

$$\begin{aligned}F \circ G\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = F\left(\begin{pmatrix} 3xy^2 \\ x^2 \end{pmatrix}\right) &= \begin{pmatrix} (3xy^2)^2 + x^2 \\ 3xy^2x^2 \end{pmatrix} \\ &= \begin{pmatrix} 9x^2y^4 + x^2 \\ 3x^3y^2 \end{pmatrix}.\end{aligned}$$

Найдем дифференциал композиции $F \circ G$.

$$\begin{aligned}&\nabla_{\binom{x}{y}}(F \circ G)\left(\begin{pmatrix} s \\ t \end{pmatrix}\right) \\ &= \begin{pmatrix} 9(x+s)^2(y+t)^4 + (x+s)^2 \\ 3(x+s)^3(y+t)^2 \end{pmatrix} - \begin{pmatrix} 9x^2y^4 + x^2 \\ 3x^3y^2 \end{pmatrix} \\ &= \begin{pmatrix} 9(2xs)y^4 + 9x^2(4y^3t) + 2xs \\ 3(3x^2s)y^2 + 3x^3(2yt) \end{pmatrix} \\ &\quad + \begin{pmatrix} 9x^2(6y^2t^2 + 4yt^3 + 4t) + 18xs[(y+1)^4 - y^4] + 9s^2(y+t)^4 \\ 3x^3t^2 + 3(3xs^2 + s^3)(y+t)^2 + 9x^2s(2yt + t^2) \end{pmatrix} \\ &= \begin{pmatrix} 18xy^4 + 2x & 36x^2y^3 \\ 9x^2y^2 & 6x^3y \end{pmatrix}\begin{pmatrix} s \\ t \end{pmatrix} + o\left(\begin{pmatrix} s \\ t \end{pmatrix}\right).\end{aligned}$$

Отсюда

$$d(F \circ G)_{\binom{x}{y}} = \begin{pmatrix} 18xy^4 + 2x & 36x^2y^3 \\ 9x^2y^2 & 6x^3y \end{pmatrix}.$$

Согласно цепному правилу, это должно быть равно $dF_{G(\binom{x}{y})} \circ dG_{\binom{x}{y}}$. Проверяем:

$$\begin{aligned} dF_{G(\binom{x}{y})} \circ dG_{\binom{x}{y}} &= \begin{pmatrix} 2(3xy^2) & 1 \\ x^2 & 3xy^2 \end{pmatrix} \begin{pmatrix} 3y^2 & 6xy \\ 2x & 0 \end{pmatrix} \\ &= \begin{pmatrix} 18xy^4 + 2x & 36x^2y^3 \\ 3x^2y^2 + 6x^2y^2 & 6x^3y \end{pmatrix} \\ &= d(F \circ G)_{\binom{x}{y}}. \end{aligned}$$

В следующих параграфах мы займемся некоторыми важными следствиями цепного правила. Но сначала рассмотрим еще несколько более «абстрактных» примеров на применение цепного правила и введем несколько обозначений.

5.3. Примеры на цепное правило

Рассмотрим отображение умножения $g : \mathbb{R}^2 \to \mathbb{R}^1$, заданное формулой

$$g\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = xy.$$

Если $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$ и $\mathbf{h} = \begin{pmatrix} r \\ s \end{pmatrix}$, то

$$\begin{aligned} g(\mathbf{v} + \mathbf{h}) = (x + r)(y + s) &= xy + xs + yr + rs \\ &= g(\mathbf{v}) + xs + yr + o(\mathbf{h}). \end{aligned}$$

Откуда

$$dg_{\mathbf{v}}(\mathbf{h}) = xs + yr,$$

а соответствующая матрица (с одной строкой и двумя столбцами) равна

$$(y,\, x).$$

Пусть задано отображение $f : \mathbb{R}^1 \to \mathbb{R}^2$. Его можно изобразить как кривую на плоскости или, что проще, представить как две вещественных функции от одной вещественной переменной, т. е.

$$f(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}.$$

Тогда

$$f(t+h) = \begin{pmatrix} x(t+h) \\ y(t+h) \end{pmatrix} = \begin{pmatrix} x(t) + x'(t)h + o(h) \\ y(t) + y'(t)h + o(h) \end{pmatrix},$$

следовательно,

$$f(t+h) - f(t) = h \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix} + o(h),$$

или

$$f'(t) = \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix}.$$

Умножение матрицы

$$dg_{f(t)} = (y(t), x(t)) \quad \text{на} \quad df_t = f'(t) = \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix}$$

дает

$$d(g \circ f)_t = (d \circ f)'(t) = x'(t)y(t) + x(t)y'(t).$$

Но $(g \circ f)(t) = x(t)y(t)$. Таким образом, в этом случае цепное правило можно истолковать как применение формулы Лейбница для производной от произведения двух функций.

Прежде чем двигаться дальше, удобно ввести некоторые новые обозначения. Вместо того, чтобы писать

$$dg_{\mathbf{v}}(\mathbf{h}) = yr + xs, \quad \text{где} \quad \mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \mathbf{h} = \begin{pmatrix} r \\ s \end{pmatrix}, \quad g\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = xy,$$

удобно всю информацию заключить в формулу

$$d(xy) = y\,dx + x\,dy.$$

В этом уравнении символ dx в правой части следует понимать как линейное отображение $\mathbb{R}^2 \to \mathbb{R}^1$: отображение, приписывающее каждому вектору его первую координату. Итак,

$$dx(\mathbf{h}) = r, \quad \text{если} \quad \mathbf{h} = \begin{pmatrix} r \\ s \end{pmatrix}$$

и аналогично

$$dy(\mathbf{h}) = s.$$

В выражении $y\,dx$ символ y — это функция вектора $\mathbf{v}$, приписывающая вектору $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$ его вторую координату. Следовательно, члены типа $y\,dx$ зависят от переменных двух типов, от переменной $\mathbf{v}$, говорящей нам, где вычисляется производная, и от переменной $\mathbf{h}$, являющейся мерой небольшого отклонения. Стоящее в левой части формулы выражение $d(xy)$ есть краткая запись $dg_{\mathbf{v}}$, где функция g определяется условием $g(\mathbf{v}) = xy$ для вектора $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$. Используя цепное правило по аналогии с первым примером, мы говорим:

> Рассмотрим x как функцию[1] в пространстве $\mathbb{R}^2$, которая каждому вектору приписывает его первую координату. Тогда, по определению нашего отображения f, можно написать $(x \circ f)(t) = x(t)$. Согласно цепному правилу, $d((x \circ f)_t = x'(t)\,dt$, функция $x'(t)$ вычисляется в той точке t, где требуется найти дифференциал, dt определяет малое приращение. Следовательно, когда мы рассматриваем x как функцию t, заданную отображением f, мы производим «замену» $dx = x'\,dt$, где x' зависит от t. Аналогично, если мы рассматриваем y как функцию от t, задаваемую отображением f, цепное правило приписывает нам сделать подстановку $dy = y'dt$.

Тогда можно написать

$$d(g \circ f) = yx'\,dt + xy'\,dt,$$

где справа величины x, y, x' и y' следует заменить явными функциями от t.

Например, положим $x(t) = t + \sin t, \quad y(t) = e^{2t}$. Тогда

$$dg = d(xy) = y\,dx + x\,dy, \quad df = d\begin{pmatrix} t + \sin t \\ e^{2t} \end{pmatrix} = \begin{pmatrix} 1 + \cos t \\ 2e^{2t} \end{pmatrix} dt$$

[1] Вероятно, полезно еще раз прочитать то место в параграфе 1.3, где подробно обсуждается, как координату x можно рассматривать в качестве функции.

и

$$d(g \circ f) = d((t + \sin t)(e^{2t})) = (e^{2t}(1 + \cos t) + 2e^{2t}(t + \sin t))\, dt.$$

Давайте изобразим цепное правило на диаграмме (рис. 5.5). Нам даны два дифференцируемых отображения $f : V \to W$ и $g : W \to Z$, из которых можно составить композицию $g \circ f : V \to Z$. На некоторую точку $\mathbf{v}$ в пространстве V действуем отображением f и получаем $f(\mathbf{v})$, потом действуем отображением g и получаем $g(f(\mathbf{v}))$. При вычислении дифференциала $d(g \circ f)_{\mathbf{v}}(\mathbf{h})$ мы можем двигаться последовательно, сначала подействовав линейным отображением $df_{\mathbf{v}}$ на вектор $\mathbf{h}$ и затем подействовав линейным отображением $dg_{f(\mathbf{v})}$ на полученный образ.

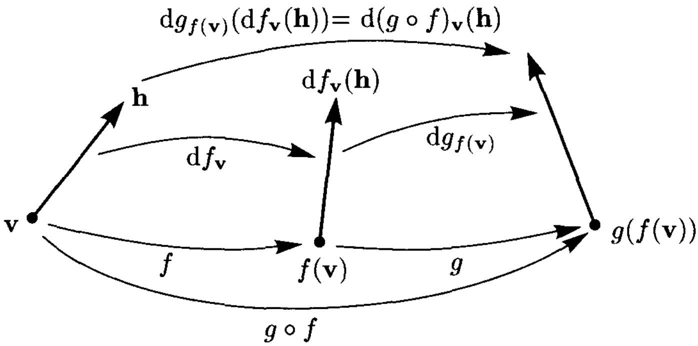


Рис. 5.5

А теперь давайте выполним более сложные вычисления с цепным правилом. Рассмотрим векторные пространства V, W и т. д. большого числа измерений. Конечно, формально говоря, нам бы следовало отложить эти вычисления, пока не изучен соответствующий раздел линейной алгебры. Но мы, тем не менее, рекомендуем разобрать их сейчас. Начнем с вычисления производной от произведения двух матриц. Пусть V есть векторное пространство, образованное парами матриц $n \times n$. Вектор в таком пространстве имеет вид

$$\mathbf{v} = \begin{pmatrix} A \\ B \end{pmatrix},$$

где A и B — матрицы размера $n \times n$. Векторы в этом пространстве складываются и умножаются на скаляр покомпонентно.

$$\text{Если } \mathbf{v} = \begin{pmatrix} A \\ B \end{pmatrix} \text{ и } \mathbf{v}' = \begin{pmatrix} A' \\ B' \end{pmatrix},$$

$$\text{то } \mathbf{v} + \mathbf{v}' = \begin{pmatrix} A + A' \\ B + B' \end{pmatrix} \text{ и } a\mathbf{v} = \begin{pmatrix} aA \\ aB \end{pmatrix}.$$

Очевидно, что пространство V имеет размерность $2n^2$.

Пусть W обозначает другое векторное пространство матриц размером $n \times n$. Определим отображение $g : V \to W$ формулой $g\left(\begin{pmatrix} A \\ B \end{pmatrix}\right) = AB$.

Если $\mathbf{v} = \begin{pmatrix} A \\ B \end{pmatrix}$ и $\mathbf{h} = \begin{pmatrix} X \\ Y \end{pmatrix}$, то

$$\begin{aligned} g(\mathbf{v} + \mathbf{h}) - g(\mathbf{v}) &= (A + X)(B + Y) - AB \\ &= XB + AY + XY = XB + AY + o(\mathbf{h}). \end{aligned}$$

Следовательно,

$$dg_{\mathbf{v}}(\mathbf{h}) = XB + AY. \tag{5.16}$$

В процессе вычислений будем для удобства пользоваться обозначениями, где опущены индекс $\mathbf{v}$ и аргумент $\mathbf{h}$. Тогда уравнение (5.16) имеет вид

$$d(AB) = (dA)B + A(dB). \tag{5.17}$$

В этих обозначениях величина AB в левой части уравнения — это удобное обозначение функции g. Величина dA справа есть производная функции, которая вектору $\begin{pmatrix} A \\ B \end{pmatrix}$ ставит в соответствие матрицу A. Эта производная, вычисленная в точке $\begin{pmatrix} A \\ B \end{pmatrix}$, для вектора $\begin{pmatrix} X \\ Y \end{pmatrix}$ дает значение X. Таким образом, dA — линейное отображение, которое каждому вектору $\begin{pmatrix} X \\ Y \end{pmatrix}$ ставит в соответствие величину X. Поэтому, например, $(dA)B$ есть линейное отображение, которое вектору $\begin{pmatrix} X \\ Y \end{pmatrix}$ ставит в соответствие значение XB. В этом смысле уравнение (5.17) есть краткая запись уравнения (5.16).

Далее, пусть f обозначает отображение из W в V согласно формуле

$$f(A) = \begin{pmatrix} A \\ A \end{pmatrix}.$$

Так как f — линейное отображение, то его производная не зависит от A, является линейным отображением, и вычисление дает

$$df_A(X) = \begin{pmatrix} X \\ X \end{pmatrix}.$$

В дифференциальных обозначениях это следует записать так:

$$d\begin{pmatrix} A \\ A \end{pmatrix} = \begin{pmatrix} dA \\ dA \end{pmatrix},$$

где опять dA — линейное отображение, которое любому элементу X ставит в соответствие значение X.

А теперь рассмотрим отображение $h : W \to W$, определяемое формулой

$$h(A) = A^2.$$

Очевидно, что $h(A) = g(f(A))$ или $h = g \circ f$. Действие цепного правила в этом случае изображено на рис. 5.6.

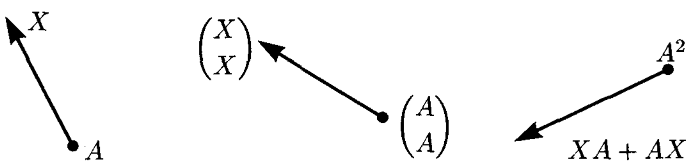


Рис. 5.6

Это значит, что

$$dh_A(Z) = d(g \circ f)_A(Z) = dg_{f(A)}(df_A(Z)) = dg_{f(A)}\begin{pmatrix} Z \\ Z \end{pmatrix} = ZA + AZ.$$

Мы могли бы записать эту выкладку в «дифференциальных» обозначениях, а именно, в уравнении (5.17) сделаем «замены»: $A = A$ и $B = A$. В результате получаем

$$d(A^2) = (dA)A + A(dA).$$

(Отметим еще раз: вследствие некоммутативной природы матричного умножения эта формула дает правильное обобщение формулы $d(x^2) = 2x\,dx$. Формула $d(A^2) = 2A\,dA$ неверна!)

Разложение Борна

Рассмотрим отображение (inv), которое каждой обратимой матрице ставит в соответствие ее обратную, т. е.

$$(inv)(A) = A^{-1}.$$

Отображение (inv) определено только на подмножестве W, состоящем из всех обратимых матриц. *Предполагая*, что (inv) дифференцируемо в области определения, покажем, как вычислять производную (inv), используя цепное правило. Пусть отображение f задано формулой

$$f(A) = \begin{pmatrix} A \\ A^{-1} \end{pmatrix},$$

или, символически,

$$f = \begin{pmatrix} (id) \\ (inv) \end{pmatrix}.$$

Напоминаем, что отображение g, как и выше, определено формулой

$$g\begin{pmatrix} A \\ B \end{pmatrix} = AB.$$

Тогда $(g\circ f)(A) = AA^{-1} = \mathbb{I}$, где $\mathbb{I}$ — единичная матрица. Другими словами, композиция $g\circ f$ постоянна и, следовательно, $d(g\circ f) = 0$. Согласно цепному правилу,

$$df_A(X) = \begin{pmatrix} d_A(id)(X) \\ d_A(inv)(X) \end{pmatrix} = \begin{pmatrix} X \\ d(inv)_A(X) \end{pmatrix}.$$

Воспользуемся цепным правилом еще раз:

$$0 = [d_{f(A)}g](d_A f(X)) = XA^{-1} + A(d_A(inv)(X)).$$

Умножая это уравнение слева на A^{-1} и решая его относительно $d_A(inv)(X)$, получаем

$$d_A(inv)(X) = -A^{-1}XA^{-1}.$$

В дифференциальных обозначениях эта процедура записывается следующим образом. Так как $AA^{-1} = \mathbb{I}$, мы имеем $d(AA^{-1}) = 0$. Заменяя A и A^{-1} на A и B в формуле $d(AB) = (dA)B + A(dB)$, получаем

$$0 = d(AA^{-1}) = (dA)A^{-1} + Ad(A^{-1}).$$

Решая это уравнение относительно $d(A^{-1})$, получаем формулу

$$\boxed{d(A^{-1}) = -A^{-1}(dA)A^{-1}.} \tag{5.18}$$

(Соотношение (5.18) дает правильное обобщение формулы дифференциала $d(1/x) = -(1/x^2)\,dx$ на случай матриц.) Дадим немного другое объяснение формуле (5.18). Предположим, что A — обратимая матрица, т. е. $\operatorname{Det} A \neq 0$. Тогда, если X является матрицей, элементы которой достаточно малы и $\operatorname{Det}(A+X) \neq 0$, то матрица $A+X$ тоже обратима. В этом случае можно написать

$$A + X = (\mathbb{I} + XA^{-1})A.$$

Если X достаточно мала, то матрица XA^{-1} тоже будет малой и ряд

$$(\mathbb{I} + XA^{-1})^{-1} = \mathbb{I} - (XA^{-1}) + (XA^{-1})^2 - (XA^{-1})^3 + \cdots$$

будет сходиться. Тогда мы получим

$$\begin{aligned}(A+X)^{-1} = [(\mathbb{I} + XA^{-1})A]^{-1} &= A^{-1}(\mathbb{I} + XA^{-1})^{-1} \\ &= A^{-1}(\mathbb{I} - (XA^{-1}) + (XA^{-1})^2 - \cdots)\end{aligned}$$

или

$$\begin{aligned}(A+X)^{-1} = A^{-1} - A^{-1}XA^{-1} &+ A^{-1}XA^{-1}XA^{-1} \\ &- A^{-1}XA^{-1}XA^{-1}XA^{-1} + \cdots .\end{aligned}$$

В физической литературе этот ряд называется *борновским разложением* по имени известного физика-теоретика Макса Борна. Формула (5.18) получается из разложения Борна, если опустить все члены более высокого порядка малости, чем X. В физической литературе формула (5.18) известна как *первое борновское приближение*, играющее чрезвычайно важную роль в теории рассеяния. Как мы уже показали, нам не надо знать все борновское разложение для получения первого борновского приближения. Его можно получить непосредственно из цепного правила.

С другой стороны, разложение Борна означает, что

$$(A+X)^{-1} - A^{-1} = -A^{-1}XA^{-1} + o(X).$$

Это *доказывает*, что функция (inv) дифференцируема. Именно это нам пришлось *предположить*, когда мы применяли цепное правило.

Пусть B — постоянная матрица. Рассмотрим отображение $f(A) = ABA^{-1}$. Тогда

$$\begin{aligned} d(ABA^{-1}) &= (dA)BA^{-1} + AB(dA^{-1}) \\ &= (dA)BA^{-1} - ABA^{-1}(dA)A^{-1}. \end{aligned}$$

Это можно записать по-другому:

$$d_A f(X) = XBA^{-1} - ABA^{-1}XA^{-1}.$$

Предположим, что $t \to A(t)$ есть некоторая дифференцируемая кривая в пространстве матриц. Пусть

$$C(t) = A(t)BA(t)^{-1},$$

где B — постоянная матрица. Кроме того, предполагаем, что $A(t)$ обратима для всех t. Применяя цепное правило и предыдущие формулы, получаем

$$C'(t) = A'(t)BA(t)^{-1} - A(t)BA(t)A'(t)A(t)^{-1}.$$

Допустим, что $A(0) = \mathbb{I}$ и $A'(0) = X$. Тогда, полагая $t = 0$ в предыдущей формуле, имеем

$$\boxed{C'(0) = XB - BX.}$$

Это весьма важная формула в математике и физике. Правая часть называется *коммутатором* матриц X и B и обозначается $[X, B]$, т.е.

$$[X, B] = XB - BX.$$

В качестве примера возьмем $A(t) = \exp tX$. Тогда

$$A(t) = \mathbb{I} + tX + \frac{1}{2}t^2X^2 + \cdots .$$

Очевидно, что $A(0) = I$ и $A'(0) = X$, поэтому можно воспользоваться полученной формулой. Итак, мы имеем

$$A(t)^{-1} = (\exp tX)^{-1} = \exp(-tX) = \mathbb{I} - tX + \frac{1}{2}t^2X^2 + \cdots ,$$

поэтому

$$\begin{aligned} A(t)BA(t)^{-1} &= (\mathbb{I} + tX + \frac{1}{2}t^2X^2 + \cdots)B(\mathbb{I} - tX + \frac{1}{2}t^2X^2 + \cdots) \\ &= B + t(XB - BX) + \frac{1}{2}t^2(X^2B - 2XBX + BX^2) + \cdots . \end{aligned}$$

Собрав вместе члены порядка два и выше по t, получаем

$$A(t)BA(t)^{-1} = B + t[X, B] + o(t).$$

Движение Кеплера

Мы уже видели, что цепное правило влечет правило Лейбница для производной произведения, причем это верно и для произведения матриц, когда операция умножения не обладает свойством коммутативности. Давайте используем ту же аргументацию для векторного произведения в пространстве $\mathbb{R}^3$. (Сейчас мы напомним это определение.) В качестве следствия мы получим второй закон Кеплера, описывающий движение планет.

В трехмерном пространстве векторное произведение определено следующим образом:

$$\text{Если} \quad \mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \text{и} \quad \mathbf{w} = \begin{pmatrix} p \\ q \\ r \end{pmatrix}, \quad \text{то} \quad \mathbf{v} \times \mathbf{w} = \begin{pmatrix} yr - zq \\ zp - xr \\ xq - yp \end{pmatrix}.$$

Из этого определения вытекают свойства векторного произведения:

$$(\mathbf{v}_1+\mathbf{v}_2)\times\mathbf{w} = \mathbf{v}_1\times\mathbf{w}+\mathbf{v}_2\times\mathbf{w}, \quad \mathbf{v}\times(\mathbf{w}_1+\mathbf{w}_2) = \mathbf{v}\times\mathbf{w}_1+\mathbf{v}\times\mathbf{w}_2,$$

$$(a\mathbf{v})\times\mathbf{w} = \mathbf{v}\times(a\mathbf{w}) = a(\mathbf{v}\times\mathbf{w}) \quad \text{и} \quad \mathbf{v}\times\mathbf{v} = \mathbf{0}.$$

Из первых трех уравнений следует, что знак «$\times$» действует как знак умножения. Можно доказать, что

$$d(\mathbf{v}\times\mathbf{w}) = d\mathbf{v}\times\mathbf{w}+\mathbf{v}\times d\mathbf{w}.$$

В частности, если $\mathbf{v}(t)$ и $\mathbf{w}(t)$ — кривые в пространстве $\mathbb{R}^3$, и если мы положим

$$\mathbf{u}(t) = \mathbf{v}(t)\times\mathbf{w}(t),$$

то

$$\mathbf{u}'(t) = \mathbf{v}'(t)\times\mathbf{w}(t)+\mathbf{v}(t)\times\mathbf{w}'(t).$$

Допустим, что $\mathbf{r}(t)$ обозначает положение частицы в момент времени t, и пусть $\mathbf{p}(t)$ обозначает импульс этой частицы в момент t. Тогда вектор

$$\boldsymbol{\mu}(t) = \mathbf{p}(t)\times\mathbf{r}(t)$$

называется *угловым моментом* частицы относительно начала координат (в момент времени t). Предположим, что частица имеет массу m, и на нее действует сила $\mathbf{F}(t) = c(t)\mathbf{r}(t)$, направленная вдоль вектора, идущего от начала координат к частице. Поскольку $\mathbf{r}(t)$, $\mathbf{p}(t)$ и $\mathbf{F}(t)$ связаны формулами

$$\mathbf{r}'(t) = (1/m)\mathbf{p}(t) \quad \text{и} \quad \mathbf{p}'(t) = \mathbf{F}(t) = c(t)\mathbf{r}(t),$$

то

$$\begin{aligned}\boldsymbol{\mu}'(t) &= \mathbf{p}'(t)\times\mathbf{r}+\mathbf{p}(t)\times\mathbf{r}'(t)\\ &= c(t)\mathbf{r}(t)\times\mathbf{r}(t)+(1/m)\mathbf{p}(t)\times\mathbf{p}(t) = 0.\end{aligned}$$

Другими словами, вектор $\boldsymbol{\mu}$ должен быть постоянным. Этот закон называется *законом сохранения углового момента*. Для простоты предположим, что $\boldsymbol{\mu}\neq\mathbf{0}$. Из определения векторного произведения легко показать, что для любых векторов $\mathbf{v}$ и $\mathbf{w}$ выполняется соотношение $(\mathbf{v}\times\mathbf{w})\cdot\mathbf{w} = 0$. Так как $\boldsymbol{\mu} = \mathbf{p}(t)\times\mathbf{r}(t)$, мы получаем, что $\boldsymbol{\mu}\cdot\mathbf{r}(t)=0$ для всех значений t. Другими словами, частица всегда двигается в фиксированной плоскости, которая перпендикулярна вектору $\boldsymbol{\mu}$. В пространстве $\mathbb{R}^3$ повернем систему координат так, чтобы $\boldsymbol{\mu}$ был направлен вдоль оси z, тогда частица будет двигаться в плоскости xy. Итак,

$$\mathbf{r}(t)=\begin{pmatrix}x(t)\\y(t)\\0\end{pmatrix},\quad \text{тогда}\quad \mathbf{p}(t)=\begin{pmatrix}x'(t)\\y'(t)\\0\end{pmatrix}$$

$$\text{и}\quad \boldsymbol{\mu}=m\begin{pmatrix}0\\0\\x'(t)y(t)-x(t)y'(t)\end{pmatrix}.$$

Таким образом, условие сохранения $\boldsymbol{\mu}$ значит, что выражение $x'(t)y(t)-x(t)y'(t)$ постоянно. Чтобы понять смысл этого условия, нарисуем траекторию частицы в плоскости xy (рис. 5.7). С точностью до членов $o(h)$ область, ограниченная траекторией $\begin{pmatrix}x(t)\\y(t)\end{pmatrix}$ и вектором $\begin{pmatrix}x(t+h)\\y(t+h)\end{pmatrix}$, совпадает с площадью треугольника, построенного на векторах $\begin{pmatrix}x(t)\\y(t)\end{pmatrix}$ и $\begin{pmatrix}x(t+h)\\y(t+h)\end{pmatrix}$, т. е. с этой точностью мы можем пренебречь площадью, заштрихованной на рис. 5.7.

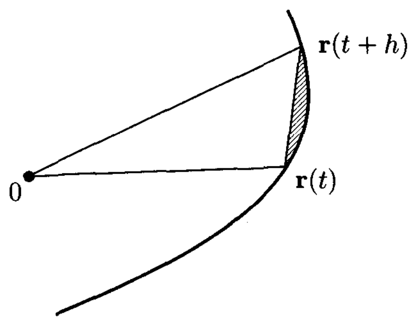


Рис. 5.7

С точностью до знака площадь треугольника равна

$$\frac{1}{2}(x(t+h)y(t)-x(t)y(t+h)).$$

Но

$$x(t+h) = x(t) + hx'(t) + o(h) \quad \text{и} \quad y(t+h) = y(t) + hy'(t) + o(h).$$

Тогда площадь треугольника равна

$$\frac{1}{2}(x'(t)y(t) - x(t)y'(t))h + o(h).$$

Это значит, что скорость изменения площади, описываемой радиусом-вектором движущейся частицы, постоянна. Так формулируется второй закон Кеплера. Используя цепное правило, мы видим, что второй закон Кеплера и то, что частица движется в фиксированной плоскости, является результатом действия силы, направленной вдоль ее радиуса-вектора. Тот факт, что движение планет происходит в фиксированных плоскостях и за равные промежутки времени их радиусы-векторы описывают равные площади, определяется тем, что действующая на них сила направлена к Солнцу. Изложенный вывод второго закона Кеплера принадлежит Ньютону.

5.4. Частные производные и дифференциальные формы

В этом параграфе мы введем понятия и обозначения, удобные для применения цепного правила. Рассмотрим дифференцируемую функцию $f : \mathbb{R}^k \to \mathbb{R}$. Например, пусть $k = 3$ и

$$f\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = x^2y^3z^4.$$

Тогда $df_{\mathbf{v}}$ есть линейное отображение $\mathbb{R}^3 \to \mathbb{R}$. Следовательно, $df_{\mathbf{v}}$ можно рассматривать как вектор-строку. Утверждается, что для любой точки $\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ вектор-строка определяется формулой

$$df_{\mathbf{v}} = (2xy^3z^4,\ 3x^2y^2z^4,\ 4x^2y^3z^3).$$

Чтобы убедиться в этом, вычислим его проекции на каждый из векторов $\begin{pmatrix}1\\0\\0\end{pmatrix}$, $\begin{pmatrix}0\\1\\0\end{pmatrix}$ и $\begin{pmatrix}0\\0\\1\end{pmatrix}$. Например,

$$f\left(\mathbf{v}+s\begin{pmatrix}1\\0\\0\end{pmatrix}\right)-f(\mathbf{v})=df_{\mathbf{v}}\begin{pmatrix}s\\0\\0\end{pmatrix}+o(s)$$
$$=s\,df_{\mathbf{v}}\begin{pmatrix}1\\0\\0\end{pmatrix}+o(s).$$

Это следует из определения $df_{\mathbf{v}}$ и из того факта, что дифференциал $df_{\mathbf{v}}[\mathbf{h}]$ линеен по $\mathbf{h}$. Тогда

$$f\left(\mathbf{v}+s\begin{pmatrix}1\\0\\0\end{pmatrix}\right)-f\left(\begin{pmatrix}x\\y\\z\end{pmatrix}\right)=f\left(\begin{pmatrix}x+s\\y\\z\end{pmatrix}\right)-f\left(\begin{pmatrix}x\\y\\z\end{pmatrix}\right)$$

и предел этого выражения при $s\to 0$ является производной f по переменной x при фиксированных y и z. Он называется *частной производной* f по x и обозначается $\partial f/\partial x$. Если

$$f\left(\begin{pmatrix}x\\y\\z\end{pmatrix}\right)=x^2y^3z^4,$$

то элементарное вычисление дает

$$\frac{\partial f}{\partial x}\left(\begin{pmatrix}x\\y\\z\end{pmatrix}\right)=2xy^3z^4.$$

Таким образом,

$$df_{\mathbf{v}}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}\right)=\frac{\partial f}{\partial x}.$$

Аналогично, $df_{\mathbf{v}}\left(\begin{pmatrix}0\\1\\0\end{pmatrix}\right)=\dfrac{\partial f}{\partial y}$ и $df_{\mathbf{v}}\left(\begin{pmatrix}0\\0\\1\end{pmatrix}\right)=\dfrac{\partial f}{\partial z}$. Итак, для функции f мы получаем формулу

$$df_{\mathbf{v}}=\left(\frac{\partial f}{\partial x}(\mathbf{v}),\ \frac{\partial f}{\partial y}(\mathbf{v}),\ \frac{\partial f}{\partial z}(\mathbf{v})\right).$$

Непосредственная проверка для $f=x^2y^3z^4$ дает тот же результат. Есть более удобный способ организации этой информации. Напоминаем, что мы уже имели обозначение dx для линейной функции, приписывающей каждому вектору его первую компоненту, т. е. dx задается матрицей

$$(1,\ 0,\ 0)$$

и, аналогично, для dy и dz — матрицы

$$(0,\ 1,\ 0)\quad \text{и}\quad (0,\ 0,\ 1).$$

Теперь можно записать равенство

$$\begin{aligned}df_{\mathbf{v}}&=\left(\frac{\partial f}{\partial x}(\mathbf{v}),\ \frac{\partial f}{\partial y}(\mathbf{v}),\ \frac{\partial f}{\partial z}(\mathbf{v})\right)\\&=\frac{\partial f}{\partial x}(\mathbf{v})(1,\ 0,\ 0)+\frac{\partial f}{\partial y}(\mathbf{v})(0,\ 1,\ 0)+\frac{\partial f}{\partial z}(\mathbf{v})(0,\ 0,\ 1)\end{aligned}$$

в форме

$$df=\frac{\partial f}{\partial x}\,dx+\frac{\partial f}{\partial y}\,dy+\frac{\partial f}{\partial z}\,dz.$$

Таким образом,

$$d(x^2y^3z^4)=2xy^3z^4\,dx+3x^2y^2z^4\,dy+4x^2y^3z^3\,dz.$$

Справа стоит сумма трех членов, являющихся функциями, умноженными на dx, dy и dz. Такая сумма называется *линейной дифференциальной формой*. Смысл этой формы в том, что она определяет правило, по которому каждой точке пространства $\mathbb{R}^3$ приписывается вектор-строка.

Это формула естественна в том смысле, что если взять функцию

$$f\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = x,$$

то

$$\frac{\partial f}{\partial x} \equiv 1, \quad \frac{\partial f}{\partial y} \equiv 0, \quad \frac{\partial f}{\partial z} \equiv 0,$$

и тогда мы получаем

$$df = dx.$$

С этими обозначениями цепное правило сводится к формальным подстановкам. Покажем, что мы имеем в виду. Рассмотрим отображение $\phi : \mathbb{R}^2 \to \mathbb{R}^2$, определяемое формулой

$$\phi\left(\begin{pmatrix} r \\ \theta \end{pmatrix}\right) = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}.$$

Пусть $f : \mathbb{R}^2 \to \mathbb{R}$ есть некая функция, скажем,

$$f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = x^3 + y^2 x.$$

Тогда

$$f \circ \phi\left(\begin{pmatrix} r \\ \theta \end{pmatrix}\right) = r^3 \cos\theta.$$

Отображение $d\phi_{\left(\begin{smallmatrix} r \\ \theta \end{smallmatrix}\right)}$ будет матрицей размера 2×2, скажем,

$$d\phi_{\left(\begin{smallmatrix} r \\ \theta \end{smallmatrix}\right)} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Чтобы найти верхнюю строку этой матрицы, мы должны умножить ее слева на вектор $(1,\, 0)$. Тогда

$$(1,\, 0)\begin{pmatrix} a & b \\ c & d \end{pmatrix} = (a,\, b).$$

В нашем случае $(1,\, 0)$ есть dx. Согласно цепному правилу,

$$dx_{\phi\left(\left(\begin{smallmatrix} r \\ \theta \end{smallmatrix}\right)\right)} \circ d\phi_{\left(\begin{smallmatrix} r \\ \theta \end{smallmatrix}\right)} = d(x \circ d\phi)_{\left(\begin{smallmatrix} r \\ \theta \end{smallmatrix}\right)}.$$

Но $x \circ \phi\left(\begin{pmatrix} r \\ \theta \end{pmatrix}\right) = r\cos\theta$, следовательно, мы получаем вектор-строку

$$\begin{aligned} d(x \circ \phi)_{\binom{r}{\theta}} &= \cos\theta\, dr - r\sin\theta\, d\theta \\ &= (\cos\theta,\ -r\sin\theta). \end{aligned}$$

Итак, $a = \cos\theta$, $b = -r\sin\theta$. Аналогично

$$\begin{aligned} (c,\, d) = (0,\, 1)\begin{pmatrix} a & b \\ c & d \end{pmatrix} &= dy_{\phi(\binom{r}{\theta})} \circ d\phi_{\binom{r}{\theta}} \\ &= d(y \circ \phi)_{\binom{r}{\theta}} = \sin\theta\, dr + r\cos\theta\, d\theta = (\sin\theta,\ r\cos\theta). \end{aligned}$$

Откуда следует, что $d\phi_{\binom{r}{\theta}}$ задается матрицей

$$\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}.$$

Тогда

$$df = (3x^2 + y^2)\, dx + (2xy)\, dy$$

и

$$d(f \circ \phi) = 3r^2\cos\theta\, dr - r^3\sin\theta\, d\theta.$$

В принципе, согласно цепному правилу, можно перемножить матрицы и получить матрицу для дифференциала $d(f \circ \phi)$:

$$\begin{aligned} (3r^2\cos^2\theta + r^2\sin^2\theta,\ 2r^2\sin\theta\cos\theta) &\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \\ &= (3r^2\cos\theta,\ -r^3\sin\theta). \end{aligned}$$

Конечно, это правильно. Однако, можно сделать иначе — подставить

$$\begin{aligned} x &= r\cos\theta, \quad dx = \cos\theta\, dr - r\sin\theta\, d\theta, \\ y &= r\sin\theta, \quad dy = \sin\theta\, dr + r\cos\theta\, d\theta \end{aligned}$$

в выражение

$$df = (3x^2 + y^2)\, dx + (2xy)\, dy,$$

после этого все перемножить, собрать вместе коэффициенты и получить то же самое выражение для $d(f \circ \phi)$.

Другими словами, можно считать, что отображение ϕ превращает x в функцию r и θ, т.е. x надо заменить на функцию $x \circ \phi = r\cos\theta$ и вычислить ее дифференциал.

Чтобы проделать эти вычисления, полезно вспомнить, что

$$d(gh) = g\,dh + h\,dg.$$

Например, в пространстве $\mathbb{R}^2$

$$dg = \frac{\partial g}{\partial x}\,dx + \frac{\partial g}{\partial y}\,dy,$$

откуда

$$hdg = h\frac{\partial g}{\partial x}\,dx + h\frac{\partial g}{\partial y}\,dy.$$

Тогда

$$d[(gh)\circ\phi] = (g\circ\phi)\,d(h\circ\phi) + (h\circ\phi)\,d(g\circ\phi).$$

Таким образом, в нашем примере

$$F\begin{pmatrix} x \\ y \end{pmatrix} = x^3 + y^2x = (x^2+y^2)x = gh\begin{pmatrix} x \\ y \end{pmatrix},$$

где $g = x^2 + y^2$ и $h = x$. Тогда $f\circ\phi = r^2 \cdot r\cos\theta$ и, следовательно,

$$\begin{aligned} d(f\circ\phi) &= (r\cos\theta)2r\,dr + r^2(\cos\theta\,dr - r\sin\theta\,d\theta) \\ &= 3r^2\cos\theta\,dr - r^3\sin\theta\,d\theta. \end{aligned}$$

Эта процедура легко обобщается: пусть $y_1, \dots, y_l$ обозначают координатные функции в пространстве $\mathbb{R}^l$, где точка задается столбцом

$$\begin{pmatrix} y_1 \\ \vdots \\ y_l \end{pmatrix}.$$

Пусть $f : \mathbb{R}^l \to \mathbb{R}$ есть дифференцируемая функция. Тогда

$$df = \frac{\partial f}{\partial y_1}\,dy_1 + \cdots + \frac{\partial f}{\partial y_l}\,dy_l.$$

Предположим, что $x_1, \ldots, x_k$ обозначают координаты в пространстве $\mathbb{R}^k$. Пусть $\phi : \mathbb{R}^k \to \mathbb{R}^l$ — дифференцируемое отображение. Определим $\phi_1 = y_1 \circ \phi$, $\phi_2 = y_2 \circ \phi$ и так далее. Тогда

$$\phi(\mathbf{v}) = \begin{pmatrix} \phi_1(\mathbf{v}) \\ \vdots \\ \phi_l(\mathbf{v}) \end{pmatrix}, \quad \text{где} \quad \mathbf{v} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}.$$

Далее можно написать

$$\begin{aligned} d\phi_1 &= \frac{\partial \phi_1}{\partial x_1}\, dx_1 + \cdots + \frac{\partial \phi_1}{\partial x_k}\, dx_k, \\ &\vdots \\ d\phi_l &= \frac{\partial \phi_l}{\partial x_1}\, dx_1 + \cdots + \frac{\partial \phi_l}{\partial x_k}\, dx_k, \end{aligned}$$

и линейное отображение $d\phi_{\mathbf{v}}$ задается матрицей

$$\begin{pmatrix} \dfrac{\partial \phi_1}{\partial x_1}(\mathbf{v}) & \cdots & \dfrac{\partial \phi_1}{\partial x_k}(\mathbf{v}) \\ \vdots & & \\ \dfrac{\partial \phi_l}{\partial x_1}(\mathbf{v}) & \cdots & \dfrac{\partial \phi_l}{\partial x_k}(\mathbf{v}) \end{pmatrix},$$

или в более простой форме

$$d\phi = \begin{pmatrix} \dfrac{\partial \phi_1}{\partial x_1} & \cdots & \dfrac{\partial \phi_1}{\partial x_k} \\ \vdots & & \\ \dfrac{\partial \phi_l}{\partial x_1} & \cdots & \dfrac{\partial \phi_l}{\partial x_k} \end{pmatrix}.$$

Согласно цепному правилу,

$$d(f \circ \phi) = \frac{\partial f}{\partial y_1} \circ \phi\, d\phi_1 + \cdots + \frac{\partial f}{\partial y_l} \circ \phi\, d\phi_l,$$

где использованы выражения $d\phi_i = \dfrac{\partial \phi_i}{\partial x_1}\, dx_1 + \cdots + \dfrac{\partial \phi_i}{\partial x_k}\, dx_k$.

Закончим этот параграф доказательством двух теорем.

Чтобы определить частную производную, мы использовали дифференцируемость f в точке $\mathbf{p}$. Можно показать, что существование частных производных по x и по y еще не означает дифференцируемость f в $\mathbf{p}$. Достаточные условия дифференцируемости f в точке p дает следующая теорема.

Теорема: *Пусть* $f : \mathbb{R}^2 \to \mathbb{R}^1$ *имеет непрерывные частные производные* $\dfrac{\partial f}{\partial x}$ *и* $\dfrac{\partial f}{\partial y}$ *в точке* $\mathbf{p}$. *Тогда* f *дифференцируема в точке* $\mathbf{p}$.

Доказательство. Если f дифференцируема в $\mathbf{p}$, то линейное отображение $df_{\mathbf{p}}$ должно иметь вид

$$df_{\mathbf{p}}\left(\begin{pmatrix} s \\ t \end{pmatrix}\right) = s\frac{\partial f}{\partial x}\bigg|_{\mathbf{p}} + t\frac{\partial f}{\partial y}\bigg|_{\mathbf{p}}.$$

Итак, если существуют $\dfrac{\partial f}{\partial x}\bigg|_{\mathbf{p}}$ и $\dfrac{\partial f}{\partial y}\bigg|_{\mathbf{p}}$, то f дифференцируема в точке $\mathbf{p}$ тогда и только тогда, когда

$$\nabla_{\mathbf{p}} f\left(\begin{pmatrix} s \\ t \end{pmatrix}\right) = s\frac{\partial f}{\partial x}\bigg|_{\mathbf{p}} + t\frac{\partial f}{\partial y}\bigg|_{\mathbf{p}} + o(\mathbb{R}^2, \mathbb{R}^1).$$

Запишем приращение функции f в виде

$$\nabla_{\mathbf{p}} f\left(\begin{pmatrix} s \\ t \end{pmatrix}\right) = f(x_{\mathbf{p}} + s, y_{\mathbf{p}} + t) - f(x_{\mathbf{p}}, y_{\mathbf{p}})$$
$$= f(x_{\mathbf{p}} + s, y_{\mathbf{p}} + t) - f(x_{\mathbf{p}}, y_{\mathbf{p}} + t) + f(x_{\mathbf{p}}, y_{\mathbf{p}} + t) - f(x_{\mathbf{p}}, y_{\mathbf{p}}).$$

По теореме Лагранжа о среднем значении из одномерного дифференциального исчисления можно записать:

$$f(x_{\mathbf{p}} + s,\, y_{\mathbf{p}} + t) - f(x_{\mathbf{p}},\, y_{\mathbf{p}} + t) = s\frac{\partial f}{\partial x}\bigg|_{\left(\begin{smallmatrix} x_0 \\ y_{\mathbf{p}}+t \end{smallmatrix}\right)}$$
$$f(x_{\mathbf{p}},\, y_{\mathbf{p}} + t) - f(x_{\mathbf{p}},\, y_{\mathbf{p}}) = t\frac{\partial f}{\partial y}\bigg|_{\left(\begin{smallmatrix} x_{\mathbf{p}} \\ y_0 \end{smallmatrix}\right)}$$

для некоторых x_0 и y_0, удовлетворяющих условиям $x_{\mathbf{p}} < x_0 < x_{\mathbf{p}} + s$ и $y_{\mathbf{p}} < y_0 < y_{\mathbf{p}} + t$. Тогда

$$\nabla_{\mathbf{p}} f\left(\begin{pmatrix} s \\ t \end{pmatrix}\right) = s\frac{\partial f}{\partial x}\bigg|_{\left(\begin{smallmatrix} x_0 \\ y_{\mathbf{p}}+t \end{smallmatrix}\right)} + t\frac{\partial f}{\partial y}\bigg|_{\left(\begin{smallmatrix} x_{\mathbf{p}} \\ y_0 \end{smallmatrix}\right)},$$

так что

$$\nabla_{\mathbf{p}} f\left(\begin{pmatrix} s \\ t \end{pmatrix}\right) - s\frac{\partial f}{\partial x}\bigg|_{\mathbf{p}} - t\frac{\partial f}{\partial y}\bigg|_{\mathbf{p}} =$$

$$= s\left\{\frac{\partial f}{\partial x}\bigg|_{\left(\begin{smallmatrix} x_0 \\ y_{\mathbf{p}}+t \end{smallmatrix}\right)} - \frac{\partial f}{\partial x}\bigg|_{\left(\begin{smallmatrix} x_{\mathbf{p}} \\ y_{\mathbf{p}} \end{smallmatrix}\right)}\right\} + t\left\{\frac{\partial f}{\partial y}\bigg|_{\left(\begin{smallmatrix} x_{\mathbf{p}} \\ y_0 \end{smallmatrix}\right)} - \frac{\partial f}{\partial y}\bigg|_{\left(\begin{smallmatrix} x_{\mathbf{p}} \\ y_{\mathbf{p}} \end{smallmatrix}\right)}\right\}.$$

Если $\left\|\begin{pmatrix} s \\ t \end{pmatrix}\right\|$ стремится к нулю, то ввиду непрерывности функций $\partial f/\partial x$ и $\partial f/\partial y$ коэффициенты при s и t тоже стремятся к нулю, т. е. относятся к категории $I(\mathbb{R}^2, \mathbb{R}^1)$. Поскольку обе функции s и t являются $O(\mathbb{R}^2, \mathbb{R}^1)$, то все выражение есть $o(\mathbb{R}^2, \mathbb{R}^1)$, что и требовалось доказать.

Теорема: *Пусть* $f : \mathbb{R}^2 \to \mathbb{R}^1$ *дифференцируема в точке* $\mathbf{p}$ *и имеет непрерывные смешанные производные* $\dfrac{\partial}{\partial y}\left(\dfrac{\partial f}{\partial x}\right)$ *и* $\dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial y}\right)$ *в точке* $\mathbf{p}$. *Тогда*

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\bigg|_{\mathbf{p}} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)\bigg|_{\mathbf{p}}.$$

Интуитивная идея, лежащая в основе доказательства этой важной теоремы, очень проста. Возьмем четыре точки $\begin{pmatrix} x \\ y \end{pmatrix}$, $\begin{pmatrix} x+s \\ y \end{pmatrix}$, $\begin{pmatrix} x \\ y+t \end{pmatrix}$ и $\begin{pmatrix} x+s \\ y+t \end{pmatrix}$. Они расположены на плоскости так:

$$\begin{array}{ll} \bullet\begin{pmatrix} x \\ y+t \end{pmatrix} & \bullet\begin{pmatrix} x+s \\ y+t \end{pmatrix} \\ \bullet\begin{pmatrix} x \\ y \end{pmatrix} & \bullet\begin{pmatrix} x+s \\ y \end{pmatrix} \end{array}$$

Рассмотрим выражение, составленное из значений функции f в этих точках:

$$f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) - f\left(\begin{pmatrix} x+s \\ y \end{pmatrix}\right) + f\left(\begin{pmatrix} x+s \\ y+t \end{pmatrix}\right) - f\left(\begin{pmatrix} x \\ y+t \end{pmatrix}\right).$$

Его можно сгруппировать двумя способами: либо

$$f\left(\begin{pmatrix} x+s \\ y+t \end{pmatrix}\right) - f\left(\begin{pmatrix} x+s \\ y \end{pmatrix}\right) - \left(f\left(\begin{pmatrix} x \\ y+t \end{pmatrix}\right) - f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right)\right), \quad (5.19)$$

либо

$$f\left(\begin{pmatrix} x+s \\ y+t \end{pmatrix}\right) - f\left(\begin{pmatrix} x \\ y+t \end{pmatrix}\right) - \left(f\left(\begin{pmatrix} x+s \\ y \end{pmatrix}\right) - f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right)\right). \quad (5.20)$$

Очевидно, что они равны. Выражение (5.20) равно

$$\nabla_{\left(\begin{smallmatrix} x \\ y+t \end{smallmatrix}\right)} f\left(\begin{pmatrix} s \\ 0 \end{pmatrix}\right) - \nabla_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)} f\left(\begin{pmatrix} s \\ 0 \end{pmatrix}\right).$$

Учтем, что

$$\nabla_{\left(\begin{smallmatrix} x \\ y+t \end{smallmatrix}\right)} f\left(\begin{pmatrix} s \\ 0 \end{pmatrix}\right) = s\left.\frac{\partial f}{\partial x}\right|_{\left(\begin{smallmatrix} x \\ y+t \end{smallmatrix}\right)} + o(s)$$

и

$$\nabla_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)} f\left(\begin{pmatrix} s \\ 0 \end{pmatrix}\right) = s\left.\frac{\partial f}{\partial x}\right|_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)} + o(s).$$

Если бы мы могли «пренебречь» поправочными членами $o(s)$, то мы бы имели

$$\begin{aligned} \nabla_{\left(\begin{smallmatrix} x \\ y+t \end{smallmatrix}\right)} f\left(\begin{pmatrix} s \\ 0 \end{pmatrix}\right) - \nabla_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)} f\left(\begin{pmatrix} s \\ 0 \end{pmatrix}\right) &= s\left(\frac{\partial f}{\partial x}\left(\begin{pmatrix} x \\ y+t \end{pmatrix}\right) - \frac{\partial f}{\partial x}\left(\begin{pmatrix} x \\ y \end{pmatrix}\right)\right) \\ &= s\,\nabla_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)}\left(\frac{\partial f}{\partial x}\right)\begin{pmatrix} o \\ t \end{pmatrix} = st\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) + s\,o(t). \end{aligned}$$

Если те же операции проделать с выражением (5.19), то получим

$$ts\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)\begin{pmatrix} x \\ y \end{pmatrix} + t\,o(s).$$

Разделив полученные ответы на st, получаем (без учета последних, «малых» слагаемых)

$$\frac{\partial}{\partial x}\frac{\partial f}{\partial y}=\frac{\partial}{\partial y}\frac{\partial f}{\partial x}.$$

Чтобы приведенное рассуждение стало строгим, нам необходимо внимательно исследовать поправочные члены. Мы можем сделать это с помощью теоремы о среднем значении из дифференциального исчисления функций одной переменной. Положим

$$g(y)=f\left(\begin{pmatrix} x+s \\ y \end{pmatrix}\right)-f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right).$$

Функция g дифференцируема по y и выражение (5.20) равно

$$g(y+t)-g(y).$$

Согласно теореме о среднем значении

$$g(y+t)-g(y)=tg'(\bar{y}),$$

где $\bar{y}$ — некоторая точка между y и $y+t$. Далее,

$$\begin{aligned} g'(\bar{y}) &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}(g(\bar{y}+\varepsilon)-g(\bar{y})) \\ &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(f\left(\begin{pmatrix} x+s \\ \bar{y}+\varepsilon \end{pmatrix}\right)-f\left(\begin{pmatrix} x+s \\ \bar{y} \end{pmatrix}\right)-f\left(\begin{pmatrix} x \\ \bar{y}+\varepsilon \end{pmatrix}\right)+f\left(\begin{pmatrix} x \\ \bar{y} \end{pmatrix}\right)\right) \\ &= \frac{\partial}{\partial y}f\left(\begin{pmatrix} x+s \\ \bar{y} \end{pmatrix}\right)-\frac{\partial}{\partial y}f\left(\begin{pmatrix} x \\ \bar{y} \end{pmatrix}\right). \end{aligned}$$

По предположению теоремы функция $\partial f/\partial y$ дифференцируема. Еще раз воспользуемся теоремой о среднем значении

$$\frac{\partial}{\partial y}f\left(\begin{pmatrix} x+s \\ \bar{y} \end{pmatrix}\right)-\frac{\partial}{\partial y}f\left(\begin{pmatrix} x \\ \bar{y} \end{pmatrix}\right)=s\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)\left(\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}\right).$$

Таким образом, выражение (5.20) равно

$$st\frac{\partial}{\partial x}\frac{\partial}{\partial y}f\left(\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}\right).$$

По условию теоремы функция $\dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial y}\right)$ непрерывна. Поэтому для любого $\varepsilon > 0$ можно найти такое $\delta > 0$, что

$$\left|\frac{\partial}{\partial x}\frac{\partial}{\partial y}f\left(\begin{pmatrix}\bar{x}\\ \bar{y}\end{pmatrix}\right) - \frac{\partial}{\partial x}\frac{\partial}{\partial y}f\left(\begin{pmatrix}x\\ y\end{pmatrix}\right)\right| < \varepsilon, \quad \text{если} \quad |s| + |t| < \delta.$$

Таким образом, (5.20) преобразуется к

$$st\left(\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)\left(\begin{pmatrix}x\\ y\end{pmatrix}\right) + r_1\right),$$

где $|r_1| < \varepsilon$, если $|s| + |t| < \delta$. Аналогично получаем, что (5.19) равняется

$$st\left(\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\left(\begin{pmatrix}x\\ y\end{pmatrix}\right) + r_2\right).$$

Будем считать, что $|st| > 0$. Разделив полученные результаты на st, получаем неравенство

$$\left|\frac{\partial}{\partial x}\frac{\partial f}{\partial y}\left(\begin{pmatrix}x\\ y\end{pmatrix}\right) - \frac{\partial}{\partial y}\frac{\partial f}{\partial x}\left(\begin{pmatrix}x\\ y\end{pmatrix}\right)\right| < r_3,$$

где

$$|r_3| \le |r_1| + |r_2| \le 2\varepsilon, \quad \text{если } |s| + |t| < \delta.$$

Левая часть неравенства есть просто разность между двумя числами, она не зависит от δ, следовательно, r_3 может быть сделано сколь угодно малым и, мы получаем равенство смешанных производных, т. е.

$$\frac{\partial}{\partial x}\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}\frac{\partial f}{\partial x}.$$

5.5. Производные по направлению

Пусть I — интервал в пространстве $\mathbb{R}^1$, содержащий 0. Возьмем $\gamma : I \to V$, дифференцируемую в 0. (Как обычно, в качестве V можно взять любое векторное пространство, для наглядности представляйте себе случай $V = \mathbb{R}^2$.) Предположим, что $\gamma(0) = \mathbf{x}$, и пусть $\gamma'(0)$ обозначает вектор $d\gamma_0(1)$, т. е.

$$\gamma'(0) = d\gamma_0(1) = \lim_{t \to 0}[(1/t)(\gamma(t) - \gamma(0))].$$

Вектор $\gamma'(0)$ называется *касательным вектором* к кривой γ в точке $t = 0$. Если γ_1 — вторая кривая, для которой выполняются условия $\gamma_1(0) = \gamma(0)$ и $\gamma_1'(0) = \gamma'(0)$ (рис. 5.8) то мы будем говорить, что γ и γ_1 касаются при $t = 0$, или что они согласуются в нуле с точностью первого порядка.

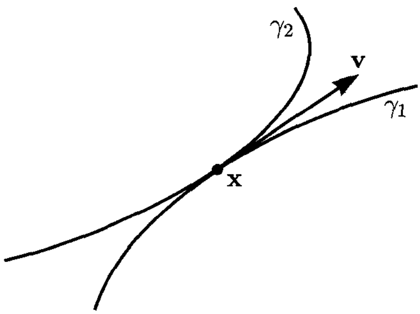


Рис. 5.8

Если γ касается γ_1 в нуле, и γ_1 касается γ_2 в нуле, то очевидно, что γ касается γ_2 в нуле. Другими словами, мы определили отношение эквивалентности для дифференцируемых кривых: две кривые эквивалентны, если они согласуются в нуле с точностью первого порядка. Если $\gamma'(0) = \mathbf{v}$, то пара $\{\mathbf{x}, \mathbf{v}\}$ определяет класс эквивалентности. Мы представляем этот класс эквивалентности как вектор $\mathbf{v}$ с началом в точке $\mathbf{x}$, и называем этот вектор *касательным вектором* в точке $\mathbf{x}$. Любые $\mathbf{x}$ и $\mathbf{v}$ определяют некоторый класс эквивалентности, потому что всегда можно взять прямую линию

$$\gamma(t) = \mathbf{x} + t\mathbf{v},$$

для которой выполняются условия $\gamma(0) = \mathbf{x}$ и $\gamma'(0) = \mathbf{v}$. Иногда мы будем пользоваться греческой буквой, например, $\boldsymbol{\xi}$, для обозначения касательного вектора в точке $\mathbf{x}$, т. е. $\boldsymbol{\xi}$ обозначает пару $\mathbf{x}$ и $\mathbf{v}$.

Предположим, что $V = \mathbb{R}^2$, тогда кривая γ задается двумя функциями $x \circ \gamma$ и $y \circ \gamma$, обычно обозначаемыми $x(t)$ и $y(t)$. Например,

$$\begin{aligned} x(t) &= t \sin t + 1 \\ y(t) &= e^t \end{aligned}$$

определяют кривую

$$\gamma(t) = \begin{pmatrix} t \sin t + 1 \\ e^t \end{pmatrix}$$

с начальными условиями

$$\gamma(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{и} \quad \gamma'(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Заметим, что

$$\begin{aligned} d(x \circ \gamma) &= d(t \sin t + 1) = (\sin t + t \cos t)\, dt \\ d(y \circ \gamma) &= d(e^t) = e^t\, dt. \end{aligned}$$

Поэтому функция

$$\gamma'(t) = \begin{pmatrix} \sin t + t \cos t \\ e^t \end{pmatrix}$$

может быть восстановлена по коэффициентам при dt в $d(x \circ \gamma)$ и $d(y \circ \gamma)$.

Пусть функция $f : V \to \mathbb{R}$ определена в некоторой окрестности точки $\mathbf{p}$. Для каждой кривой γ, удовлетворяющей условию $\gamma(0) = \mathbf{p}$, функция $f \circ \gamma$ определена вблизи 0 в пространстве $\mathbb{R}$. Если f — дифференцируемая функция в $\mathbf{p}$ и γ дифференцируема в нуле, то, согласно *цепному правилу*, $f \circ \gamma$ будет (вещественной) функцией, дифференцируемой в нуле, и ее производная определяется формулой

$$(f \circ \gamma)'(0) = df_{\mathbf{p}}(\gamma'(0)).$$

Используя обозначения, принятые для дифференциальных форм в пространстве $\mathbb{R}^2$, в выражении для df заменим $d(x \circ \gamma)$ на

dx и $d(\gamma \circ \gamma)$ на dy. Например, если взять

$$\begin{aligned} f\begin{pmatrix} x \\ y \end{pmatrix} &= x^2 + y^2, \\ df &= 2x\,dx + 2y\,dy, \\ d(f \circ \gamma) &= 2(t\sin t + 1)(\sin t + t\cos t)\,dt + 2e^t \cdot e^t\,dt \\ &= 2(t\sin t + 1)(\sin t + t\cos t)\,dt + 2e^{2t}\,dt. \end{aligned}$$

Коэффициент при dt равен $(f \circ \gamma)'(t)$. Полагая $t = 0$, получаем $(f \circ \gamma)'(0)$.

Заметим, что $(f \circ \gamma)'(0)$ зависит от $\mathbf{p}$ и $\gamma'(0)$, но при этом не требуется другая информация о кривой γ, т.е. эта величина зависит только от касательного вектора $\boldsymbol{\xi}$. Обозначим эту величину $D_{\boldsymbol{\xi}}f$ и будем называть ее *производной по направлению* $\boldsymbol{\xi}$ от функции f. Итак,

$$D_{\boldsymbol{\xi}}f = df_{\mathbf{p}}(\mathbf{v}), \quad \boldsymbol{\xi} = \{\mathbf{p}, \mathbf{v}\}.$$

Например, если $\mathbf{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, то $D_{\boldsymbol{\xi}}f = \dfrac{\partial f}{\partial x}(\mathbf{p})$. Пусть f_1 и f_2 — две функции, дифференцируемые в точке $\mathbf{x}$, и пусть $f = f_1 + f_2$. Кривая γ проходит через точку $\mathbf{x}$, причем ее касательный вектор в $\mathbf{p}$ есть $\boldsymbol{\xi}$. Из курса дифференциального исчисления функций одной переменной мы знаем, что

$$(f \circ \gamma)'(0) = (f_1 \circ \gamma)'(0) + (f_2 \circ \gamma)'(0).$$

Следовательно,

$$\boxed{D_{\boldsymbol{\xi}}(f_1 + f_2) = D_{\boldsymbol{\xi}}f_1 + D_{\boldsymbol{\xi}}f_2.}$$

Аналогично, если мы возьмем $h = f_1 f_2$, то элементарные вычисления дают

$$\begin{aligned} (h \circ \gamma)'(0) &= (f_1 \circ \gamma)'(0)(f_2 \circ \gamma)(0) + (f_1 \circ \gamma)(0)(f_2 \circ \gamma)'(0) \\ &= (f_1 \circ \gamma)'(0) f_2(\mathbf{x}) + f_1(\mathbf{x})(f_2 \circ \gamma)'(0), \end{aligned}$$

так как $(f_1 \circ \gamma)(0) = f_1(\gamma(0)) = f_1(\mathbf{x})$ и то же самое для f_2. Таким образом, мы можем записать

$$\boxed{D_{\boldsymbol{\xi}}(f_1 f_2) = (D_{\boldsymbol{\xi}}f_1)f_2 + f_1(D_{\boldsymbol{\xi}}f_2).}$$

Еще один пример на вычисление производной по направлению. Пусть $\gamma : \mathbb{R} \to \mathbb{R}^2$, где

$$\gamma(t) = \begin{pmatrix} t-1 \\ t^2+2t+2 \end{pmatrix}, \quad f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = x^2y + y^3.$$

Тогда $df_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)}$ и $\boldsymbol{\xi} = \{\gamma(0), \gamma'(0)\}$ определяются формулами

$$df_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)} = (2xy, x^2+3y^2), \quad \boldsymbol{\xi} = \left\{\begin{pmatrix} -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right\}.$$

Отсюда

$$D_{\boldsymbol{\xi}}(f) = df_{\left(\begin{smallmatrix} -1 \\ 2 \end{smallmatrix}\right)}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right) = (-4, 13)\begin{pmatrix} 1 \\ 2 \end{pmatrix} = -4 + 26 = 22.$$

Покажем, что это равно $(f \circ \gamma')(0)$. Для этого заметим, что

$$\begin{aligned} f \circ \gamma(t) &= (t-1)^2(t^2+2t+2) + (t^2+2t+2)^3, \\ (f \circ \gamma)'(t) &= 2(t-1)(t^2+2t+2) + (t-1)^2(2t+2) \\ &\quad + 3(t^2+2t+2)^2(2t+2). \end{aligned}$$

Следовательно,

$$(f \circ \gamma)'(0) = 2(-1)(2) + (-1)^2(2) + 3(2)^2(2) = 22.$$

Вычислим производную по направлению от произведения двух функций. Пусть $g : \mathbb{R}^2 \to \mathbb{R}$, где $g\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = x^2 - y^2$. Тогда отображение произведения $fg : \mathbb{R}^2 \to \mathbb{R}$ имеет вид

$$fg\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) g\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = (x^2y + y^3)(x^2 - y^2).$$

Дифференциалы dg и $d(fg)$ равны

$$\begin{aligned} dg_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)} &= (2x, -2y) \\ d(fg)_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)} &= (2xy(x^2-y^2) + (x^2y+y^3)(2x),\ (x^2+3y^2)(x^2-y^2) \\ &\quad + (x^2y+y^3)(-2y)). \end{aligned}$$

Тогда мы имеем

$$\begin{aligned} D_{\boldsymbol{\xi}}(fg) &= d(fg)_{\binom{-1}{2}}\left(\binom{1}{2}\right) \\ &= (-4(-3)+10(-2),\ 13(-3)+10(-4))\binom{1}{2} \\ &= -8-158=-166. \end{aligned}$$

Согласно правилу дифференцирования произведения, производная $D_{\boldsymbol{\xi}}(fg)$ должна быть равна

$$D_{\boldsymbol{\xi}}(fg) = D_{\boldsymbol{\xi}}(f)g\binom{-1}{2} + f\binom{-1}{2}D_{\boldsymbol{\xi}}(g),$$

где

$$g\binom{-1}{2} = (-1)^2 - 2^2 = -3, \quad f\binom{-1}{2} = (-1)^2 2 + 2^3 = 10$$

$$D_{\boldsymbol{\xi}}(g) = dg_{\binom{-1}{2}}\left(\binom{1}{2}\right) = (-2,\ -4)\binom{1}{2} = -10.$$

Следовательно,

$$D_{\boldsymbol{\xi}}(fg) = 22(-3) + 10(-10) = -166,$$

что совпадает с ранее полученным результатом. Множество касательных векторов в точке $\mathbf{x}$ удобно рассматривать как векторное пространство, называемое *касательным* пространством в точке $\mathbf{x}$ и обозначаемое $TV_{\mathbf{x}}$. Тогда если $\boldsymbol{\xi} = \{\mathbf{x}, \mathbf{v}\}$ и $\boldsymbol{\eta} = \{\mathbf{x}, \mathbf{w}\}$ — два касательных вектора в точке $\mathbf{x}$, то их сумма равна $\boldsymbol{\xi}+\boldsymbol{\eta} = \{\mathbf{x}, \mathbf{v}+\mathbf{w}\}$. Кроме того, если $\boldsymbol{\xi} = \{\mathbf{x}, \mathbf{v}\}$ и a — произвольное вещественное число, то $a\boldsymbol{\xi} = \{\mathbf{x}, a\mathbf{v}\}$. Таким образом, пространство $TV_{\mathbf{x}}$ похоже на пространство V, за исключением того, что все выражения имеют немой индекс $\mathbf{x}$. Сейчас читателю эти обозначения могут показаться слишком громоздкими, но позднее станет ясно, насколько они удобны.

Если $\boldsymbol{\xi} = \{\mathbf{x}, \mathbf{v}\}$ и $\boldsymbol{\eta} = \{\mathbf{x}, \mathbf{w}\}$, то

$$D_{\boldsymbol{\xi}+\boldsymbol{\eta}}f = df_{\mathbf{x}}[\mathbf{v}+\mathbf{w}] = df_{\mathbf{x}}x[\mathbf{v}] + df_{\mathbf{x}}[\mathbf{w}] = D_{\boldsymbol{\xi}}f + D_{\boldsymbol{\eta}}f,$$

следовательно,

$$\boxed{D_{\boldsymbol{\xi}+\boldsymbol{\eta}} f = D_{\boldsymbol{\xi}} f + D_{\boldsymbol{\eta}} f.}$$

Аналогично,

$$\boxed{D_{a\boldsymbol{\xi}} f = a D_{\boldsymbol{\xi}} f.}$$

5.6. Операция переноса

Пусть $\phi : V \to W$, где $\phi(\mathbf{x}) = \mathbf{y}$ — дифференцируемая функция. Если $f : W \to \mathbb{R}$ определена в окрестности $\mathbf{y}$, то функция $f \circ \phi$ определена в окрестности $\mathbf{x}$. Чтобы подчеркнуть точку зрения, которая будет центральной в этой книге, обозначим эту функцию $\phi^* f$ и назовем ее *переносом*[2] функции f с помощью ϕ. Таким образом,

$$\phi^* f = f \circ \phi,$$
$$(\phi^* f)(\mathbf{x}) = f(\phi(\mathbf{x})).$$

Будем считать функцию ϕ фиксированной, а f переменной, так что ϕ^* переводит все функции на пространстве W в функции на пространстве V (назад). Заметим, что

$$\phi^*(f_1 + f_2) = \phi^* f_1 + \phi^* f_2$$

и

$$\phi^*(f_1 f_2) = (\phi^* f_1)(\phi^* f_2),$$

следовательно, ϕ^* сохраняет все алгебраические операции.

Если, например, $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ — отображение, задающее переход от полярных координат к декартовым согласно формуле

$$\phi\left(\begin{pmatrix} r \\ \theta \end{pmatrix}\right) = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix},$$

[2] В английском оригинале эта операция имеет выразительное название *pullback*, что в буквальном переводе означает *отступление, отход назад*. Русский термин *перенос* несколько менее информативен, так как в нем нет указания на то, что область определения функции f под действием переноса перемещается в направлении, *противоположном* перемещению точек при отображении ϕ. — *Прим. ред.*

то можно написать

$$\phi^* x = r\cos\theta, \quad \phi^* y = r\sin\theta.$$

Здесь нам хотелось бы остановиться и объяснить смысл этих уравнений. У нас есть две плоскости: $r\theta$ и xy (рис. 5.9). Пусть отображение ϕ каждой точке плоскости $r\theta$ ставит в соответствие точку на плоскости xy. Мы рассматриваем x как функцию на плоскости xy: каждой точке плоскости эта функция ставит в соответствие координату x.

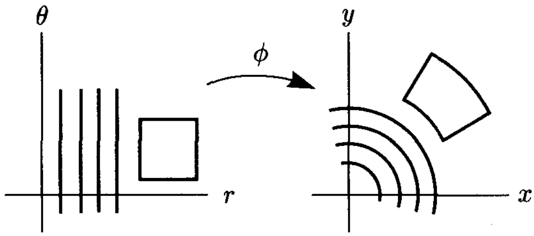


Рис. 5.9

Тогда $\phi^* x$ становится функцией на плоскости $r\theta$, т. е.

$$\phi^* x = r\cos\theta, \quad \text{и, аналогично,} \quad \phi^* y = r\sin\theta.$$

Теперь мы хотим определить перенос с помощью ϕ для дифференциальных форм. Начнем с определения переноса для базисных форм dx и dy. По определению,

$$\phi^* dx = d(\phi^* x) = \cos\theta\, dr - r\sin\theta\, d\theta,$$
$$\phi^* dy = d(\phi^* y) = \sin\theta\, dr + r\cos\theta\, d\theta.$$

Тогда для любой линейной дифференциальной формы типа

$$x^2\, dx + y^2\, dy$$

можно определить

$$\begin{aligned}\phi^*(x^2\, dx + y^2\, dy) &= \phi^*(x^2)\cdot\phi^*\, dx + \phi^*(y^2)\phi^*\, dy \\ &= (r^2\cos^2\theta)(\cos\theta\, dr - r\sin\theta\, d\theta) + (r^2\sin^2\theta)(\sin\theta\, dr + r\cos\theta\, d\theta) \\ &= r^2(\cos^3\theta + \sin^3\theta)\, dr + r^3(\sin^2\theta\cos\theta - \cos^2\theta\sin\theta)\, d\theta.\end{aligned}$$

Другими словами, для дифференциальной формы

$$a\,dx + b\,dy,$$

где a и b — функции, определяем

$$\boxed{\phi^*(a\,dx + b\,dy) = \phi^*(a)\phi^*(dx) + \phi^*(b)\phi^*(dy),}$$

где потом следует собрать вместе коэффициенты при dr и $d\theta$. Заметим, что если f — произвольная функция на плоскости xy, то

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy.$$

Откуда, воспользовавшись цепным правилом, получаем

$$\begin{aligned}
\phi^* df &= \phi^*\left(\frac{\partial f}{\partial x}\right)\phi^*\,dx + \phi^*\left(\frac{\partial f}{\partial y}\right)\phi^*\,dy \\
&= \left(\frac{\partial f}{\partial x}\circ\phi\right) d(x\circ\phi) + \left(\frac{\partial f}{\partial y}\circ\phi)\right) d(y\circ\phi) \\
&= d(f\circ\phi)
\end{aligned}$$

Итак,

$$\boxed{\phi^*\,df = d(\phi^* f).}$$

Эта формула справедлива в общем случае для дифференцируемых отображений $\mathbb{R}^k \to \mathbb{R}^l$.

В качестве примера рассмотрим отображение $\phi : \mathbb{R}^2 \to \mathbb{R}^3$, заданное формулой

$$\phi\left(\begin{pmatrix} r \\ s \end{pmatrix}\right) = \begin{pmatrix} r^2 \\ rs \\ s^2 \end{pmatrix}.$$

Тогда если x, y, z обозначают координаты в пространстве $\mathbb{R}^3$, то

$$\begin{aligned}
\phi^* x &= r^2, \quad \text{отсюда} \quad \phi^*\,dx = 2r\,dr, \\
\phi^* y &= sr, \quad \text{отсюда} \quad \phi^*\,dy = s\,dr + r\,ds, \\
\phi^* z &= s^2, \quad \text{отсюда} \quad \phi^*\,dz = 2s\,ds.
\end{aligned}$$

Для любой функции f в пространстве $\mathbb{R}^3$ справедлива формула

$$df = \frac{\partial f}{\partial x}\, dx + \frac{\partial f}{\partial y}\, dy + \frac{\partial f}{\partial z}\, dz$$

и мы можем вычислить $\phi^*(df)$ двумя способами: вычислять

$$d(\phi^* f)$$

или

$$\phi^*(df) = \phi^*\left(\frac{\partial f}{\partial x}\right)\phi^*\, dx + \phi^*\left(\frac{\partial f}{\partial y}\right)\phi^*\, dy + \phi^*\left(\frac{\partial f}{\partial z}\right)\phi^*\, dz.$$

Например, пусть

$$f\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = y^2 - xz.$$

Тогда

$$df = -z\, dx + 2y\, dy - x\, dz.$$

Подстановка дает

$$\phi^* f = 0$$

и

$$\phi^*\, df = d(\phi^* f) = 0.$$

Если вычислять $\phi^* df$ непосредственно, получаем

$$\begin{aligned} \phi^*\, df &= -s^2 \cdot 2r\, dr + 2sr(s\, dr + r\, ds) - r^2 \cdot 2s\, ds \\ &= [-s^2 \cdot 2r + 2s^2 r]\, dr + [-r^2 \cdot 2s + 2r^2 s]\, ds = 0. \end{aligned}$$

Теперь понятно, как поступать в общем случае. Если $x_1, \ldots, x_k$ — координаты в пространстве $\mathbb{R}^k$ и $y_1, \ldots, y_l$ — координаты в пространстве $\mathbb{R}^l$, то дифференцируемое отображение $\phi : \mathbb{R}^k \to \mathbb{R}^l$ определяется формулой

$$\phi(\mathbf{v}) = \begin{pmatrix} \phi_1(\mathbf{v}) \\ \vdots \\ \phi_l(\mathbf{v}) \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}.$$

Тогда

$$\phi^* y_1 = \phi_1, \quad \phi^*\, dy_1 = d\phi_1 = \frac{\partial\phi_1}{\partial x_1}\, dx_1 + \cdots + \frac{\partial\phi_1}{\partial x_k}\, dx_k,$$
$$\ldots \quad \ldots$$
$$\phi^* y_l = \phi_l, \quad \phi^*\, dy_l = d\phi_l.$$

Операция ϕ^* переносит функции на пространстве $\mathbb{R}^l$ в функции на пространстве $\mathbb{R}^k$, соответственно она переводит линейные дифференциальные формы из пространства $\mathbb{R}^l$ в линейные дифференциальные формы пространства $\mathbb{R}^k$. При этом перенос ϕ^* сохраняет все алгебраические операции: сложение и умножение двух функций, сложение двух форм, умножение функции на форму. Более того, согласно цепному правилу,

$$\phi^*\, df = d(\phi^* f)$$

для любой функции f в пространстве $\mathbb{R}^l$.

Предположим, что мы имеем $\phi : V \to W$ и $\psi : W \to Z$. Можно сделать композицию двух отображений и получить

$$\psi \circ \phi : V \to Z.$$

Если g — произвольная функция на пространстве Z, то можно образовать

$$\psi^* g = g \circ \psi,$$

т. е. функцию на пространстве W, и тогда

$$\phi^*(\psi^* g) = (g \circ \psi) \circ \phi$$

является функцией на пространстве V. Согласно ассоциативному закону, для композиции имеем

$$(g \circ \psi) \circ \phi = g \circ (\psi \circ \phi).$$

Следовательно,

$$(\psi \circ \phi)^* g = \phi^*(\psi^* g),$$

так что

$$\boxed{(\psi \circ \phi)^* = \phi^* \circ \psi^*.}$$

Обратите внимание на обратный порядок функций!

Допустим, что $V = \mathbb{R}^k$, $W = \mathbb{R}^l$ и $Z = \mathbb{R}^m$. В каждом из пространств определены координаты $x^1, \ldots, x^k$; $y^1, \ldots, y^l$; $z^1, \ldots, z^m$ соответственно. Тогда если

$$\omega = a_1\,dz^1 + \cdots + a_m\,dz^m$$

дифференциальная форма в пространстве $\mathbb{R}^m$, то $\phi^*\omega$ будет дифференциальной формой в пространстве $\mathbb{R}^l$, а $\phi^*\psi^*\omega$ – дифференциальной формой в пространстве $\mathbb{R}^k$. Из цепного правила следует, что

$$\phi^*\psi^*\,df = \phi^*(d(\psi^* f)) = d(\phi^*\psi^* f) = d[(\psi \circ \phi)^* f].$$

Операции ϕ^* и ψ^* сохраняют все алгебраические действия, то же самое можно сказать и про $(\psi \circ \phi)^*$,

$$\phi^*\psi^*(g\,df) = (\psi \circ \phi)^*(g\,df).$$

Наиболее общий вид дифференциальной формы представляется суммой членов типа $g\,df$, т. е.

$$a_1\,dz^1 + \cdots + a_m\,dz^m,$$

следовательно, уравнение

$$\boxed{(\psi \circ \phi)^*(\omega) = \phi^*\psi^*\omega} \tag{5.21}$$

справедливо для всех линейных дифференциальных форм.

Пусть $\gamma : \mathbb{R} \to V$ — кривая, проходящая через точку $\mathbf{x}$, т. е. $\gamma(0) = \mathbf{x}$, а кривая $\phi \circ \gamma$ проходит через точку $\mathbf{y}$. Если γ — дифференцируемая функция в нуле, то $\phi \circ \gamma$ тоже дифференцируема в нуле и, согласно цепному правилу,

$$(\phi \circ \gamma)'(0) = d\phi_{\mathbf{x}}(\gamma'(0)).$$

Правая часть этого уравнения зависит только от касательного вектора $\boldsymbol{\xi}$, относящегося к кривой γ. Поэтому $d\phi_{\mathbf{x}}$ отображает касательные векторы в точке $\mathbf{x}$ в касательные векторы в точке $\phi(\mathbf{x})$:

$$d\phi_{\mathbf{x}} : TV_{\mathbf{x}} \to TW_{\phi(\mathbf{x})},$$

где мы определили

$$d\phi_{\mathbf{x}}\boldsymbol{\xi} = \{\phi(\mathbf{x}), d\phi_{\mathbf{x}}(\mathbf{v})\}, \quad \text{если} \quad \boldsymbol{\xi} = \{\mathbf{x}, \mathbf{v}\}.$$

Дифференциал $d\phi_{\mathbf{x}}$ можно истолковать как преобразование инфинитезимальных кривых, проходящих через точку $\mathbf{x}$, в инфинитезимальные кривые, проходящие через $\phi(\mathbf{x})$.

Пусть теперь функция $f : W \to \mathbb{R}$ дифференцируема в $\phi(\mathbf{x})$. По ассоциативному закону для композиции имеем:

$$(\phi^* f) \circ \gamma = f \circ \phi \circ \gamma = f \circ (\phi \circ \gamma).$$

Дифференцируя это уравнение при $t = 0$, получаем

$$\boxed{D_{\boldsymbol{\xi}}(\phi^* f) = D_{d\phi_{\mathbf{x}}\boldsymbol{\xi}} f.} \tag{5.22}$$

Это значит, что мы можем перенести функцию f назад с помощью ϕ и потом вычислить производную по направлению $\boldsymbol{\xi}$, или мы можем пройти вперед с помощью $d\phi_{\mathbf{x}}$ и потом взять производную по направлению от f по $d\phi_{\mathbf{x}}\boldsymbol{\xi}$. В обоих случаях мы получим один и тот же ответ.

Положив $\boldsymbol{\xi} = \{\mathbf{x}, \mathbf{v}\}$, полученное равенство можно переписать в виде

$$D_{\{\mathbf{x},\mathbf{v}\}}(\phi^* f) = D_{d\phi_{\mathbf{x}}(\mathbf{x}), d\phi_{\mathbf{x}}[\mathbf{v}]}(f),$$

или, равносильно,

$$d(f \circ \phi)_{\mathbf{x}}[\mathbf{v}] = df_{\phi(\mathbf{x})}[d\phi_{\mathbf{x}}[\mathbf{v}]].$$

Последнее равенство означает, что

$$d(f \circ \phi)_{\mathbf{x}} = df_{\phi(\mathbf{x})} \circ d\phi_{\mathbf{x}},$$

а это не что иное, как частный случай цепного правила.

В качестве примера на применение этого равенства возьмем $f : W \to \mathbb{R}$ и $\phi : V \to W$, определяемые формулами

$$f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = x^2 y, \quad \phi\left(\begin{pmatrix} r \\ \theta \end{pmatrix}\right) = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}.$$

Тогда

$$df_{\binom{x}{y}} = (2xy,\, x^2),$$
$$d\phi_{\binom{r}{\theta}} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix},$$

и мы получаем, что

$$\begin{aligned} D_{\left\{\phi\left(\binom{r}{\theta}\right), d\phi\left(\binom{r}{\theta}\right)[\mathrm{v}]\right\}}(f) &= df_{\phi\binom{r}{\theta}} \circ d\phi_{\binom{r}{\theta}}[v] \\ &= (2(r\cos\theta)(r\sin\theta),\, (r\cos\theta)^2)\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}\begin{pmatrix} v_r \\ v_\theta \end{pmatrix} \\ &= (2r^2\cos\theta\sin\theta,\, r^2\cos^2\theta)\begin{pmatrix} v_r\cos\theta - v_\theta r\sin\theta \\ v_r\sin\theta + rv_\theta\cos\theta \end{pmatrix} \\ &= 2r^2\cos\theta\sin\theta(v_r\cos\theta - rv_\theta\sin\theta) + r^2\cos^2\theta(v_r\sin\theta + rv_\theta\cos\theta) \\ &= 3r^2\cos^2\theta\sin\theta\, v_r + r^3(-2\cos\theta\sin^2\theta + \cos^3\theta)v_\theta. \end{aligned}$$

Теперь надо показать, что последнее выражение равно

$$D_{\left\{\binom{r}{\theta}, \mathrm{v}\right\}}(\phi^* f).$$

Для этого заметим, что

$$\phi^* f\left(\begin{pmatrix} r \\ \theta \end{pmatrix}\right) = f\left(\begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}\right) = (r\cos\theta)^2 r\sin\theta = r^3\cos^2\theta\sin\theta.$$

Следовательно,

$$d(\phi^* f)_{\binom{r}{\theta}} = (3r^2\cos^2\theta\sin\theta,\, r^3(2\cos\theta(-\sin\theta)\sin\theta + \cos^3\theta)).$$

Отсюда, наконец, получаем:

$$\begin{aligned} D_{\left\{\binom{r}{\theta}, \mathrm{v}\right\}}(\phi^* f) = d(\phi^* f)_{\binom{r}{\theta}}\left(\begin{pmatrix} v_r \\ v_\theta \end{pmatrix}\right) = \\ = 3r^2\cos^2\theta\sin\theta\, v_r + r^3(-2\cos\theta\sin^2\theta + \cos^3\theta)v_\theta. \end{aligned}$$

Резюме

A. Дифференциалы и частные производные

Вы должны уметь давать определение дифференциала df функции f в терминах «o» и «O».

Вы должны знать и уметь применять правила дифференцирования суммы, произведения и композиции функций.

Вы должны уметь выражать дифференциал функции через частные производные и уметь строить матрицу, представляющую дифференциал функции $f : \mathbb{R}^2 \to \mathbb{R}^2$.

B. Преобразования координат

Для заданного преобразования, которое вводит новые координаты на плоскости, надо уметь пользоваться цепным правилом для выражения дифференциалов и частных производных в этих новых координатах.

C. Применения дифференциалов

Вы должны знать, как получить уравнение прямой или плоскости, касательной к графику функции в выбранной точке.

Вы должны уметь пользоваться цепным правилом для решения задач с участием функций от нескольких переменных.

Задачи

5.1. Покажите, что если $f : V \to W$ дифференцируема в точке $\boldsymbol{\alpha}$ и $T : W \to Z$ линейна, то композиция $T \circ f$ дифференцируема в точке $\boldsymbol{\alpha}$ и
$$d(T \circ f)_{\boldsymbol{\alpha}} = T \circ df_{\boldsymbol{\alpha}}.$$

5.2. Пусть $F : V \to \mathbb{R}$ дифференцируема в $\boldsymbol{\alpha}$ и функция $f : \mathbb{R} \to \mathbb{R}$ имеет производную в $a = F(\boldsymbol{\alpha})$. Докажите, что $f \circ F$ дифференцируема в точке $\boldsymbol{\alpha}$ и
$$d(f \circ F)_{\boldsymbol{\alpha}} = f'(a)\, dF_{\boldsymbol{\alpha}}.$$

5.3. Пусть $F : V \to W$ и $G : W \to V$ — непрерывные отображения, причем $G \circ F(\mathbf{v}) = \mathbf{v}$ и $F \circ G(\mathbf{w}) = \mathbf{w}$ для всех $\mathbf{v}$ в пространстве

V и для всех $\mathbf{w}$ в пространстве W. Допустим, что F дифференцируема в точке $\boldsymbol{\alpha}$, а G дифференцируема в точке $\boldsymbol{\beta} = F(\boldsymbol{\alpha})$. Докажите, что

$$dG_{\boldsymbol{\beta}} = (dF_{\boldsymbol{\alpha}})^{-1}.$$

5.4. Пусть функция $f : V \to \mathbb{R}$ дифференцируема в точке $\boldsymbol{\alpha}$. Покажите, что $g = f^n$ дифференцируема в точке $\boldsymbol{\alpha}$ и что

$$dg_{\boldsymbol{\alpha}} = nf^{n-1}\, df_{\boldsymbol{\alpha}}.$$

5.5. Пусть $\gamma : \mathbb{R} \to \mathbb{R}^2$ обозначает кривую $\gamma(t) = \begin{pmatrix} e^t \\ \sin t \end{pmatrix}$. Пусть $F : \mathbb{R}^2 \to \mathbb{R}^2$ — отображение, где

$$F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} 3x^2y \\ x^2y^3 \end{pmatrix}.$$

(a) Вычислите вектор, касательный к кривой γ при $t = 0$ и $t = \pi/2$.

(b) Найдите производную по направлению отображения F для каждого из полученных касательных векторов.

5.6. Пусть $f : \mathbb{R}^2 \to \mathbb{R}^2$ и $g : \mathbb{R}^2 \to \mathbb{R}^2$ определены формулами

$$f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x^2y \\ y^3 \end{pmatrix}, \quad g\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} \cos xy \\ \sin xy \end{pmatrix}.$$

Проверьте цепное правило для отображения $g \circ f : \mathbb{R}^2 \to \mathbb{R}^2$.

5.7. Пусть $g : V \to W$ — отображение, где $g\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x^2e^y \\ \cos xy \end{pmatrix}$, и $\lambda : \mathbb{R} \to V$ — прямая линия

$$\lambda(t) = \begin{pmatrix} 3 \\ -1 \end{pmatrix} + t\begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

(a) Найдите касательный вектор в точке $\begin{pmatrix} 3 \\ -1 \end{pmatrix}$ для кривой $g \circ \lambda$.

(b) Производную по направлению $D_{\left(\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right),\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)}(g)$ вычислите двумя способами.

5.8. Пусть $\phi : V \to W$ — отображение, задаваемое формулой $\phi : \begin{pmatrix} r \\ \theta \end{pmatrix} \mapsto \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}$, а функция $f : W \to \mathbb{R}$ определена формулой $f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto x^3y^4$. Проверьте, что

$$D_{\boldsymbol{\xi}}(\phi^* f) = D_{d\phi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})} f$$

для всех касательных векторов $\boldsymbol{\xi} = (\boldsymbol{\alpha}, \mathbf{v})$.

5.9. Определим отображения $F : \mathbb{R}^2 \to \mathbb{R}^2$, $G : \mathbb{R}^2 \to \mathbb{R}^2$, $f : \mathbb{R}^1 \to \mathbb{R}^2$ и $g : \mathbb{R}^2 \to \mathbb{R}^1$ равенствами

$$F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x^3 + y^2 \\ xy \end{pmatrix}, \quad G\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} 2x^2y \\ y^2 \end{pmatrix},$$

$$f(t) = \begin{pmatrix} t^2 + 1 + \cos t \\ 3t + 2 \end{pmatrix}, \quad g\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = (x^3y).$$

Проверьте, что:

(a) $d(F \circ G)_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)} = dF_{G\left(\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)\right)} \circ dG_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)}$

(b) $d(G \circ F)_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)} = dG_{F\left(\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)\right)} \circ dF_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)}$

(c) $d(g \circ f)_{(t)} = dg_{f(t)} \circ dF_{(t)}$

(d) $d(f \circ g)_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)} = df_{g\left(\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)\right)} \circ dg_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)}$.

5.10. Пусть функция $f : \mathbb{R}^2 \to \mathbb{R}$ дифференцируема в некоторой окрестности $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ и удовлетворяет условию $f\left(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}\right) \neq 0$. Покажите, что в некоторой окрестности $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ отображение $g\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = 1 \Big/ f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right)$ дифференцируемо и что

$$dg_{\left(\begin{smallmatrix} x_0 \\ y_0 \end{smallmatrix}\right)} = -df_{\left(\begin{smallmatrix} x_0 \\ y_0 \end{smallmatrix}\right)} \Big/ \left(f\left(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}\right)\right)^2.$$

5.11. Функция f определена на плоскости через аффинные координаты x и y следующим образом:

$$f(P) = \sqrt{|x(P)y(P)|}.$$

(a) *Непрерывна* функция f в начале координат $P_0(x = 0, y = 0)$ или нет? Дайте точный ответ через определение непрерывности.

(b) *Дифференцируема* ли функция f в начале координат? Дайте точный ответ через определение дифференцируемости.

5.12. Так называемые параболические координаты на плоскости выражаются через декартовы координаты x и y согласно формуле

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \sqrt{x^2 + y^2} - x \\ \sqrt{x^2 + y^2} + x \end{pmatrix}.$$

(a) Выразите $\begin{pmatrix} du \\ dv \end{pmatrix}$ через $\begin{pmatrix} dx \\ dy \end{pmatrix}$, используя матрицы 2×2, потом обратите эти матрицы и выразите $\begin{pmatrix} dx \\ dy \end{pmatrix}$ через $\begin{pmatrix} du \\ dv \end{pmatrix}$.

(b) Получите обратное преобразование координат, выразив $\begin{pmatrix} x \\ y \end{pmatrix}$ через $\begin{pmatrix} u \\ v \end{pmatrix}$. Выразите дифференциал $\begin{pmatrix} dx \\ dy \end{pmatrix}$ через $\begin{pmatrix} du \\ dv \end{pmatrix}$.

(c) Покажите, что кривые $u = \text{const}$ и $v = \text{const}$ являются параболами, перпендикулярными друг другу в точке пересечения. Нарисуйте семейства этих кривых.

(d) Рассмотрите функцию f, определенную на плоскости формулой $f(\mathbf{p}) = 1/(u(\mathbf{p}) + v(\mathbf{p}))$. Выразите $d_{\mathbf{q}}f$, где $\mathbf{q}$ — точка с координатами $u = 4$ и $v = 16$, сначала через du и dv, а потом через dx и dy.

(e) Предположим, что траектория движущейся частицы определена формулами: $\boldsymbol{\alpha} : \mathbb{R} \to \mathbb{R}^2$, где

$$\begin{pmatrix} dx \\ dy \end{pmatrix} \circ \boldsymbol{\alpha} = \begin{pmatrix} t^2 + t \\ t^3 \end{pmatrix}.$$

Вычислите производные от $x \circ \boldsymbol{\alpha}$, $y \circ \boldsymbol{\alpha}$, $u \circ \boldsymbol{\alpha}$ и $v \circ \boldsymbol{\alpha}$ при $t = 2$.

5.13. Пусть координаты на плоскости u и v связаны с декартовыми координатами x и y уравнениями

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2uv \\ u^2 - v^2 \end{pmatrix}.$$

(a) Вычислите производную (матрицу Якоби) этого преобразования в точке $u = 2, v = 1$ (т. е. выразите $\begin{pmatrix} dx \\ dy \end{pmatrix}$ через $\begin{pmatrix} du \\ dv \end{pmatrix}$ в этой точке).

(b) Рассмотрите функцию $f(u, v) = u^2v^3$. В координатах x и y получите уравнение касательной к кривой $f(u, v) = 4$ в точке $u = 2, v = 1$ (т. е. $x = 4, y = 3$). (Не пытайтесь искать u и v как функции x и y; просто используйте цепное правило.)

(c) Предположим, что частица движется вдоль траектории

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} t^3/2 \\ 3t^2/4 \end{pmatrix}.$$

В момент времени $t = 2$ частица проходит через точку $\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. С какой скоростью изменяются координаты u и v, т. е. чему равны du/dt и dv/dt в этот момент?

5.14. Пусть $\mathbb{A}$ обозначает аффинную плоскость. Точка P_0 лежит на этой плоскости. Постройте функцию $f : \mathbb{A} \to \mathbb{R}$, для которой $f(P_0) = 0$, такую, что для произвольных аффинных координат $s(P)$, $t(P)$ частные производные $(\partial f/\partial s)_t$ и $(\partial f/\partial t)_s$ определены и равны нулю в точке P_0, но при этом f не дифференцируема в P_0. (Указание: если заменить термины «дифференцируема» на термин «непрерывна» и «частная производная» на «предел функции от одной переменной», то ответ был бы

$$f(P) = \begin{cases} 0 & \text{в точке} \quad P_0, \\ \dfrac{x^2y}{x^4 + y^2} & \text{в других точках,} \end{cases}$$

где $x(P_0)$ и $y(P_0)$ равны нулю.)

5.15. Для некоторого кратера высота над уровнем моря определяется формулой

$$z(x, y) = \sqrt{x^2 + 4y^2} \quad (x,\ y,\ z \quad \text{измерены в километрах}).$$

Температура по Фаренгейту находится из уравнения

$$T(x, y) = 100 + 2x - \frac{1}{4}x^2y^2.$$

(a) Выразите dz и dT через dx и dy в точке $x = 3$, $y = 2$.

(b) Получите уравнение касательной плоскости к кратеру в точке $x = 3, y = 2$.

(c) Вдоль какого направления температура изменяется быстрее всего в точке $x = 3$, $y = 2$? Если двигаться вдоль этого направления, то с какой скоростью изменяется температура в зависимости от высоты над уровнем моря (с точностью до градуса на километр)?

Глава 6

Теоремы многомерного дифференциального исчисления

В этой главе мы продолжаем изучение дифференциального исчисления. Приводим векторные варианты теоремы о среднем значении, формулы Тейлора и теоремы существования обратной функции. Обсуждаем поведение в окрестности критической точки и множители Лагранжа. Эту главу можно прочитать быстро, не очень вникая в детали доказательств, но упражнения надо проделать.

6.1. Теорема о среднем значении

Это одна из немногих теорем, которые для функций нескольких переменных мы не сможем доказать с такой же степенью строгости, как это делается для функций, зависящих только от одной переменной. Сначала вспомним формулировку теоремы для случая одной переменной. Утверждается, что если f — непрерывно дифференцируемая функция на некотором отрезке $[a, b]$, то существует точка z, лежащая внутри этого отрезка, такая, что

$$f(b) - f(a) = f'(z)(b - a). \tag{6.1}$$

Значение z в общем случае трудно определить явно, поэтому обычно теорема о среднем значении используется в несколько ином виде, как неравенство:

$$\text{Если } f'(x) \le m \text{ для всех } x \in [a,b], \text{ то } f(b) - f(a) \le m(b-a). \tag{6.2}$$

Так как $f'(z) \le m$, то это неравенство немедленно следует из теоремы о среднем значении, сформулированной выше. В то же время его легко доказать непосредственно с помощью фундаментальной теоремы интегрального исчисления. Действительно,

$$f(b) - f(a) = \int_a^b f'(s)\,ds \le \int_a^b m\,ds \le m(b-a). \tag{6.3}$$

Чтобы обобщить теорему на случай большего числа измерений, удобно немного переписать аргумент функции в (6.3):

$$\begin{aligned} f(b) - f(a) &= \int_0^1 \frac{d}{dt} f(a + t(b-a))\,dt = \int_0^1 f'(a + t(b-a))(b-a)\,dt \\ &\le (b-a) \int_0^1 f'(a+t(b-a))\,dt \le m(b-a). \end{aligned} \tag{6.4}$$

(Заметим, что второе равенство следует из цепного правила.) Преимущество записей (6.2) или (6.4) по сравнению с (6.1) состоит в том, что их можно сразу же распространить на случай, когда f является отображением из $\mathbb{R}$ в $\mathbb{R}^k$. Предположим, что f — именно такое отображение:

$$f(t) = \begin{pmatrix} f_1(t) \\ \vdots \\ f_k(t) \end{pmatrix}.$$

Тогда $f'(t) = \lim_{h\to 0}(1/h)(f(t+h) - f(t)) = df_t[1]$. Здесь мы рассматриваем 1 как вектор в пространстве $\mathbb{R}^1$, если использовать ранее введенное обозначение. Очевидно, что вектор $f'(t)$ определяется формулой

$$f'(t) = \begin{pmatrix} f_1'(t) \\ \vdots \\ f_k'(t) \end{pmatrix}.$$

Для этого вектора можно написать равенство, аналогичное (6.4), т. е.

$$f(b)-f(a)=\int_0^1 \frac{d}{dt}f(a+t(b-a))\,dt=\left(\int_0^1 f'(a+t(b-a))\,dt\right)(b-a). \tag{6.5}$$

Под интегралом от векторной функции g мы подразумеваем вектор, компонентами которого будут интегралы от его компонент:

$$\text{если}\quad g=\begin{pmatrix} g_1 \\ \vdots \\ g_k \end{pmatrix},\quad \text{то}\quad \int g(t)\,dt=\begin{pmatrix} \int g_1(t)\,dt \\ \vdots \\ \int g_k(t)\,dt \end{pmatrix}.$$

Конечно, у нас есть определение интеграла в виде предела аппроксимирующей суммы:

$$\int_0^1 g(t)\,dt=\lim_{n\to\infty}(1/n)\sum_{i=1}^{n} g(i/n).$$

Поскольку каждая компонента этой аппроксимирующей суммы векторов есть аппроксимирующая сумма для интеграла от соответствующей компоненты функции, то оба определения интеграла для векторной функции от одной переменной совпадают. Для векторов существует неравенство $\|\mathbf{v}_1+\mathbf{v}_2\|\le\|\mathbf{v}_1\|+\|\mathbf{v}_2\|$. Аналогичное неравенство можно написать для аппроксимирующей суммы и, переходя к пределу, мы получаем неравенство для интеграла

$$\left\|\int g(t)\,dt\right\|\le\int\|g(t)\|\,dt.$$

Положим $g=f'$ и подставим в (6.5). Тогда

$$\|f(b)-f(a)\|\le m(b-a),\quad \text{если}\quad \|f'(t)\|\le m\quad \text{для всех}\quad t\in[a,b]. \tag{6.6}$$

Именно здесь можно увидеть трудность, возникающую при попытке обобщить (6.1) на функцию, определенную в пространстве $\mathbb{R}^k$. Мы можем применить (6.1) к каждой компоненте f_j функции f. Для каждой компоненты мы получим

$$f_j(b)-f_j(a)=f_j'(z_j)(b-a).$$

Однако значения z_j могут быть разными для разных j. Вообще говоря, не существует такой точки z, которая подошла бы для всех компонент f_j. Поэтому аналогия с (6.1) в многомерном случае не проходит. Что касается формулы (6.5), она остается верной.

Давайте обобщим утверждение (6.6) на случай, когда f есть отображение $V \to W$, при этом V — не обязательно одномерное пространство. Мы уже знаем, что обобщением $f'(x)$ является df_x. Пусть df_x — линейное преобразование. Нам необходимо понять, что значит $\|A\|$, если A — линейное преобразование из V в W. По определению,

$$\|A\| = \max_{\|\mathbf{u}\|=1} \|A\mathbf{u}\|$$

или, что то же самое,

$$\|A\| = \max_{\mathbf{v}\neq\mathbf{0}} \frac{\|A\mathbf{v}\|}{\|\mathbf{v}\|}.$$

Тогда $\|A\mathbf{v}\| \leq \|A\| \cdot \|\mathbf{v}\|$ для всех $\mathbf{v}$ и $\|A\|$ будет наименьшим числом, обладающим этим свойством[1], т. е. если

$$\|A\mathbf{v}\| \leq k\|\mathbf{v}\| \quad \text{для всех} \quad \mathbf{v}, \quad \text{то} \quad \|A\| \leq k.$$

Если A_1 и A_2 — два линейных преобразования из V и W, то

$$\|(A_1+A_2)\mathbf{v}\| = \|A_1\mathbf{v}+A_2\mathbf{v}\| \leq \|A_1\mathbf{v}\|+\|A_2\mathbf{v}\| \leq (\|A_1\|+A_2\|)\cdot\|\mathbf{v}\|,$$

откуда

$$\|A_1 + A_2\| \leq \|A_1\| + \|A_2\|. \tag{6.7}$$

Пусть $[\mathbf{a}, \mathbf{b}]$ обозначает отрезок прямой, соединяющей любые точки $\mathbf{a}$ и $\mathbf{b}$ в пространстве V (рис. 6.1). Это значит, что отрезок $[\mathbf{a}, \mathbf{b}]$ состоит из всех точек вида $\mathbf{a}+t(\mathbf{b}-\mathbf{a})$ для $0 \leq t \leq 1$ (естественное обобщение одномерных обозначений). Мы хотим доказать следующее. Предположим, что $f : V \to W$ дифференцируема во всех точках $[\mathbf{a}, \mathbf{b}]$, и дифференциал функции $df_{\mathbf{x}}$ является непрерывной функцией от $\mathbf{x}$ на этом отрезке. Далее, предположим, что

$$\|df_{\mathbf{x}}\| \leq m \quad \text{для всех} \quad \mathbf{x} \in [\mathbf{a}, \mathbf{b}].$$

[1]Число $\|A\|$ называют *нормой* линейного преобразования $A : V \to W$. — *Прим. ред.*

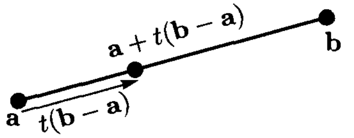


Рис. 6.1

Тогда
$$\|f(\mathbf{b}) - f(\mathbf{a})\| \le m\|\mathbf{b} - \mathbf{a}\|. \tag{6.8}$$

Доказательство. Пусть $h : [0,1] \to [\mathbf{a}, \mathbf{b}]$ — отображение, заданное формулой
$$h(t) = \mathbf{a} + t(\mathbf{b} - \mathbf{a}).$$

Обозначим $F = f \circ h$, так что $F(t) = f(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))$. Тогда F — дифференцируемое отображение $[0,1] \to W$ и, согласно цепному правилу,
$$dF_t = df_{h(t)} \circ dh_t.$$

Далее, $dh_t[1] = \mathbf{b} - \mathbf{a}$, значит,
$$F'(t) = dF_t[1] = df_{h(t)}[\mathbf{b} - \mathbf{a}].$$

Отсюда
$$f(\mathbf{b}) - f(\mathbf{a}) = F(1) - F(0) = \int_0^1 F'(t)\,dt = \int_0^1 df_{h(t)}[\mathbf{b} - \mathbf{a}]\,dt.$$

Следовательно,
$$f(\mathbf{b}) - f(\mathbf{a}) = A(\mathbf{b} - \mathbf{a}), \tag{6.9}$$

где A — линейное преобразование
$$A = \int_0^1 df_{h(t)}\,dt. \tag{6.10}$$

В формуле (6.10) мы интегрируем функцию, которая каждому значению t приписывает линейное преобразование $df_{h(t)}$. Мы можем толковать эти интегралы точно так же, как мы толковали векторные интегралы. Например, так как в пространствах V и W имеются стандартные базисы, то каждое линейное преобразование можно определять матрицей. Интеграл от матричной функции g, где $g(t) = (g_{ij}(t))$, задается матрицей, у которой

ij-й матричный элемент есть интеграл от численной функции g_{ij}. Или, как и раньше, интеграл можно рассматривать как предел от аппроксимирующих сумм. Тогда из (6.7) следует, что

$$\left\| \int g(t)\,dt \right\| \le \int \|g(t)\|\,dt.$$

В частности, выполнив такую оценку (6.10) и предположив, что $\|df_{\mathbf{x}}\| \le m$ для всех $\mathbf{x} \in [\mathbf{a}, \mathbf{b}]$, можно сделать вывод:

$$\|A\| \le m,$$

и с учетом свойств нормы $\|A\|$ из равенства (6.9) получаем справедливость оценки (6.8).

6.2. Производные высших порядков и формула Тейлора

Пусть $f : \mathbb{R}^2 \to \mathbb{R}$ — дифференцируемая функция. Если частные производные $\partial f/\partial x$ и $\partial f/\partial y$ непрерывно дифференцируемы, то можно образовать конструкции

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right), \quad \text{обозначается} \quad \frac{\partial^2 f}{\partial x^2},$$

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right), \quad \text{обозначается} \quad \frac{\partial^2 f}{\partial y^2},$$

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right), \quad \text{обозначается} \quad \frac{\partial^2 f}{\partial x \partial y}.$$

Мы уже знаем, что

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial y}\frac{\partial f}{\partial x}.$$

Аналогично можно определить частные производные более высоких порядков, если они существуют, и получить соответствующие равенства для смешанных производных. Например,

$$\frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\right) = \frac{\partial}{\partial y}\left(\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)\right) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)\right)$$

и так далее. Значение производных второго и более высоких порядков связано с формулой Тейлора, которую мы сейчас получим и докажем.

Для упрощения обозначений пока сформулируем ее для начала координат. Предположим, что f имеет непрерывные производные первого и второго порядков в окрестности $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Для удобства мы будем писать $f(x, y)$ вместо $f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right)$. Тогда

$$\begin{aligned} f(x, y) &= f(0,0) + \int_0^1 \frac{d}{dt} f(tx, ty)\, dt \\ &= f(0,0) + \int_0^1 \left(\frac{\partial f}{\partial x}(tx, ty)x + \frac{\partial f}{\partial y}(tx, ty)y \right) dt. \end{aligned}$$

Пусть

$$f_1(x, y) = \int_0^1 \frac{\partial f}{\partial x}(tx, ty)\, dt \quad \text{и} \quad f_2(x, y) = \int_0^1 \frac{\partial f}{\partial y}(tx, ty)\, dt.$$

Тогда

$$f(x, y) = f(0,0) + x f_1(x, y) + y f_2(x, y),$$

где f_1 и f_2 — дифференцируемые[2] функции, или, более компактно,

$$f = f(0,0) + x f_1 + y f_2.$$

Отметим, что

$$f_1(0,0) = \frac{\partial f}{\partial x}(0,0) \quad \text{и} \quad f_2(0,0) = \frac{\partial f}{\partial y}(0,0).$$

Те же самые соображения применим для f_1 и f_2:

$$f_1(x, y) = f_1(0,0) + x f_{11}(x, y) + y f_{12}(x, y),$$

где

$$f_{11}(x, y) = \int_0^1 \frac{\partial f_1}{\partial x}(tx, ty)\, dt \quad \text{и} \quad f_{12}(x, y) = \int_0^1 \frac{\partial f_1}{\partial y}(tx, ty)\, dt.$$

[2]Разумеется, это требует доказательства. — *Прим. ред.*

Аналогично

$$f_2(x,y) = f_2(0,0) + xf_{21}(x,y) + yf_{22}(x,y).$$

Итак,

$$f = f(0,0) + xf_1(0,0) + yf_2(0,0) + x^2 f_{11} + xy(f_{12} + f_{21}).$$

Если f имеет непрерывные[3] производные вплоть до третьего порядка, мы можем повторить процедуру еще раз и получить:

$$\begin{aligned} f = {} & f(0,0) + xf_1(0,0) + yf_2(0,0) + x^2 f_{11}(0,0) \\ & + xy(f_{12}(0,0) + f_{21}(0,0)) + y^2 f_{22}(0,0) \\ & + x^3 f_{111} + x^2 y(f_{112} + f_{121} + f_{211}) \\ & + xy^2(f_{122} + f_{212} + f_{221}) + y^3 f_{222}, \end{aligned}$$

где все функции f_{111}, f_{112}, и т. д. непрерывны. Вычислив вторые производные от обеих частей этого уравнения, получаем

$$\begin{aligned} 2f_{11}(0,0) &= \frac{\partial^2 f}{\partial x^2}(0,0), \\ f_{12}(0,0) + f_{21}(0,0) &= \frac{\partial f}{\partial x \partial y}(0,0), \\ 2f_{22}(0,0) &= \frac{\partial f^2}{\partial y^2}(0,0). \end{aligned}$$

Итак, мы доказали, что

$$\begin{aligned} f(x,y) = {} & f(0,0) + x\frac{\partial f}{\partial x}(0,0) + y\frac{\partial f}{\partial y}(0,0) + \frac{1}{2}x^2\frac{\partial^2 f}{\partial x^2}(0,0) \\ & + xy\frac{\partial^2 f}{\partial x \partial y}(0,0) + \frac{1}{2}y^2\frac{\partial^2 f}{\partial y^2}(0,0) + O\left(\left\|\begin{pmatrix} x \\ y \end{pmatrix}\right\|^3\right). \end{aligned} \tag{6.11}$$

[3] Авторы опускают здесь доказательство непрерывности, так же, как ниже в формуле (6.11) опускают оценку остатка, записанного сразу как $O\left(\left\|\begin{pmatrix} x \\ y \end{pmatrix}\right\|^3\right)$. — *Прим. ред.*

Очевидно, что если f имеет непрерывные производные более высоких порядков, то эту процедуру можно продолжать и дальше. Понятно также, что подобные соображения применимы не только в $\mathbb{R}^2$, но и в произвольном пространстве $\mathbb{R}^k$. Наконец, начало координат $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ можно заменить любым вектором $\mathbf{u}$ и $\begin{pmatrix} x \\ y \end{pmatrix}$ заменить на $\mathbf{u} + \mathbf{v}$. В результате придем к следующей теореме:

Пусть функция $f : \mathbb{R}^k \to \mathbb{R}$ имеет непрерывные частные производные вплоть до порядка $n+1$. Тогда существует такой полином P_n, зависящий от координат вектора $\mathbf{v}$, что

$$f(\mathbf{u} + \mathbf{v}) = P_n(\mathbf{v}) + O(\|\mathbf{v}\|^{n+1}).$$

Коэффициенты $P_n(\mathbf{v})$ определяются путем последовательного дифференцирования по $\mathbf{v}$ этого тождества и последующей подстановки $\mathbf{v} = \mathbf{0}$.[4]

Если $f : \mathbb{R}^2 \to \mathbb{R}^1$, то матрица, составленная из частных производных второго порядка,

$$H_{\mathbf{p}}(f) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2}(\mathbf{p}) & \dfrac{\partial^2 f}{\partial x \partial y}(\mathbf{p}) \\ \dfrac{\partial^2 f}{\partial x \partial y}(\mathbf{p}) & \dfrac{\partial^2 f}{\partial y^2}(\mathbf{p}) \end{pmatrix}$$

называется матрицей *Гессе*, а соответствующая квадратичная форма называется *гессианом* и обозначается $d^2 f_{\mathbf{p}}$. Если функция f и точка $\mathbf{p}$ зафиксированы, матрицу $H_{\mathbf{p}}(f)$ будем обозначать просто H. Итак, формулу (6.11) можно переписать в форме

$$f(\mathbf{p} + \mathbf{v}) = f(\mathbf{p}) + df_{\mathbf{p}}(\mathbf{v}) + \frac{1}{2} d^2 f_{\mathbf{p}}(\mathbf{v}) + o(\|\mathbf{v}\|^2), \tag{6.12}$$

где

$$d^2 f_{\mathbf{p}} = \mathbf{v}^T H \mathbf{v} = (v_1, v_2) H \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}. \tag{6.13}$$

[4]Можно доказать, что в условиях этой теоремы все частные производные от остатка $O\left(\|\mathbf{v}\|^{n+1}\right)$ до порядка n включительно при $\mathbf{v} = 0$ равны нулю. — *Прим. ред.*

Гессиан $d^2 f_{\mathbf{p}}$ является квадратичной формой, и поэтому к нему применим анализ, изложенный в главе 4. Например, если $f(P) = [x(P)]^2 y(P)$, то в точке $x = 2$, $y = 3$ частные производные равны

$$\frac{\partial f}{\partial x} = 2xy = 12, \quad \frac{\partial f}{\partial y} = x^2 = 4,$$
$$\frac{\partial^2 f}{\partial x^2} = 2y = 6, \quad \frac{\partial^2 f}{\partial x \partial y} = 2x = 4, \quad \frac{\partial^2 f}{\partial y^2} = 0.$$

Следовательно, дифференциал df в точке $\mathbf{p} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ определяется числами $(12, 4)$, а гессиан — матрицей $\begin{pmatrix} 6 & 4 \\ 4 & 0 \end{pmatrix}$.

6.3. Максимум и минимум

Гессиан особенно полезен при анализе поведения функции в окрестности критической точки, т. е. точки, где df равен нулю. Если $\mathbf{p}_0 = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ — критическая точка, то

$$f(\mathbf{p}_0 + \mathbf{v}) = f(\mathbf{p}_0) + \frac{1}{2} d^2 f_{\mathbf{p}}(\mathbf{v}).^5$$

Если квадратичная форма $d^2 f_{\mathbf{p}}$ *положительно определена* (в этом случае H имеет два положительных собственных значения), то из формулы Тейлора следует, что $f(\mathbf{p}_0 + \mathbf{v}) > f(\mathbf{p}_0)$ для достаточно малых $\mathbf{v} \neq \mathbf{0}$ и функция f имеет *минимум* в точке $\mathbf{p}_0$. Если $d^2 f_{\mathbf{p}}$ *отрицательно определена* (H имеет два отрицательных собственных значения), то $f(\mathbf{p}_0 + \mathbf{v}) < f(\mathbf{p}_0)$ и f достигает *максимума* в $\mathbf{p}_0$. Наконец, если собственные значения H имеют

[5] Авторы опустили в этой формуле остаток $O\left(\left\|\begin{pmatrix} x \\ y \end{pmatrix}\right\|^3\right)$, так что написанное равенство — приближенное, а не точное. Поэтому, хотя сформулированные в следующем абзаце утверждения верны, приведенные рассуждения следует считать лишь наводящими соображениями, а не строгим выводом достаточных условий максимума, минимума и седловой точки. — *Прим. ред.*

разные знаки, то $d^2 f(\mathbf{v})$ может иметь и положительные, и отрицательные значения. В этом случае точка $\mathbf{p}_0$ является *седловой* точкой функции f. Если одно или два собственных значения H равны нулю, т. е. гессиан — сингулярный, то чтобы установить, будет ли функция f иметь максимум или минимум в точке $\mathbf{p}_0$, мы должны исследовать производные более высокого порядка.

Рассмотрим пример использования гессиана. Мы должны найти и классифицировать критические точки функции

$$f = 3x^2 + 2y^3 - 6xy.$$

Вычислим частные производные по x и по y и приравняем их к нулю.

$$\frac{\partial f}{\partial x} = f_1'(x, y) = 6x - 6y = 0, \quad \text{т. е.} \quad x = y,$$
$$\frac{\partial f}{\partial y} = f_2'(x, y) = 6y^2 - 6x = 0, \quad \text{т. е.} \quad x = y^2.$$

Следовательно, функция f имеет критические точки: $x = 0$, $y = 0$ и $x = 1$, $y = 1$. Чтобы построить гессиан, вычислим частные производные второго порядка:

$$\frac{\partial^2 f}{\partial x^2} = 6, \quad \frac{\partial^2 f}{\partial x \partial y} = -6, \quad \frac{\partial^2 f}{\partial y^2} = 12y.$$

Следовательно, в точке $x = 0$, $y = 0$ матрица гессиана равна

$$H = \begin{pmatrix} 6 & -6 \\ -6 & 0 \end{pmatrix}.$$

Определитель этого гессиана отрицательный, следовательно, его собственные значения имеют противоположные знаки, и функция $f(x, y)$ имеет в начале координат седловую точку. Чтобы проверить это, заметим, что f имеет положительные значения при $y = 0$ и малых x. А при $x = y$ функция имеет отрицательные значения в окрестности начала координат.

В точке $x = 1$, $y = 1$ матрица гессиана равна

$$H = \begin{pmatrix} 6 & -6 \\ -6 & 12 \end{pmatrix}.$$

Определитель этого гессиана положителен, поэтому собственные значения имеют одинаковый знак. Так как след определителя положителен, то собственные значения гессиана положительны. Следовательно, в точке $x = 1$, $y = 1$ функция имеет локальный минимум.

На аффинной плоскости дифференциал второго порядка $d^2 f$ обладает одним свойством, которое не зависит от выбора системы координат, а именно, число положительных, отрицательных и нулевых собственных значений гессиана. На евклидовой плоскости функция f обладает еще одним свойством, не зависящим от выбора системы координат. Чтобы получить его, сравним значение функции, усредненное по малой окружности вокруг точки $\mathbf{p}_0$, со значением функции в точке $\mathbf{p}_0$ (рис. 6.2).

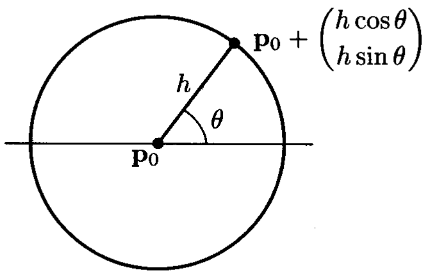


Рис. 6.2

Из формулы (6.12) следует, что

$$f(\mathbf{p}_0 + \mathbf{v}) = f(\mathbf{p}_0) + df[\mathbf{v}] + \frac{1}{2} d^2 f(\mathbf{v}) + \text{остаток}.$$

Поскольку $df[-\mathbf{v}] = -df[\mathbf{v}]$, среднее значение $df[\mathbf{v}]$ для любой окружности с центром в точке $\mathbf{p}_0$ равно нулю. Чтобы усреднить $\frac{1}{2} d^2 f(\mathbf{v})$, положим $\mathbf{v} = \begin{pmatrix} h\cos\theta \\ h\sin\theta \end{pmatrix}$. Воспользуемся формулой (6.13):

$$\begin{aligned} \frac{1}{2} d^2 f(\mathbf{v}) &= \frac{1}{2}(h\cos\theta, h\sin\theta) \begin{pmatrix} \partial^2 f/\partial x^2 & \partial^2 f/\partial x \partial y \\ \partial^2 f/\partial x \partial y & \partial^2 f/\partial y^2 \end{pmatrix} \begin{pmatrix} h\cos\theta \\ h\sin\theta \end{pmatrix} \\ &= \frac{1}{2} h^2 \left[\frac{\partial^2 f}{\partial x^2} \cos^2\theta + 2 \frac{\partial^2 f}{\partial x \partial y} \cos\theta \sin\theta + \frac{\partial^2 f}{\partial y^2} \sin^2\theta \right]. \end{aligned}$$

Средние значения $\cos^2\theta$ и $\sin^2\theta$ на отрезке $[0, 2\pi]$ равны $\frac{1}{2}$, среднее значение $\cos\theta\sin\theta$ равно нулю. Поэтому

$$\langle\frac{1}{2}d^2 f(\mathbf{v})\rangle = \frac{1}{4}h^2\left[\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right].$$

Следовательно,

$$\langle f(\mathbf{p}_0 + \mathbf{v})\rangle = f(\mathbf{p}_0) + \frac{1}{4}h^2\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right) + \text{остаток}.$$

Величина $\partial^2 f/\partial x^2 + \partial^2 f/\partial y^2$, определяющая «в среднем» поведение функции f при движении от точки $\mathbf{p}_0$, называется *лапласианом* f. Для любой системы координат, полученной вращением системы x, y, лапласиан имеет одинаковое значение. Поэтому уравнение

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0,$$

называемое *уравнением Лапласа*, описывает свойство функции на евклидовой плоскости, не зависящее от выбора системы координат. Не удивительно, что такое уравнение часто описывает функции, имеющие физический смысл: электрический потенциал, температура и т. д.

Прежде чем закончить исследование вопроса о максимуме и минимуме функции на плоскости, рассмотрим проблему *условного* экстремума: найти максимум и минимум функции $f(\mathbf{p})$ на кривой, заданной уравнением уравнением $g(\mathbf{p}) = \text{const}$. [6]

Необходимое условие того, что функция достигает условного экстремума в точке $\mathbf{p}_0$, имеет вид: $df_{\mathbf{p}_0}(\mathbf{v}) = 0$ для любого вектора $\mathbf{v}$, касательного к кривой $g(\mathbf{p}) = \text{const}$ (рис. 6.3). Но такой вектор удовлетворяет уравнению $dg(\mathbf{v}) = 0$. Следовательно, в точке $\mathbf{p}_0$, где достигается максимум или минимум исследуемой функции, значение $df_{\mathbf{p}_0}$ должно быть *пропорционально* $dg_{\mathbf{p}_0}$, т. е. $df = \lambda\, dg$. Таким образом, проблема условного экстремума привела нас к методу, известному как метод *множителей Лагранжа*.

[6] В дальнейшем используется, что f, g дифференцируемы, а также условие, что $dg_{\mathbf{p}_0}$ не равен тождественно нулю. — *Прим. ред.*

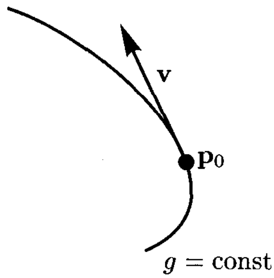


Рис. 6.3

Рассмотрим функцию $f-\lambda g$, где λ — неопределенный скалярный коэффициент, называемый *множителем Лагранжа*. Если в точке $\mathbf{p}_0$ достигается максимум или минимум, то в этой точке при некотором λ должно выполняться равенство $d(f-\lambda g)=0$. Кроме уравнения $g=\text{const}$, это дает еще два, из которых определяются неизвестные значения x, y и λ. Чтобы выяснить, какой экстремум (максимум или минимум) мы получили (или никакого), сначала найдем вектор $\mathbf{v}$, для которого $dg(\mathbf{v})=0$. Этот вектор будет касательным к кривым $f=\text{const}$ и $g=\text{const}$ в критической точке $\mathbf{p}_0$. Затем рассмотрим функцию $h=f-\lambda g$, у которой, по построению, точка $\mathbf{p}_0$ будет критической на плоскости. Найдем наилучшее квадратичное приближение к h в окрестности этой критической точки и вычислим его для нашего вектора $\mathbf{v}$. Если величина $\frac{1}{2}d^2h(\mathbf{v})$ положительна, то $h(\mathbf{p})>h(\mathbf{p}_0)$ во всех точках кривой вблизи $\mathbf{p}_0$. Действительно, зададим кривую $g=0$ параметрически: $p=p(t)$, т.е. мы возьмем функцию $p:\mathbb{R}\to\mathbb{R}^2$, для которой

$$g(p(t))\equiv 0, \quad p(0)=\mathbf{p}_0, \quad p'(0)=\mathbf{v}.$$

(Это всегда можно сделать, как следует из теоремы о неявной функции, которую мы докажем в следующем параграфе.) Тогда

$$h\circ p=f\circ p, \quad \text{так как} \quad g\circ p\equiv 0.$$

Далее,

$$(h\circ p)'(0)=dh_{\mathbf{p}_0}(\mathbf{v})=0,$$
$$(f\circ p)''=(h\circ p)''(0)=d^2h_{\mathbf{p}_0}(\mathbf{v}).$$

Если $(h \circ p)''(0) > 0$, то f имеет минимум в точке $\mathbf{p}_0$ на линии $g = 0$. Аналогично, f имеет максимум на линии $g = 0$, если $d^2h(\mathbf{v}) < 0$.

Например, мы хотим найти экстремум квадратичной формы

$$Q(x, y) = 8x^2 - 12xy + 17y^2$$

на окружности $G(x, y) = x^2 + y^2 = 1$. Положив дифференциал $Q - \lambda G$ равным нулю, получаем

$$\begin{aligned} 16x - 12y - 2\lambda x &= 0, \\ -12x + 34y - 2\lambda y &= 0. \end{aligned}$$

Исключив из этих уравнений λ, получаем

$$16 - 12(y/x) = -12(x/y) + 34$$

или

$$(x/y) - (y/x) = 3/2.$$

Отсюда

$$x^2 - \frac{3}{2}xy - y^2 = 0,$$

и, следовательно,

$$(x - 2y)\left(x + \frac{1}{2}y\right) = 0.$$

Таким образом, квадратичная форма Q может принимать экстремальные значения в точках пересечения прямых $x = 2y$ и $x = -\dfrac{1}{2}y$ с окружностью $x^2 + y^2 = 1$ (рис. 6.4), т. е. в точках

$$\mathbf{p}_1 = \pm \begin{pmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{pmatrix} \quad \text{и} \quad \mathbf{p}_2 = \pm \begin{pmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{pmatrix}.$$

Исследуем теперь, имеются ли экстремумы в этих точках, и какие именно. Вычислим дифференциалы

$$\begin{aligned} dQ &= (16x - 12y)\,dx + (-12x + 34y)\,dy, \\ dG &= 2x\,dx + 2y\,dy. \end{aligned}$$

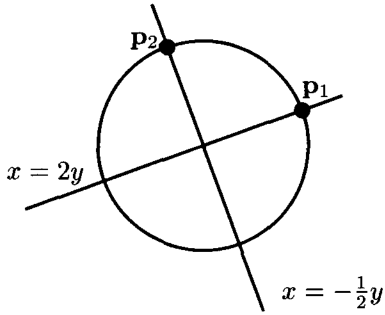


Рис. 6.4

В точке $\mathbf{p}_1$, где $x = 2y$, получаем

$$dQ = 10x\,dx + 5x\,dy, \qquad dG = 2x\,dx + x\,dy.$$

Как мы и ожидали, dQ пропорционален dG с множителем $\lambda = 5$. Вектор $\mathbf{v}$, для которого $dQ(\mathbf{v}) = dG(\mathbf{v}) = 0$, равен $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$. В этом случае гессиан квадратичной формы $Q - 5G$ равен

$$H = \begin{pmatrix} 3 & -6 \\ -6 & 12 \end{pmatrix}.$$

Вычисляем $(1, -2)\begin{pmatrix} 3 & -6 \\ -6 & 12 \end{pmatrix}\begin{pmatrix} 1 \\ -2 \end{pmatrix} = 75$ и приходим к выводу, что Q имеет *минимум* на окружности в точке $\mathbf{p}_1$.

В другой критической точке $\mathbf{p}_2$, где $y = -2x$, получаем

$$dQ = 40x\,dx - 80x\,dy, \quad dG = 2x\,dx - 4x\,dy,$$

следовательно, $\lambda = 20$. Вектор, для которого $dQ(\mathbf{v}) = dG(\mathbf{v}) = 0$, равен $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Вычислив гессиан функции $Q - 20G$ для этого вектора, получаем

$$(2, 1)\begin{pmatrix} -12 & -6 \\ -6 & -3 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = (2, 1)\begin{pmatrix} -30 \\ -15 \end{pmatrix} = -75,$$

значит, Q имеет локальный максимум на окружности $x^2 + y^2 = 1$ в точке $\mathbf{p}_2$.

6.4. Теорема об обратной функции

Пусть U и V — векторные пространства одинаковой размерности и пусть $f : U \to V$ — дифференцируемое отображение с условием $f(\mathbf{p}_0) = \mathbf{q}_0$. Мы хотели бы знать, при каких условиях существует такое отображение $g : V \to U$, что $g \circ f = (id)$. Прежде чем дать формулировку соответствующей теоремы, исследуем сначала ряд необходимых ограничений этой проблемы.

Если мы ожидаем, что отображение g тоже должно быть дифференцируемым, то согласно цепному правилу

$$dg_{f(\mathbf{p})} \circ df_{\mathbf{p}} = (id).$$

Тогда *линейное* отображение $df_{\mathbf{p}}$ должно быть обратимым. Если это так, то справедлива формула

$$dg_{f(\mathbf{p})} = [df_{\mathbf{p}}]^{-1}.$$

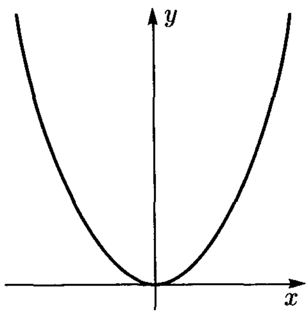


Рис. 6.5

Обсудим случай, когда df не является обратимым отображением: $U = V = \mathbb{R}^1$ и $f(x) = x^2$ (рис. 6.5). В начале координат $df_{\mathbf{0}} = 0$, и возникает проблема с $g(y) = \sqrt{y}$ вблизи $y = 0$. По существу, возникают три разных проблемы. Прежде всего, $\sqrt{y}$ не определен (в области вещественных чисел) для $y < 0$. Более точно, не существует точки $y < 0$, являющейся образом f. Во-вторых, корень квадратный из y определен неоднозначно: для данного $y > 0$

существует два значения x, удовлетворяющих уравнению $x^2 = y$. И последнее: производная $\sqrt{y}$ неограниченно растет при $y \to 0$. Чтобы обойти вторую из перечисленных трудностей, проделаем следующую процедуру.

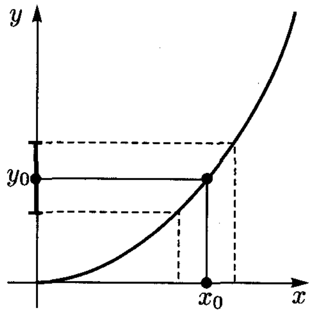


Рис. 6.6

Предположим, что мы взяли некоторое значение $x_0 \neq 0$, для которого $x_0^2 = y_0$. Для определенности, пусть $x_0 > 0$. Тогда в достаточно малой окрестности точки y_0 (настолько малой, чтобы не включать $y = 0$) существует единственная обратная функция: квадратный корень, удовлетворяющая конкретному требованию: значения функции должны быть достаточно близки к x_0. (В этом случае слова «достаточно близки» значат, что функция не может быть отрицательной. Если мы выбрали положительное значение корня, то функция определена однозначно (рис.6.6).)

Мы можем не только утверждать существование квадратного корня, но и дать алгоритм его вычисления с любой наперед заданной точностью. Здесь приводится один из возможных алгоритмов — метод Ньютона, который формулируется в общем случае.

Допустим, что у нас есть отображение $f : U \to V$ с условием $f(\mathbf{p}_0) = \mathbf{q}_0$. Пусть задано некоторое значение $\mathbf{q}$ вблизи $\mathbf{q}_0$, и мы хотим найти такое $\mathbf{p}$ вблизи $\mathbf{p}_0$, что $f(\mathbf{p}) = \mathbf{q}$. Найти $\mathbf{p}$ — это то же самое, что найти $\mathbf{p} - \mathbf{p}_0$. Итак, мы хотели бы иметь равенство

$$f(\mathbf{p}_0 + \mathbf{p} - \mathbf{p}_0) = \mathbf{q}.$$

Однако,

$$\begin{aligned} f(\mathbf{p}_0 + \mathbf{p} - \mathbf{p}_0) &= f(\mathbf{p}_0) + df_{\mathbf{p}_0}(\mathbf{p} - \mathbf{p}_0) + o(\mathbf{p} - \mathbf{p}_0) \\ &= \mathbf{q}_0 + df_{\mathbf{p}_0}(\mathbf{p} - \mathbf{p}_0) + o(\mathbf{p} - \mathbf{p}_0). \end{aligned}$$

Если опустить член $o(\mathbf{p} - \mathbf{p}_0)$, то получится приближенное уравнение

$$\mathbf{q} - \mathbf{q}_0 = df_{\mathbf{p}_0}(\mathbf{p} - \mathbf{p}_0),$$

или, так как $df_{\mathbf{p}_0}$ имеет обратное отображение,

$$\mathbf{p} = \mathbf{p}_0 + df_{\mathbf{p}_0}^{-1}(\mathbf{q} - \mathbf{q}_0).$$

Это значит, что мы получили

$$\mathbf{p}_1 = \mathbf{p}_0 + df_{\mathbf{p}_0}^{-1}(\mathbf{q} - \mathbf{q}_0)$$

в качестве приближенного решения. Теперь определим

$$\mathbf{q}_1 = f(\mathbf{p}_1)$$

и начнем все сначала. На следующем шаге получим

$$\mathbf{p}_2 = \mathbf{p}_1 + df_{\mathbf{p}_1}^{-1}(\mathbf{q} - \mathbf{q}_1),$$

и так далее[7]. В этом и состоит *метод Ньютона.*

Вернемся к рассмотренному выше примеру. Пусть $U = V = \mathbb{R}^1$ и $f(p) = p^2$. Тогда df_p есть умножение на $2p$ и, следовательно, $df_p^{-1}(w) = w/2p$. В этом случае

$$p_1 = p_0 + \frac{1}{2p_0}(q - q_0).$$

Например, если взять $p_0 = 3$, то $q_0 = 9$ и $q = 10$. Тогда

$$p_1 = 3 + \frac{1}{6}(10 - 9) = 3.1666\ldots, \quad q_1 = p_1^2 = 10.02777\ldots$$

[7]При выполнении некоторых условий последовательность $\mathbf{p}_1, \mathbf{p}_2, \ldots$, вычисляемая по формуле $\mathbf{p}_{i+1} = \mathbf{p}_i + df_{\mathbf{p}_i}^{-1}(\mathbf{q} - \mathbf{q}_i)$, имеет предел $\mathbf{p}$, который является решением уравнения $f(\mathbf{p}) = \mathbf{q}$. Авторы говорят об этом ниже, в пункте «Сходимость метода Ньютона». — *Прим. ред.*

(Заметим, что p_1 дает значительно лучшее приближение для $\sqrt{10}$, чем $p_0 = 3$.)

Следующее приближение:

$$\begin{aligned} p_2 &= 3.1\bar{6} + \frac{1}{2 \times 3.1\bar{6}}(10 - 10.02\bar{7}) \\ &= 3.1622816\ldots \end{aligned}$$

При этом

$$q_2 = p_2^2 = 10.000024,$$

а число p_2 дает ответ с точностью до четырех десятичных знаков ($\sqrt{10} = 3.1622776\ldots$).

Рассмотрим еще пример. Пусть $U = V = \mathbb{R}^2$ и отображение f дается формулой

$$f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x^3 - y^3 \\ 2xy \end{pmatrix}.$$

Тогда $df_{\begin{pmatrix} x \\ y \end{pmatrix}}$ есть линейное преобразование, матрица которого равна

$$\begin{pmatrix} 3x^2 & -3y^2 \\ 2y & 2x \end{pmatrix}.$$

Пусть $\mathbf{p}_0 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Тогда

$$\mathbf{q}_0 = \begin{pmatrix} 8 - 1 \\ 2 \cdot 2 \cdot 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 4 \end{pmatrix},$$

$$df_{\mathbf{p}_0} = \begin{pmatrix} 12 & -3 \\ 2 & 4 \end{pmatrix}, \quad (df_{\mathbf{p}_0})^{-1} = \frac{1}{54}\begin{pmatrix} 4 & 3 \\ -2 & 12 \end{pmatrix}.$$

Возьмем $\mathbf{q} = \begin{pmatrix} 7.5 \\ 3.8 \end{pmatrix}$ и будем решать уравнение $f(\mathbf{p}) = \mathbf{q}$ по методу Ньютона. Тогда

$$\begin{aligned} \mathbf{p}_1 &= \mathbf{p}_0 + (df_{\mathbf{p}_0})^{-1}(\mathbf{q} - \mathbf{q}_0) \\ &= \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \frac{1}{54}\begin{pmatrix} 4 & 3 \\ -2 & 12 \end{pmatrix}\begin{pmatrix} 0.5 \\ -0.2 \end{pmatrix} = \begin{pmatrix} 2.026 \\ 0.937 \end{pmatrix}. \end{aligned}$$

В результате получаем $\mathbf{q}_1 = \begin{pmatrix} 7.493 \\ 3.796 \end{pmatrix}$, что уже достаточно близко к точному решению: $\mathbf{p} = \begin{pmatrix} 2.026643\ldots \\ 0.937510\ldots \end{pmatrix}$. Заметим, что каждый последующий шаг этого алгоритма означает вычисление значений преобразования $(df_{\mathbf{p}_i})^{-1}$ в различных точках $\mathbf{p}_i$.

Ниже мы сформулируем математическую теорему, согласно которой при некоторых предположениях относительно f метод Ньютона дает последовательность точек $\mathbf{p}_i$, которая сходится к решению $\mathbf{p}$, когда $f(\mathbf{p}) = \mathbf{q}$ при условии, что $\mathbf{q}$ достаточно близко к $\mathbf{q}_0$.

Существует другой алгоритм вычисления квадратного корня, называемый *методом Пикара.* Однако сходится он гораздо медленнее метода Ньютона. Согласно методу Пикара, полагаем

$$L = (df_{\mathbf{p}_0})^{-1}$$

и тогда

$$\begin{aligned} \mathbf{p}_1 &= \mathbf{p}_0 + L(\mathbf{q} - \mathbf{q}_0), \quad \mathbf{q}_1 = f(\mathbf{p}_1), \\ \mathbf{p}_2 &= \mathbf{p}_1 + L(\mathbf{q} - \mathbf{q}_1), \quad \mathbf{q}_2 = f(\mathbf{p}_2), \quad \text{и так далее.} \end{aligned}$$

Например, для $f : \mathbb{R} \to \mathbb{R}$ с $f(p) = p^2$, $q = 10$ и $p_0 = 3$ мы имеем

$$p_1 = 3.1\bar{6}, \quad q_1 = 10.02\bar{7},$$

как и в методе Ньютона. Однако

$$\begin{aligned} p_2 &= 3.1\bar{6} + \frac{1}{6}(-0.02\bar{7}) = 3.162036, \\ q_2 &= 9.9984817 \quad \text{и так далее.} \end{aligned}$$

Преимущество метода Пикара в том, что L вычисляется один раз.

Сравнительно легко доказать сходимость методов Пикара и Ньютона при небольших ограничениях на f. Получение хорошей оценки скорости сходимости метода Ньютона потребует, как мы увидим, более сильных ограничениях на f.

Доказательства сходимости. Здесь мы сформулируем предположения относительно функции f, которые нам необходимы для доказательства сходимости обоих методов. Напоминаем, что мы называем функцию f дифференцируемой в точке $\mathbf{p}$, если

$$f(\mathbf{p}^1) = f(\mathbf{p}) + df_{\mathbf{p}}(\mathbf{p}^1 - \mathbf{p}) + o(\mathbf{p}^1 - \mathbf{p}).$$

Это значит, что для любого данного ε можно найти такое δ, что

$$\|f(\mathbf{p}^1) - f(\mathbf{p}) - df_{\mathbf{p}}(\mathbf{p}^1 - \mathbf{p})\| \leq \varepsilon\|\mathbf{p}^1 - \mathbf{p}\| \tag{6.14}$$

при условии, что $\|\mathbf{p}^1 - \mathbf{p}\| \leq \delta$. Величина δ, стоящая в этом неравенстве, может зависеть от точки $\mathbf{p}$. Предположим, что f *равномерно дифференцируема* в том смысле, что для любого ε можно найти такое δ, что (6.14) выполняется для всех точек $\mathbf{p}$ и $\mathbf{p}^1$, принадлежащих небольшому шарику с центром в $\mathbf{p}_0$. Итак, мы предполагаем, что существует такое $a > 0$, что для данного ε можно найти такое δ, чтобы неравенство (6.14) выполнялось при условии:

$$\|\mathbf{p}^1 - \mathbf{p}\| \leq \delta, \qquad \|\mathbf{p} - \mathbf{p}_0\| \leq a, \qquad \|\mathbf{p}^1 - \mathbf{p}_0\| \leq a.$$

Предположим также, что $df_{\mathbf{p}}^{-1}$ равномерно ограничена, т. е. что существует некоторая постоянная K, для которой

$$\|df_{\mathbf{p}}^{-1}\| < K \quad \text{для всех} \quad \|\mathbf{p} - \mathbf{p}_0\| \leq a. \tag{6.15}$$

Кроме того, будем считать, что и сама функция $df_{\mathbf{p}}$ равномерно ограничена, т. е. что

$$\|df_{\mathbf{p}}\| < M \quad \text{для всех} \quad \|\mathbf{p} - \mathbf{p}_0\| \leq a.$$

Наконец, для метода Пикара предположим, что $df_{\mathbf{p}}$ непрерывна в том смысле, что для любого $\varepsilon > 0$ существует такое δ, что

$$\|df_{\mathbf{p}} - df_{\mathbf{p}^1}\| \leq \frac{1}{2}\varepsilon,$$

$$\text{если} \quad \|\mathbf{p} - \mathbf{p}^1\| \leq \delta, \quad \|\mathbf{p} - \mathbf{p}_0\| \leq a, \quad \|\mathbf{p}^1 - \mathbf{p}_0\| \leq a, \tag{6.16}$$

и нам потребуется только существование $(df_{\mathbf{p}_0})^{-1}$.

Сходимость метода Ньютона. Рассмотрим метод Ньютона. Переход от точки $\mathbf{p}_i$ к точке $\mathbf{p}_{i+1}$ определяется формулой

$$\mathbf{p}_{i+1} = \mathbf{p}_i + (df_{\mathbf{p}_i})^{-1}(\mathbf{q} - f(\mathbf{p}_i)). \tag{6.17}$$

Дифференцируемость $f(\mathbf{p})$ в точке $\mathbf{p}_{i-1}$ означает, что

$$\begin{aligned} f(\mathbf{p}_i) &= f(\mathbf{p}_{i-1} + \mathbf{p}_i - \mathbf{p}_{i-1}) \\ &= f(\mathbf{p}_{i-1}) + df_{\mathbf{p}_{i-1}}(\mathbf{p}_i - \mathbf{p}_{i-1}) + o(\mathbf{p}_i - \mathbf{p}_{i-1}) \end{aligned} \tag{6.18}$$

Равенство (6.17) можно переписать в виде

$$df_{\mathbf{p}_i}(\mathbf{p}_{i+1} - \mathbf{p}_i) = \mathbf{q} - f(\mathbf{p}_i),$$

откуда заменой i на $(i-1)$ мгновенно получается, что сумма первых двух слагаемых в правой части (6.18) равна $\mathbf{q}$, поэтому из (6.18) следует, что

$$f(\mathbf{p}_i) - \mathbf{q} = o(\mathbf{p}_i - \mathbf{p}_{i-1}).$$

Это значит, что для любого $\varepsilon > 0$ можно подобрать $\delta > 0$ так, чтобы при $\|\mathbf{p}_i - \mathbf{p}_{i-1}\| < \delta$ выполнялось неравенство

$$\|f(\mathbf{p}_i) - \mathbf{q}\| < \varepsilon\|\mathbf{p}_i - \mathbf{p}_{i-1}\|.$$

С учетом равенства (6.17) и оценки (6.15) из только что написанной формулы можно получить следующую оценку:

$$\|\mathbf{p}_{i+1} - \mathbf{p}_i\| \leq K\varepsilon\|\mathbf{p}_i - \mathbf{p}_{i-1}\|. \tag{6.19}$$

В дальнейшем будем считать, что число ε взято таким, что $K\varepsilon < \dfrac{1}{2}$.

Перейдем к оценке скорости сходимости метода Ньютона. Прежде всего,

$$\mathbf{p}_1 = \mathbf{p}_0 + df^{-1}_{\mathbf{p}_0}(\mathbf{q} - \mathbf{q}_0), \quad \mathbf{q}_0 = f(\mathbf{p}_0).$$

Поэтому, если $\|\mathbf{q} - \mathbf{q}_0\| < \delta/K$, то $\|\mathbf{p} - \mathbf{p}_0\| \leq \delta$. Если $\delta < a/2$, точка $\mathbf{p}_1$ будет удовлетворять неравенству

$$\|\mathbf{p}_1 - \mathbf{p}_0\| < \frac{1}{2}a.$$

Поскольку $\mathbf{p}_1$ лежит внутри области определения f, то можно воспользоваться алгоритмом (6.17) и получить $\mathbf{p}_2$. Из неравенства (6.19) следует, что

$$\|\mathbf{p}_2 - \mathbf{p}_1\| \leq \frac{1}{2}\delta \leq \frac{1}{2}\left(\frac{1}{2}a\right),$$

следовательно,

$$\|\mathbf{p}_2 - \mathbf{p}_0\| \leq \|\mathbf{p}_2 - \mathbf{p}_1\| + \|\mathbf{p}_1 - \mathbf{p}_0\| \leq \left(\frac{1}{2} + 1\right)\frac{1}{2}a \leq a.$$

Итак, точка $\mathbf{p}_2$ опять лежит внутри шарика радиуса a и опять можно воспользоваться алгоритмом (6.17). Получаем

$$\|\mathbf{p}_3 - \mathbf{p}_2\| \leq \frac{1}{2}\|\mathbf{p}_2 - \mathbf{p}_1\| \leq \frac{1}{4}\delta,$$

и, следовательно,

$$\|\mathbf{p}_3 - \mathbf{p}_0\| \leq \left(\frac{1}{4} + \frac{1}{2} + 1\right)\frac{1}{2}a \leq a,$$

и так далее. По индукции можно доказать, что при всех i

$$\|\mathbf{p}_i - \mathbf{p}_{i-1}\| \leq \delta/2^{i-1}$$

и

$$\|\mathbf{p}_i - \mathbf{p}_0\| \leq \left(\frac{1}{2^i} + \frac{1}{2^{i-1}} + \cdots + 1\right)\frac{1}{2}a \leq 2\cdot\frac{1}{2}a \leq a.$$

Последовательность точек $\mathbf{p}_i$ сходится[8] к некоторой точке $\mathbf{p}$, т. к.

$$\|\mathbf{p}_{i+k} - \mathbf{p}_i\| \leq \frac{1}{2^{i-1}}\left(\frac{1}{2^k} + \frac{1}{2^{k-1}} + \cdots + \frac{1}{2}\right)\delta \leq \frac{1}{2^{i-1}}\delta \to 0$$

при $i \to \infty$.

Наконец,

$$\|\mathbf{q} - f(\mathbf{p}_i)\| = \|df_{\mathbf{p}_i}(\mathbf{p}_{i+1} - \mathbf{p}_i)\| \leq M\|\mathbf{p}_{i+1} - \mathbf{p}_i\| \to 0,$$

т. е. $\lim_{i\to\infty} f(\mathbf{p}_i) = \mathbf{q}$. Так как функция f непрерывна, то

$$f(\mathbf{p}) = \mathbf{q}.$$

Единственность решения. Рассмотрим проблему единственности решения уравнения $f(\mathbf{p}) = \mathbf{q}$. Заметим, что величина K

[8] Здесь авторы используют классическое необходимое и достаточное условие сходимости, известное как «критерий Коши». — *Прим. ред.*

определяется функцией f. При желании мы можем выбрать еще меньшее значение a, не изменяя при этом K. Сделать это можно за счет выбора δ и, следовательно, меньшего значения δ/K. В частности, предположим, что a выбрано настолько малым, что (6.14) выполняется для любой пары точек $\mathbf{p}$ и $\mathbf{p}^1$, и при этом $\varepsilon K < 1$. Тогда для любой пары точек имеем

$$\|\mathbf{p} - \mathbf{p}^1\| = \|(df_{\mathbf{p}})^{-1}(df_{\mathbf{p}}(\mathbf{p} - \mathbf{p}^1))\| \le K\|df_{\mathbf{p}}(\mathbf{p} - \mathbf{p}^1)\|.$$

Если $f(\mathbf{p}) = f(\mathbf{p}^1)$, то неравенство (6.14) значит, что

$$\|df_{\mathbf{p}}(\mathbf{p}^1 - \mathbf{p}) \le \varepsilon\|\mathbf{p}^1 - \mathbf{p}\|,$$

и тогда, комбинируя эти два неравенства, получаем

$$\|\mathbf{p} - \mathbf{p}^1\| \le \varepsilon K\|\mathbf{p} - \mathbf{p}^1\|, \quad \varepsilon K < 1.$$

Это неравенство справедливо только в случае $\|\mathbf{p} - \mathbf{p}^1\| = 0$, т. е. $\mathbf{p} = \mathbf{p}^1$. Таким образом, для $\|\mathbf{p} - \mathbf{p}_0\| \le a$ может быть только одно уравнения решение $f(\mathbf{p}) = \mathbf{q}$.

Сходимость метода Пикара. Рассмотрим метод Пикара. Пусть $L = (df_{\mathbf{p}_0})^{-1}$. Тогда

$$\begin{aligned}\mathbf{p}_{i+1} &= \mathbf{p}_i + L(\mathbf{q} - f(\mathbf{p}_i)) \\ &= \mathbf{p}_i + L(\mathbf{q} - f(\mathbf{p}_{i-1}) + df_{\mathbf{p}_{i-1}}(\mathbf{p}_i - \mathbf{p}_{i-1}) + o(\mathbf{p}_i - \mathbf{p}_{i-1})),\end{aligned}$$

как и раньше. Теперь

$$\mathbf{q} - f(\mathbf{p}_{i-1}) + df_{\mathbf{p}_0}(\mathbf{p}_i - \mathbf{p}_{i-1}) = \mathbf{0},$$

следовательно,

$$\begin{aligned}\|\mathbf{q} - f(\mathbf{p}_{i-1}) + df_{\mathbf{p}_{i-1}}(\mathbf{p}_i - \mathbf{p}_{i-1})\| &= \|(df_{\mathbf{p}_{i-1}} - df_{\mathbf{p}_0})(\mathbf{p}_i - \mathbf{p}_{i-1})\| \\ &\le \frac{\varepsilon}{2}\|\mathbf{p}_i - \mathbf{p}_{i-1}\|\end{aligned}$$

при условии, что выбранное значение a достаточно мало. Кроме того, мы можем выбрать δ настолько малым, чтобы в неравенстве (6.14) можно было ε заменить на $\frac{1}{2}\varepsilon$. Тогда

$$\begin{aligned}\|\mathbf{p}_{i+1} - \mathbf{p}_i\| &\le k\left(\frac{1}{2}\varepsilon\|\mathbf{p}_i - \mathbf{p}_{i-1}\| + \frac{1}{2}\varepsilon\|\mathbf{p}_i - \mathbf{p}_{i-1}\|\right) \\ &\le K\varepsilon\|\mathbf{p}_i - \mathbf{p}_{i-1}\|,\end{aligned}$$

так что неравенство (6.19), где сделана замена $K = \|L\|$, выполняется, как и раньше. На самом деле можно воспользоваться теоремой о среднем значении, чтобы изменить обоснование метода Пикара, и таким образом избежать необязательного предположения о *равномерной дифференцируемости*. Действительно, рассмотрим отображение

$$h(\mathbf{p}) = \mathbf{p} + L(\mathbf{q} - f(\mathbf{p})),$$

где $L = (df_{\mathbf{p}_0})^{-1}$. Непрерывность дифференциала $df_{\mathbf{p}}$ позволяет выбрать значение a настолько малым, чтобы выполнялось неравенство

$$\|dh_{\mathbf{p}}\| = \|I - Ldf_{\mathbf{p}}\| \leq \frac{1}{2}.$$

Тогда, воспользовавшись теоремой о среднем значении, получаем

$$\|\mathbf{p}_{i+1} - \mathbf{p}_i\| = \|h(\mathbf{p}_i) - h(\mathbf{p}_{i-1})\| \leq \frac{1}{2}\|\mathbf{p}_i - \mathbf{p}_{i-1}\|,$$

и дальше можно рассуждать, как и раньше.

Теперь обсудим, почему на самом деле метод Ньютона сходится гораздо быстрее, чем это видно из доказанных выше оценок. Предположим, что f имеет две непрерывные производные. Тогда из формулы Тейлора следует, что

$$|f(\mathbf{p}^1) - f(\mathbf{p}) - df_{\mathbf{p}}(\mathbf{p}^1 - \mathbf{p})| \leq c\|\mathbf{p}^1 - \mathbf{p}\|^2,$$

где постоянная c определяется максимумом $|d^2 f|$. Это неравенство значительно сильнее, чем (6.14). Возвращаясь назад к доказательству (6.17) и использовав это неравенство, получаем

$$\|\mathbf{p}_{i+1} - \mathbf{p}_1\| \leq K\|\mathbf{p}_i - \mathbf{p}_{i-1}\|^2.$$

Если, например, мы начали со столь малого значения $\|\mathbf{p}_1 - \mathbf{p}_0\|$, что $K\|\mathbf{p}_1 - \mathbf{p}_0\|^{1/2} \leq 1$ (и $\|\mathbf{p}_1 - \mathbf{p}_0\| < 1$), то из последнего неравенства следует

$$\|\mathbf{p}_{i+1} - \mathbf{p}_i\| \leq K\|\mathbf{p}_i - \mathbf{p}_{i-1}\|^{1/2} \cdot \|\mathbf{p}_i - \mathbf{p}_{i-1}\|^{3/2}.$$

Тогда

$$\|\mathbf{p}_{i+1} - \mathbf{p}_i\| \leq \|\mathbf{p}_1 - \mathbf{p}_0\|^{(3/2)^n},$$

т. е. мы получаем экспоненциальную скорость убывания вместо геометрической прогрессии $\|\mathbf{p}_{i+1} - \mathbf{p}_i\| \leq (\varepsilon K)^n \|\mathbf{p}_1 - \mathbf{p}_0\|$, получающейся в методе Пикара.

Итак, резюмируем, что мы выяснили. Мы показали, что при соответствующих условиях вокруг точки $\mathbf{q}_0 = f(\mathbf{p}_0)$ существует шарик B, а вокруг точки $\mathbf{p}_0$ шарик C, такие, что для каждой точки $\mathbf{q} \in B$ существует единственная точка $\mathbf{p} \in C$, для которой $f(\mathbf{p}) = \mathbf{q}$ (рис. 6.7). Другими словами, мы определили такое отображение $g : B \to C$, что $f \circ g = (id)$ и $g \circ f = (id)$.[9]

Дифференцируемость решения. Докажем, что функция g дифференцируема. Заметим, что из единственности g вытекает ее непрерывность. Итак, предположим, что $g(\mathbf{q}) = \mathbf{p}$. Нарисуем шарик с центром в точке $\mathbf{p}$. А теперь воспользуемся результатами, уже полученными для точек $\mathbf{p}$ и $\mathbf{q}$. Это значит, что существуют шарик с центром в точке $\mathbf{q}$, и функция, обратная к f, отображающая его в шарик с центром в точке $\mathbf{p}$.

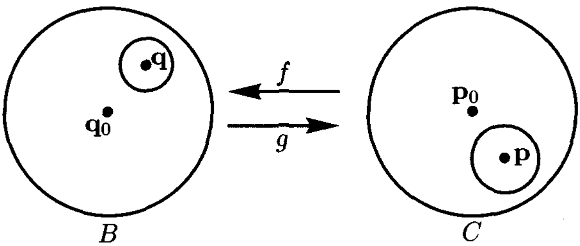


Рис. 6.7

Однако, в силу единственности, эта обратная функция должна совпадать с g. Следовательно, g отображает небольшой шарик с центром $\mathbf{q}$ в небольшой шарик с центром $\mathbf{p}$, т. е. она непрерывна. Тогда

$$\mathbf{v} = f(g(\mathbf{q}+\mathbf{v})) - f(g(\mathbf{q})) = df_{\mathbf{p}}(g(\mathbf{q}+\mathbf{v}) - g(\mathbf{q})) + o(g(\mathbf{q}+\mathbf{v}) - g(\mathbf{q})).$$

Применим отображение $(df_{\mathbf{p}})^{-1}$ к обеим частям этого равенства.

[9]Последнее предложение, означающее биективность отображения g, требует уточнения: для биективности g следует его рассматривать как отображение $g : B \to C_1$, где $C_1 = g(B)$ — образ шарика B при отображении g, который содержится в C, но не обязательно заполняет все C. — *Прим. ред.*

В результате получим

$$(df_{\mathbf{p}})^{-1}(\mathbf{v}) = g(\mathbf{q}+\mathbf{v}) - g(\mathbf{q}) + o(g(\mathbf{q}+\mathbf{v}) - g(\mathbf{q})).$$

В силу непрерывности функции g мы можем выбрать $\mathbf{v}$ настолько малым, что $\|o(g(\mathbf{q}+\mathbf{v}) - g(\mathbf{q}))\|$ будет меньше $\dfrac{1}{2}\|g(\mathbf{q}+\mathbf{v}) - g(\mathbf{q})\|$. Из написанного выше равенства следует, что

$$\|(df_{\mathbf{p}})^{-1}(\mathbf{v})\| + \|o(g(\mathbf{q}+\mathbf{v}) - g(\mathbf{q}))\| \geq \|g(\mathbf{q}+\mathbf{v}) - g(\mathbf{q})\|,$$

откуда

$$\|g(\mathbf{q}+\mathbf{v}) - g(\mathbf{q})\| \leq 2\|(df_{\mathbf{p}})^{-1}(\mathbf{v})\|,$$

т. е. $g(\mathbf{q}+\mathbf{v}) - g(\mathbf{q}) = O(\mathbf{v})$. Но тогда

$$o(g(\mathbf{q}+\mathbf{v}) - g(\mathbf{q})) = o(O(\mathbf{v})) = o(\mathbf{v}).$$

Следовательно, мы получили, что

$$g(\mathbf{q}+\mathbf{v}) - g(\mathbf{p}) = (df_{\mathbf{p}})^{-1}(\mathbf{v}) + o(\mathbf{v}),$$

т. е. функция g дифференцируема в точке $\mathbf{q}$, и ее производная равна

$$dg_{f(\mathbf{p})} = (df_{\mathbf{p}})^{-1}.$$

Тем самым мы доказали следующую теорему[10].

Теорема об обратной функции. *Пусть $f : U \to V$ — непрерывно дифференцируемая функция, где $f(\mathbf{p}_0) = \mathbf{q}_0$, и отображение $df_{\mathbf{p}_0}$ обратимо. Тогда вокруг точек $\mathbf{q}_0$ и $\mathbf{p}_0$ существуют такие шарики B и C, что имеется единственное отображение $g : B \to C$, для которого $f \circ g = (id)$. Это отображение непрерывно дифференцируемо и*

$$dg_{f(\mathbf{p})} = (df_{\mathbf{p}})^{-1}.$$

Теорема о неявной функции. Давайте рассмотрим некоторые следствия теоремы существования обратной функции. Предположим, что $G : \mathbb{R}^2 \to \mathbb{R}^1$, $G(x_0, y_0) = 0$ и

$$\frac{\partial G}{\partial y} \neq 0 \quad \text{в точке} \quad (x_0, y_0).$$

[10] В связи с формулировкой этой теоремы см. сноску 9. — *Прим. ред.*

Речь будет идти о том, чтобы решить (хотя бы локально) уравнение $g(x, y) = 0$ относительно y.

Определим отображение $f : \mathbb{R}^2 \to \mathbb{R}^2$ формулой

$$f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x \\ G(x,y) \end{pmatrix}.$$

Тогда

$$df = \begin{pmatrix} 1 & 0 \\ \dfrac{\partial G}{\partial x} & \dfrac{\partial G}{\partial y} \end{pmatrix}$$

является матрицей, несингулярной в точке $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$. Согласно теореме существования обратной функции, можно найти отображение g, для которого $f \circ g = (id)$. Тогда можно написать

$$g\left(\begin{pmatrix} u \\ v \end{pmatrix}\right) = \begin{pmatrix} F(u,v) \\ H(u,v) \end{pmatrix}$$

и уравнение $f \circ g = (id)$ преобразуется к виду

$$\begin{aligned} F(u,v) &= u, \\ G(F(u,v), H(u,v)) &= v. \end{aligned}$$

Подставив первое уравнение во второе, получим

$$G(u, H(u,v)) = v,$$

а выбрав $v = 0$ и обозначив $h(u) = H(u, 0)$, имеем

$$G(u, h(u)) = 0.$$

Функция $h(u)$ дифференцируема и является единственным решением этого уравнения. Единственность и дифференцируемость функции h составляет содержание *теоремы существования неявной функции*. Таким образом, теорема существования неявной функции от одной переменной есть *следствие* теоремы существования обратной функции, зависящей от двух переменных. Сформулируем эту теорему еще раз.

Теорема о неявной функции. *Пусть G — дифференцируемая функция, для которой $G(x_0, y_0) = 0$ и $(\partial G/\partial y)(x_0, y_0) \neq 0$. Тогда существует единственная функция $h(x)$, определенная вблизи $x = x_0$, удовлетворяющая условиям: $h(x_0) = y_0$ и $G(x, h(x)) \equiv 0$. Функция h дифференцируема и $h'(x) = -(\partial G/\partial x)/(\partial G/\partial y)$.*

Предыдущую аргументацию можно сформулировать по-другому. Простейшее отображение из $\mathbb{R}^2 \to \mathbb{R}^1$ (за исключением постоянного отображения), которое можно придумать, есть проекция на одну из осей:

$$\pi : \mathbb{R}^2 \to \mathbb{R}^1 \quad \pi\left(\begin{pmatrix} u \\ v \end{pmatrix}\right) = v.$$

Пусть $G : \mathbb{R}^2 \to \mathbb{R}^1$ — любое непрерывно дифференцируемое отображение и предположим, что $dG_{\mathbf{p}_0}$ *сюръективно* (что в нашем случае одномерной области значений означает $dG_{\mathbf{p}_0} \neq 0$). Утверждается, что существуют локальные *замены координат*, т. е. отображения $g : \mathbb{R}^2 \to \mathbb{R}^2$, которые локально определены и имеют дифференцируемые обратные отображения, так что

$$G \circ g = \pi.$$

Действительно, поскольку $dG_{\mathbf{p}_0} \neq 0$, мы можем сначала линейной заменой добиться, чтобы $\partial G/\partial y \neq 0$. Тогда вышеприведенные соображения позволяют построить отображение g, для которого $G \circ g = \pi$. Таким образом, если допускаются произвольные замены координат, то самое общее непрерывно дифференцируемое отображение, у которого $dG_{\mathbf{p}}$ сюрьективно, есть проекция на одну из осей. На рис. 6.8 вы видите пример функции $G(x, y) = (x^2 + y^2)^{1/2}$ (изображены линии $g(x, y) = \text{const}$), которая после перехода к полярным координатам имеет вид $G \circ g(\theta, r) = r$, т. е. задает проекцию точки (θ, r) на ось r.

Теперь поговорим об отображениях из $\mathbb{R}^1$ в $\mathbb{R}^2$. Простейшее нетривиальное отображение такого вида — это отображение i, которое вставляет пространство $\mathbb{R}^1$ в $\mathbb{R}^2$ как ось x:

$$i : \mathbb{R}^1 \to \mathbb{R}^2, \quad i(x) = \begin{pmatrix} x \\ 0 \end{pmatrix}.$$

Мы утверждаем, что если $G : \mathbb{R}^1 \to \mathbb{R}^2$ — любое непрерывно дифференцируемое отображение с условием $dG_{\mathbf{p}_0} \neq 0$, то локально

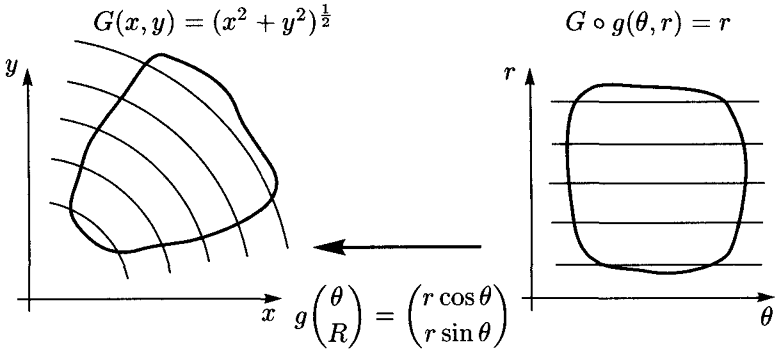


Рис. 6.8

можно найти замену координат, т. е. непрерывно дифференцируемое отображение f, имеющее дифференцируемое обратное, так что (см. рис. 6.9)

$$f \circ G = i.$$

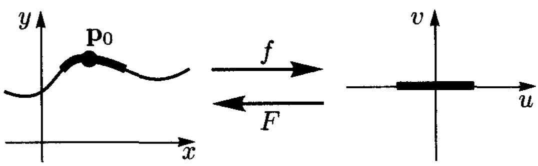


Рис. 6.9

Действительно, сначала можно выполнить линейное преобразование переменных на плоскости так, чтобы в новой состоме координат отображение G имело вид

$$G(t) = \begin{pmatrix} G_1(t) \\ G_2(t) \end{pmatrix}, \quad dG_{\mathbf{p}_0} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Теперь определим $F : \mathbb{R}^2 \to \mathbb{R}^2$ формулой

$$F\left(\begin{pmatrix} u \\ v \end{pmatrix}\right) = \begin{pmatrix} G_1(u) \\ G_2(u) + v \end{pmatrix}.$$

Тогда

$$dF_{G(\mathbf{p}_0)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

и, следовательно, F имеет (локально) непрерывно дифференцируемое обратное отображение. Из определения F и i следует, что

$$F \circ i = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix} = G.$$

Поэтому, взяв $f = F^{-1}$, получаем

$$i = f \circ G.$$

Итак, изменяя систему координат, можно локально «выпрямить любую кривую».

6.5. Поведение функции вблизи критической точки

Предположим, что $f : \mathbb{R}^2 \to \mathbb{R}$ имеет критическую точку в $\mathbf{p}_0$, т. е. $df_{\mathbf{p}_0} = 0$. Будем считать, что мы уже сделали преобразование координат так, что теперь $\mathbf{p}_0 = \mathbf{0}$ в $\mathbb{R}^2$. Допустим, что $d^2 f_\mathbf{0}$ невырождена, т. е. $\mathrm{Det}(d^2 f_\mathbf{0}) \neq 0$. Другими словами, предположим, что симметричная матричная функция

$$H = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f}{\partial x_2^2} \end{pmatrix}$$

невырождена в $\mathbf{0}$. Из результатов параграфа 4.2 мы знаем, что можно найти такое линейное преобразование координат L, что $LH(\mathbf{0})L^T$ может быть принять одну из трех форм:

$$LH(\mathbf{0})L^T = \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\ \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \\ \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}. \end{cases}$$

Допустим, что мы уже сделали линейное преобразование координат, и $d^2 f_{\mathbf{0}}$ уже записывается в виде одной из этих стандартных форм. Из доказательства разложения Тейлора мы знаем, что

$$\begin{aligned} f(x,y) &= f(\mathbf{0}) + b_{11}(x,y)x^2 + 2b_{12}(x,y)xy + b_{22}(x,y)y^2 \\ &= f(\mathbf{0}) + (x,y)\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\ &= f(\mathbf{0}) + (x,y)B\begin{pmatrix} x \\ y \end{pmatrix}, \end{aligned}$$

где b_{ij} — непрерывные функции x и y, а матричная функция B, вычисленная в начале координат, равна $d^2 f_{\mathbf{0}}$, т. е.

$$B(\mathbf{0}) = H(\mathbf{0})$$

или

$$b_{11}(\mathbf{0}) = \frac{\partial^2 f}{\partial x^2}(\mathbf{0}), \quad b_{12}(\mathbf{0}) = \frac{\partial^2 f}{\partial x \partial y}(\mathbf{0}), \quad b_{22}(\mathbf{0}) = \frac{\partial^2 f}{\partial y^2}(\mathbf{0}).$$

Матрица B симметрична. Применим процедуру Грама–Шмидта к $B(\mathbf{x})\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ и $B(\mathbf{x})\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Из непрерывности B ясно, что произведения

$$(1,0)B(\mathbf{x})\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad (1,0)B(\mathbf{x})\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (0,1)B(\mathbf{x})\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

непрерывно зависят от $\mathbf{x}$. Следовательно, для $\mathbf{x}$, достаточно близких к нулю, можно найти такую обратимую матрицу $Q(\mathbf{x})$ (заданную процедурой Грама–Шмидта), что

$$B(\mathbf{x}) = Q(\mathbf{x})^T B(\mathbf{0}) Q(\mathbf{x}).$$

Алгоритм Грама–Шмидта обеспечивает дифференцируемость функции $Q(\mathbf{x})$. Таким образом,

$$f(\mathbf{x}) = f(\mathbf{0}) + \mathbf{x}^T Q(\mathbf{x})^T H(\mathbf{0}) Q(\mathbf{x})\mathbf{x},$$

или

$$f(\mathbf{x}) = f(\mathbf{0}) + \mathbf{y}^T H(\mathbf{0})\mathbf{y},$$

где

$$\mathbf{y} = Q(\mathbf{x})\mathbf{x}.$$

Отображение $\mathbf{x} \mapsto \mathbf{y}$, заданное этой формулой, обратимо согласно теореме об обратной функции! Более подробно: пусть $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ определена формулой

$$\phi(\mathbf{x}) = Q(\mathbf{x})\mathbf{x}.$$

Тогда, дифференцируя произведение, получаем

$$d\phi_{\mathbf{0}}(\mathbf{x}) = (dQ_{\mathbf{0}}(\mathbf{x}))\mathbf{0} + Q(\mathbf{0})\mathbf{x}$$

или

$$d\phi_{\mathbf{0}} = (id).$$

Таким образом, теорема существования обратной функции гарантирует, что ϕ имеет дифференцируемую обратную

$$\mathbf{x} = \psi(\mathbf{y}).$$

Но тогда $(\psi^* f)(\mathbf{y}) = f(\mathbf{x})$, поэтому

$$\psi^* f(\mathbf{y}) = f(\mathbf{0}) + \mathbf{y}^T H(\mathbf{0})\mathbf{y}.$$

Другими словами, функция $\psi^*(f - f(0,0))$ — квадратичная форма! Итак, мы доказали, что вблизи любой невырожденной критической точки можно ввести такие координаты y_1, y_2, что

$$f(y_1, y_2) = f(\mathbf{0}) + Q(y_1, y_2),$$

где

$$Q(y_1, y_2) = \pm(y_1^2 + y_2^2) \quad \text{или} \quad Q(y_1, y_2) = -y_1^2 + y_2^2.$$

Что выбрать из этих трех вариантов, определяется нормальной формой (числом отрицательных собственных значений) второго дифференциала $d^2 f_{\mathbf{0}}$.

Это доказательство имеет общий характер, оно применимо для любой размерности. Итак, если $\mathbf{0}$ — невырожденная критическая точка f, то можно найти такие координаты, в которых

$$f(\mathbf{y}) = f(\mathbf{0}) + Q(\mathbf{y}),$$

где
$$Q(\mathbf{y}) = \pm y_1^2 \pm y_2^2 + \cdots .$$
Число знаков минус в этой формуле (называемое *индексом* Q), равно числу отрицательных собственных значений матрицы $d^2 f_{\mathbf{0}}$. Этот результат известен как *лемма Морса*. Мы воспользуемся этой леммой при изучении асимптотических интегралов в главе 21.

Резюме

A. Производные высших порядков

Вы должны уметь писать разложение Тейлора для функции, определенной на плоскости, через частные производные первого и второго порядка.

Вы должны уметь применять цепное правило, чтобы выразить частные производные второго порядка или уравнения в частных производных второго порядка через новые координаты.

B. Критические точки

Надо уметь находить критические точки функции, определенной на плоскости, и классифицировать их, т. е. определять максимум, минимум и седловую точки.

Надо уметь пользоваться методом множителей Лагранжа для нахождения условных критических значений функции нескольких переменных.

C. Обратные функции

Вы должны знать формулировку теоремы существования обратной функции и уметь пользоваться ей.

Вы должны уметь применять метод Ньютона для нахождения приближенного решения уравнения $f(\mathbf{p}) = \mathbf{q}$, где $f : \mathbb{R}^2 \to \mathbb{R}^2$.

Задачи

6.1. Пусть $F(x,y) = \begin{cases} 0 & \text{для } x = y = 0; \\ \dfrac{x^3 y}{x^2 + y^2} & \text{в других точках.} \end{cases}$

(a) Вычислите $\partial F/\partial x$ и $\partial F/\partial y$. Непрерывны ли они в $(0,0)$?

(b) Вычислите $\partial^2 F/\partial x\partial y$ и $\partial^2 F/\partial y\partial x$. Непрерывны они в $(0,0)$ или нет? (Замечание: если эти производные не непрерывны в (0,0), то их нельзя вычислить, получая общую формулу и устремляя x и y к нулю!)

(c) Покажите, что $\partial^2 F/\partial x\partial y(0,0) \neq \partial^2 F/\partial y\partial x(0,0)$.

(d) Придумайте гладкую кривую, которая проходит через начало координат и определяется формулами $x = X(t)$, $y = Y(t)$ с условием $X(0) = Y(0) = 0$ так, чтобы функция

$$G(t) = F(X(t), Y(t))$$

была не дифференцируема в начале координат.

6.2. Найдите и проклассифицируйте все критические точки функции $F : \mathbb{R}^2 \to \mathbb{R}$, определенной формулой

$$F(x,y) = x^3 + y^3 - 3xy.$$

6.3. Пусть $F(x,y) = x^2y - 3xy + \frac{1}{2}x^2 + y^2$.

(a) Найдите уравнение плоскости, касательной к функции $z = F(x,y)$ в точке $x = 2$, $y = 2$, $z = 2$.

(b) Функция $F(x,y)$ имеет три критические точки, две из которых лежат на прямой $x = y$. Найдите эти критические точки и выясните, где будет максимум, минимум или седловая точка.

6.4. Функция F в пространстве $\mathbb{R}^2$ определена формулой

$$F(x,y) = x^2 - 4xy + y^2 - 6x^{-1}.$$

(a) Получите уравнение плоскости, касательной к $z = F(x,y)$ в точке $x = -1$, $y = -2$.

(b) Найдите критические точки этой функции и выясните их природу.

6.5. Найдите и проклассифицируйте все критические точки функции $F(x,y) = y^2 + (x^2 - 3x)\log y$, определенной в верхней полуплоскости $y > 0$.

6.6. Покажите, что функция $F(x,y) = y(e^{4x} - 1) + 9x^2 + 6y^2$ имеет критическую точку в начале координат и определите природу этой точки. Опишите линии уровня $F(x,y)$ в окрестности начала координат. Нарисуйте пару типичных кривых. Опишите линии уровня $F(x,y)$ в окрестности точки $x = 0$, $y = 1$ и нарисуйте типичные кривые.

6.7. Найдите критические точки функции

$$F(x, y) = 5x^3 - 3x^2y + 6xy^2 - 4y^3 - 27x + 27y$$

и определите их природу. (Критические точки имеют целые координаты.)

6.8. (a) Найдите критические точки функции $F(x, y) = xy^2e^{-(x+y)}$.

(b) Определите природу тех критических точек, которые не находятся в начале координат. Нарисуйте как можно точнее несколько кривых уровня вблизи каждой точки.

(c) Если критическая точка совпадает с началом координат, то в ней гессиан равен нулю. Выясните природу критической точки в начале координат и нарисуйте несколько кривых уровня в его окрестности.

6.9. Пусть x и y — обычные аффинные координатные функции на плоскости. Определим другую пару координатных функций в правой полуплоскости ($x > 0$):

$$u = x^2 - y^2, \qquad v = 2xy.$$

(a) Выразите du и dv через dx и dy. Напишите матрицу, выражающую $\begin{pmatrix} du \\ dv \end{pmatrix}$ через $\begin{pmatrix} dx \\ dy \end{pmatrix}$ в точке P с координатами $x = 2$, $y = 1$, $u = 3$, $v = 4$.

(b) Найдите приближенные значения координат x и y для такой точки Q, что $u(Q) = 3.5$, $v(Q) = 4$.

(c) Пусть ϕ обозначает электрический потенциал на плоскости. Для выбранной точки $P(x = 2,\ y = 1,\ u = 3,\ v = 4)$, где $(\partial\phi/\partial u) = 2$ и $\partial\phi/\partial v = -1$, вычислите $\partial\phi/\partial x$ и $\partial\phi/\partial y$. Найдите направление, по которому функция ϕ изменяется быстрее всего.

(d) В той же точке выразите $\partial^2\phi/\partial y\partial x$ через частные производные ϕ по u и v. Конечно, ответ может включать явный вид функций от x и y.

6.10. Предположим, что координаты u и v выражаются через x и y согласно формуле

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Пусть f — дважды дифференцируемая функция на плоскости. Покажите, что

$$\frac{\partial^2 f}{\partial u^2} + \frac{\partial^2 f}{\partial v^2} = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}.$$

6.11. Полярные координаты r, θ связаны с декартовыми координатами соотношением

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}.$$

Предположим, что функция $f : \mathbb{R}^2 \to \mathbb{R}$ удовлетворяет уравнению Лапласа:

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0.$$

Выразите это уравнение через производные f по r и θ.

6.12. Пусть $f : \mathbb{R}^2 \to \mathbb{R}$ — дважды дифференцируемая функция и пусть $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}$. Выразите $\partial^2 f/\partial\theta^2$ через частные производные f по x и y.

6.13. Пусть $f : \mathbb{R}^2 \to \mathbb{R}$ — дважды дифференцируемая функция. Пусть, далее, $x \neq 0$ и

$$\begin{pmatrix} r \\ \theta \end{pmatrix} = \begin{pmatrix} \sqrt{x^2 + y^2} \\ \arctan(y/x) \end{pmatrix}.$$

Выразите $\partial^2 f/\partial x^2$ через частные производные f по r и θ.

6.14. Дано, что $\partial f/\partial x = f + \partial f/\partial y$. Покажите, что

$$\partial^2 f/\partial x^2 - \partial^2 f/\partial y^2 = f + 2(\partial f/\partial y).$$

6.15. Воспользуемся полярными координатами, данными в упражнении 6.11.

(a) Пусть функция f определена на плоскости. Допустим, что в точке с координатами $x = 3$, $y = 4$ частные производные равны $\partial f/\partial x = 2$ и $\partial f/\partial y = 1$. Вычислите $\partial f/\partial r$ и $\partial f/\partial\theta$ в этой точке.

(b) Предположим, что f удовлетворяет уравнению в частных производных

$$\frac{\partial^2 f}{\partial r^2} + \frac{1}{r}\frac{\partial f}{\partial r} + \frac{1}{r^2}\frac{\partial^2 f}{\partial\theta^2} = 0.$$

Преобразуйте его в уравнение, содержащее только частные производные f по x и y.

6.16. Предположим, что на плоскости определена функция f, удовлетворяющая уравнению Лапласа $\partial^2 f/\partial x^2 + \partial^2 f/\partial y^2 = 0$. Преобразуйте это уравнение к координатам, приведенным в упражнении 5.13. В него могут войти $\partial^2 f/\partial u^2$, $\partial^2 f/\partial v^2$, $\partial^2 f/\partial u \partial v$, $\partial f/\partial u$, $\partial f/\partial v$, u и v, но не должны входить x, y и частные производные по x и y.

6.17. Пусть есть отображение $\phi : \mathbb{R}^2 \to \mathbb{R}^2$:

$$\phi\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} e^x + e^y \\ e^x - e^y \end{pmatrix}.$$

Покажите, что для ϕ в окрестности любой точки можно получить обратную функцию. Вычислите якобиан обратного отображения.

6.18. В пространстве $\mathbb{R}^3$ рассмотрите поверхность $z = F(x, y)$, где $F(x, y) = x^2 - 2xy + 2y^2 + 3x + 4y$.

(a) Получите *наилучшее аффинное приближение* к F в окрестности точки $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

(b) Напишите уравнение плоскости, касательной к поверхности в точке $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

(c) Получите уравнение линии, перпендикулярной к поверхности в этой же точке.

(d) Уравнение $F(x, y) = 8$ определяет функцию $y = g(x)$. Вычислите $g'(1)$.

6.19. В статистической механике существует важная проблема. Рассмотрим физическую систему, которая может иметь энергии $+E$, 0 и $-E$. Пусть x обозначает вероятность системе иметь энергию $+E$ и пусть y обозначает вероятность системе иметь энергию $-E$. Тогда вероятность системе иметь нулевую энергию равна $1-x-y$. Получите максимум *энтропии* S, если она определена соотношением

$$S(x, y) = -x \log x - y \log y - (1 - x - y) \log(1 - x - y)$$

при условии, что средняя энергия равна E_0, т. е.

$$F(x, y) = xE - yE = E_0.$$

Решите эту задачу, введя множитель Лагранжа β. Покажите, что вероятность x пропорциональна $e^{-\beta E}$, а вероятность y пропорциональна $e^{\beta E}$. Найдите коэффициенты пропорциональности.

6.20. Рассмотрите функцию $\alpha : \mathbb{R}^2 \to \mathbb{R}^2$, определенную формулой

$$\alpha(P) = F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} xy^{3/2} \\ x^{3/2} - y^2 \end{pmatrix}.$$

(a) Вычислите матрицу якобиана 2×2, которая представляет линейную часть наилучшего аффинного приближения к функции α вблизи точки $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$. Вычислите приближенное значение $F\left(\begin{pmatrix} 4.2 \\ 1.1 \end{pmatrix}\right)$, воспользовавшись этой матрицей.

(b) С помощью той же матрицы получите приближенное решение уравнения $F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \left(\begin{pmatrix} 4.2 \\ 6.6 \end{pmatrix}\right)$.

6.21. Пусть $f(x, y) = \sqrt{x^2y^4 + 9x^2}$.

(a) Получите *наилучшее аффинное приближение* к этой функции вблизи точки $x = 1$, $y = 2$.

(b) Возьмем точку $x = 1$, $y = 2$. По какому направлению скорость изменения функции $f(x, y)$ максимальна?

(c) Решением системы уравнений

$$f(x, y) = 5,$$
$$x + y = 3$$

является точка $x = 1$, $y = 2$. Получите приближенное решение уравнения

$$f(x, y) = 5.34,$$
$$x + y = 3.05,$$

используя наилучшее аффинное приближение, полученное в первом пункте.

6.22. Функции s и t определены через аффинные координаты функций x и y в области $x > 0, y > 0$ следующими формулами:

$$s = xy, \qquad t = \log y - \log x.$$

(a) Выразите дифференциалы ds и dt через dx и dy в точке $x = 1$, $y = 2$.

(b) В точке $x = 1$, $y = 2$ значения функций s и t равны $s = 2$, $t = \log 2 \approx 0.693$. Используйте матрицу Якоби в этой точке и получите приближенные значения координат x и y для точки, где $s = 2.02$ и $t = 0.723$.

(c) Пусть f — дважды дифференцируемая функция, определенная на плоскости. Выразите $\partial f/\partial x$, $\partial f/\partial y$ и $\partial^2 f/\partial y \partial x$ через x, y и частные производные f по s и t.

Глава 7

Дифференциальные формы и криволинейные интегралы

Главы 7 и 8 являются введением к изучению многомерного интегрального исчисления. Глава 7 посвящена линейным дифференциальным формам и соответствующим криволинейным интегралам. Особое внимание уделено замене переменных в криволинейных интегралах. Обсуждаются некоторые другие одномерные интегралы, в частности, вычисление длины кривой.

Введение

В этой главе мы выясним геометрический смысл линейных дифференциальных форм, в частности, объектов, которые предстоит интегрировать по ориентированным траекториям, в результате чего получаются числа. Начнем с примеров. Рассмотрим форму

$$\omega = \frac{1}{2}(x\,dy - y\,dx).$$

По определению, эта форма задает правило, которое каждой точке $\begin{pmatrix} x \\ y \end{pmatrix}$ ставит в соответствие вектор-строку $\frac{1}{2}(-y, x)$. Далее, вектор-строка понимается как линейная функция векторов нашего пространства. В рассматриваемом случае вектор-строка

$\frac{1}{2}(-y, x)$ ставит в соответствие вектору $\mathbf{h} = \begin{pmatrix} r \\ s \end{pmatrix}$ число

$$\omega_{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)}[\mathbf{h}] = \frac{1}{2}(xs - yr),$$

которое равно ориентированной площади треугольника с вершинами в нуле и в точках $\begin{pmatrix} x \\ y \end{pmatrix}$, $\begin{pmatrix} x+r \\ y+s \end{pmatrix}$ (рис. 7.1).

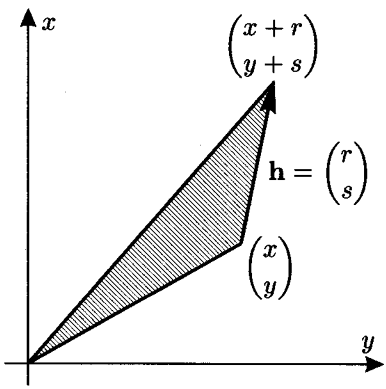


Рис. 7.1

Предположим (рис. 7.2), что у нас есть кривая $\alpha(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$. Выберем на этой кривой точки $\mathbf{p}_i$ и пусть $\mathbf{h}_i = \mathbf{p}_{i+1} - \mathbf{p}_i$. Тогда сумма $\sum \omega_{\mathbf{p}_i}[\mathbf{h}_i]$ даст нам полную ориентированную площадь изображенных на рис. 7.2 треугольников.

В пределе ломаная линия, соединяющая точки $\mathbf{p}_i$, аппроксимирует кривую. Мы ожидаем, что эта сумма будет стремиться к пределу — ориентированной площади, заметенной радиусом-вектором, движущимся вдоль нашей кривой. Мы уже встречались с этим понятием, когда изучали второй закон Кеплера.

Второй пример. В нем появляется понятие *силового поля.* Силовое поле

$$\omega = F\,dx + G\,dy + H\,dz$$

в каждой точке трехмерного пространства определяет линейную функцию с матрицей

$$(F, G, H).$$

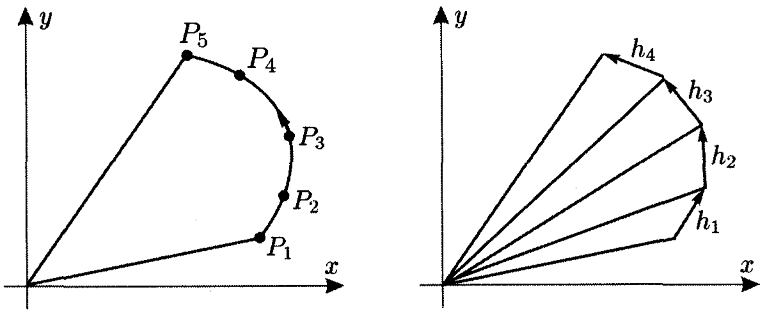


Рис. 7.2

Эта линейная функция задает элемент работы *инфинитезимального движения*, т. е. любому вектору смещения **v** в точке **p** ставит в соответствие

$$(F,\ G,\ H)\begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = Fv_x + Gv_y + Hv_z.$$

Мы ожидаем, что эта величина может быть проинтегрирована вдоль любой траектории Γ, в результате чего получается $\int_\Gamma \omega$ — работа, совершенная при движении вдоль траектории Γ. Заметим, что хотелось бы уметь получать работу при движении вдоль всех траекторий. Представим себе двумерное пространство, где действует силовое поле

$$\omega = F\,dx + G\,dy.$$

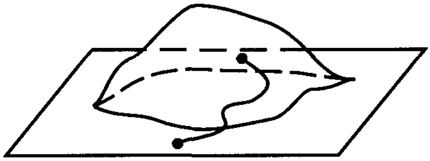


Рис. 7.3

Например, на рис. 7.3 изображено влияние гравитации при движении по поверхности $z = f(x, y)$. В этом случае силовое поле пропорционально

$$\omega = df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy.$$

Допустим, что у нас есть идеально обратимый электрический автомобиль. (Идеально обратимый означает, что вся энергия торможения возвращается в аккумуляторную батарею, при этом не учитывается ни сопротивление воздуха, ни любое другое сопротивление.) Будем измерять, сколько энергии поступает в батарею и сколько расходуется. Разницу между начальным и конечным показаниями батареи обозначим $B_F - B_I$. Кроме того, введем кинетическую энергию автомобиля в начале и в конце пути — KE_I и KE_F. Из закона сохранения энергии следует, что

$$KE_F - KE_I + B_F - B_I = \int_\Gamma \omega,$$

что равно работе, выполненной при движении. (В этом рассуждении мы предполагаем, что силы *не зависят* от скорости, т. е. мы рассматриваем такое силовое поле, для которого силы зависят только от пространственных координат.)

Отметим некоторое отличие от использования понятия силы в законах Ньютона. Если мы пустили камень катиться по поверхности, то законы Ньютона *предсказывают*, как он будет двигаться дальше: движение камня определяется уравнениями

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}, \quad \mathbf{p} = m\mathbf{v} \text{ — импульс}, \quad \mathbf{v} \text{ — скорость}.$$

Рис. 7.4

В нашем обсуждении мы интересуемся величиной энергии, использованной при движении вдоль данной кривой. Силовое поле ω приписывает энергии траекториям Г. Если $\omega = df$, то полная работа, совершенная при движении вдоль траектории, есть разность потенциалов

$$f(Q) - f(P),$$

где P и Q — начальная и конечная точки нашего пути (рис. 7.4).

Ну а теперь, имея в виду все эти соображения, перейдем на математический язык.

7.1. Ориентированные траектории

Ориентированной траекторией на плоскости (или в пространстве $\mathbb{R}^k$) мы будем называть кривую, на которой выбрано определенное направление движения. Траектория вида Γ_1 (рис. 7.5), начальная и конечная точки которой не совпадают, имеет определенное начало (точка P_a) и определенный конец (точка P_b). Перестановка начальной и конечной точек изменяет ориентацию кривой на противоположную. Замкнутая кривая вида Γ_2 (рис. 7.6) не имеет определенных начальной и конечной точек. Любая точка P может быть началом траектории и одновременно ее концом. Для таких замкнутых траекторий, тем не менее, можно задать направление движения, т. е. ориентацию. Тогда для любого фрагмента такой траектории определится его начало и конец.

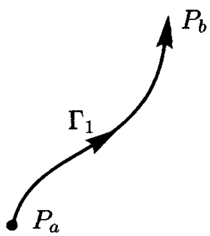


Рис. 7.5

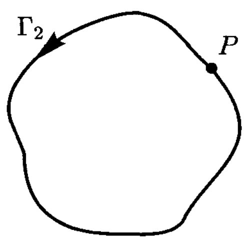


Рис. 7.6

Физически это можно истолковать так: мы имеем траекторию движения частицы, причем известно, через какие точки проходит частица и в каком порядке, но скорость ее движения мы не знаем. Вполне возможно, что какие-то отрезки пути проходятся два или более раз. Например, частица может пройти два раза по окружности единичного радиуса против часовой стрелки, или (рис. 7.7) пройти от точки P_a до точки Q, обратно до точки R и только потом к точке P_b. Такие траектории, может быть, трудно изображать кривыми с соответствующими стрелками, но они имеют физический смысл и, как мы скоро увидим, их легко описывать на языке функций.

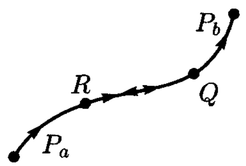


Рис. 7.7

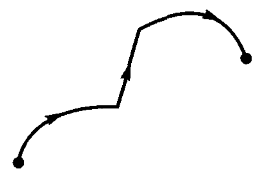


Рис. 7.8

Здесь мы ограничимся рассмотрением только *кусочно-дифференцируемых траекторий* — непрерывных кривых, для которых существуют касательные во всех точках, за исключением их конечного числа (рис. 7.8). Такая траектория может быть описана как образ отрезка вещественной прямой при непрерывном отображении

$$\alpha : \mathbb{R} \to \mathbb{R}^2,$$

которое дифференцируемо везде, кроме конечного числа точек. Функция α называется *параметризацией* траектории (рис. 7.9).

Физически можно считать, что функция α определяет положение частицы в каждый момент времени. Мы можем задавать α, определяя перенос α^* координатных функций, т. е. записывать формулы, которые дают численные значения координат x и y в зависимости от времени. Например, если частица движется вдоль круговой дуги с радиусом R из точки $\begin{pmatrix} R \\ 0 \end{pmatrix}$ в точку $\begin{pmatrix} 0 \\ R \end{pmatrix}$ (рис. 7.10), то ее траекторию можно представить в форме

$$\alpha : t \to \begin{pmatrix} R\cos t \\ R\sin t \end{pmatrix}, \quad 0 \le 2t \le \pi,$$

или

$$\alpha^* x = R\cos t, \quad \alpha^* y = R\sin t.$$

Каждая траектория может быть параметризована по-разному,

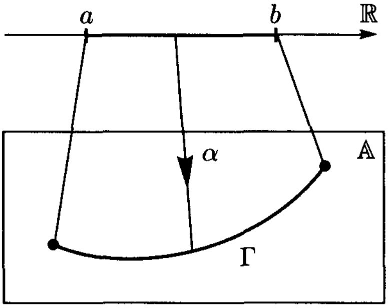


Рис. 7.9

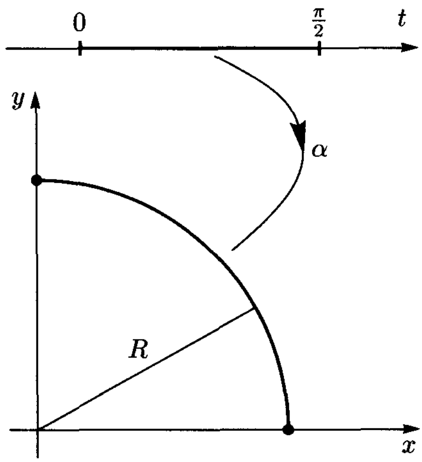


Рис. 7.10

причем каждый способ физически будет соответствовать прохождению траектории с различными скоростями в разные промежутки времени. Например, отрезок гиперболы $y^2 - x^2 = 1$ от точки $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ до точки $\begin{pmatrix} 1 \\ \sqrt{2} \end{pmatrix}$ (рис. 7.11) может быть параметризован следующими способами:

$$\begin{array}{lll} \alpha^* x = t, & \alpha^* y = \sqrt{t^2 - 1}, & 0 \le t \le 1, \\ \alpha^* x = \sinh t, & \alpha^* y = \cosh t & 0 \le t \le \operatorname{arcsinh} 1, \\ \alpha^* x = \tan t, & \alpha^* y = \sec t & 0 \le 4t \le \pi. \end{array}$$

Траектории, на которых отдельные участки кривой проходятся более одного раза, удобно записывать в параметрической форме. Например, параметризация

$$\alpha^* x = \cos t, \quad \alpha^* y = \sin t, \quad 0 \le t \le 4\pi$$

описывает путь, проходимый частицей вдоль окружности единичного радиуса против часовой стрелки два раза, а параметризация

$$\alpha^* x = \sin^2 t, \quad \alpha^* y = \sin^2 t, \quad 0 \le t \le 3\pi/2$$

описывает движение по отрезку прямой от точки $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ вперед к точке $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, обратно к $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ и потом еще раз вперед к точке $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

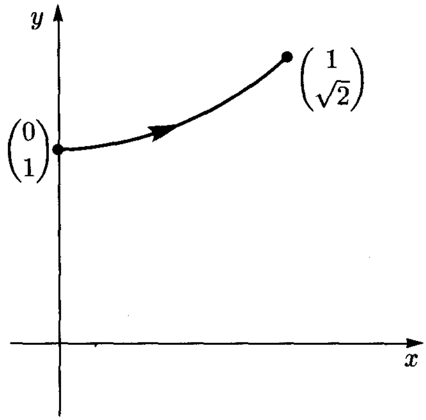


Рис. 7.11

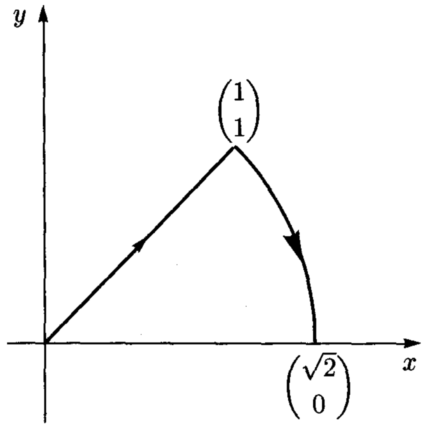


Рис. 7.12

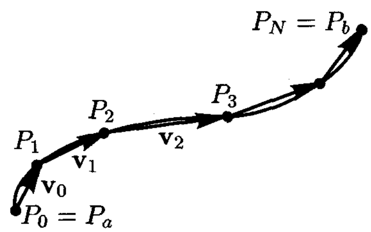


Рис. 7.13

В практических вычислениях каждый дифференцируемый отрезок пути параметризуется отдельно. Например, на рис. 7.12 изображен путь, который можно параметризовать в виде: сначала

$$\alpha^* x = t, \quad \alpha^* y = t, \quad 0 \le t \le 1,$$

а потом

$$\alpha^* x = \sqrt{2}\cos(\pi/4 - t), \quad \alpha^* y = \sqrt{2}\sin(\pi/4 - t), \quad 0 \le t \le \pi/4.$$

Кусочно-дифференцируемый путь можно аппроксимировать последовательностью векторов, уложенных вдоль него (рис. 7.13). Разобьем кривую на отдельные участки, расставив на ней по ходу движения точки P_0, P_1, P_2, ..., P_N, где P_0 — начало пути, а P_N — его конец. При этом необходимо включить в это подразбиение все точки, где кривая не имеет касательной. Затем можно ввести векторы смещения

$$\mathbf{v}_0 = P_0P_1, \quad \mathbf{v}_1 = P_1P_2, \quad \mathbf{v}_{N-1} = P_{N-1}P_N.$$

Выбрав точки P_i достаточно близко, мы можем аппроксимировать любую кусочно-гладкую траекторию с помощью ломаной линии, заданной последовательностью векторов с любой наперед заданной точностью. Следует заметить, что есть пути, которые нельзя хорошо аппроксимировать таким способом, но они не могут быть параметризованы дифференцируемыми функциями, и мы не будем их рассматривать.

7.2. Криволинейные интегралы

Теперь опишем естественный способ, когда заданная дифференциальная форма приписывает траектории некоторое вещественное число. Пусть отрезок пути характеризуется точкой P_i с соответствующим вектором $\mathbf{v}_i$. Дифференциальная форма ω приписывает этому отрезку вещественное число $\omega(P_i)[\mathbf{v}_i]$. Вычислим сумму этих чисел по всем отрезкам пути:

$$I_N = \sum_{i=0}^{N-1} \omega(P_i)[\mathbf{v}_i].$$

Эта сумма очень похожа на римановскую сумму для обычного интеграла. Пусть число точек P_i растет так, что все вектора $\mathbf{v}_i$ стремятся к нулю[1]. Вычислим предел суммы I_N в этом случае. Если такой предел существует и не зависит от выбора последовательности точек, то он называется *криволинейным интегралом*[2]

[1]Точнее говоря, следует потребовать, чтобы стремилось к нулю число, равное $\max_i \|\mathbf{v}_i\|$. — *Прим. ред.*

[2]В русскоязычной литературе такие интегралы иначе называют *криволинейными интегралами второго рода*. Конструкция «абсолютного» интеграла по кривой, рассматриваемая авторами в конце этой главы, в русскоязычной литературе называют *криволинейными интегралами первого рода*. — *Прим. ред.*

от дифференциальной формы ω по пути Γ. Его обозначают $\int_\Gamma \omega$, т. е.

$$\int_\Gamma \omega = \lim \sum_{i=0}^{N-1} \omega(P_i)[\mathbf{v}_i].$$

Мы докажем скоро, что для кусочно-дифференцируемых путей написанный предел существует и не зависит от выбора точек деления пути. Дадим формулу для $\int_\Gamma \omega$, используя операцию переноса дифференциальной формы.

Из определения суммы дифференциальных форм и произведения дифференциальной формы на вещественное число следуют три очевидных свойства криволинейных интегралов $\int_\Gamma \omega$.

(1) Интеграл линеен по ω, т. е. если $\omega = \omega_1 + \omega_2$, то $\int_\Gamma \omega = \int_\Gamma \omega_1 + \int_\Gamma \omega_2$. Если $\omega = \lambda\omega'$, то $\int_\Gamma \omega = \lambda \int_\Gamma \omega'$.

(2) Если Γ состоит из двух участков Γ_1 и Γ_2, то

$$\int_\Gamma \omega = \int_{\Gamma_1} \omega + \int_{\Gamma_2} \omega.$$

Это значит, что всегда можно разделить кусочно-дифференцируемую траекторию на дифференцируемые фрагменты, как это показано на рис. 7.14, и свести интегрирование по всей кривой к вычислению интегралов по этим фрагментам.

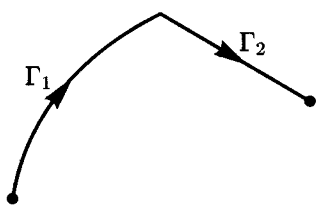


Рис. 7.14

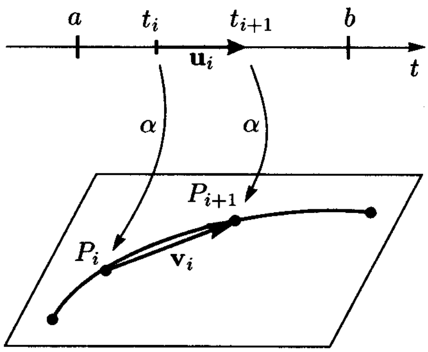


Рис. 7.15

(3) Если Γ отличается от Γ' только ориентацией, то

$$\int_{\Gamma'} \omega = -\int_\Gamma \omega.$$

Это следует из того, что, изменяя ориентацию Γ, мы меняем направление всех векторов $\mathbf{v}_i$, а поскольку ω — линейная форма, то

$$\omega(P_i)[-\mathbf{v}_i] = -\omega(P_i)[\mathbf{v}_i].$$

Перейдем к задаче нахождения численного значения криволинейного интеграла. Наша цель состоит в том, чтобы свести эту задачу к вычислению обычного интеграла по параметру, введенному вдоль пути интегрирования. Параметризация $\alpha : \mathbb{R} \to \mathbb{R}^2$ отображает отрезок вещественной прямой $[a, b]$ в траекторию Γ.

Предположим, что точка a (нижняя граница интервала $[a, b]$) отображается в начальную точку пути P_a, а верхняя граница b — в конечную точку пути P_b. Рассматривая отдельно гладкие участки нашего пути, будем считать, что α задается дифференцируемыми функциями

$$\alpha^* x = X(t), \quad \alpha^* y = Y(t),$$

для которых, согласно цепному правилу,

$$\alpha^* \, dx = X'(t) \, dt, \quad \alpha^* \, dy = Y'(t) \, dt.$$

Пусть разбиение пути соответствует следующему разбиению отрезка:

$$a = t_0 < t_1 < \cdots < t_n = b,$$

так что

$$P_i = \alpha(t_i) = \begin{pmatrix} X(t_i) \\ Y(t_i) \end{pmatrix}.$$

Возьмем дифференциальную форму

$$\omega = g \, dx + h \, dy.$$

Тогда криволинейный интеграл может быть аппроксимирован суммой

$$\begin{aligned}\sum_i \omega(P_i)[\alpha(t_{i+1}) - \alpha(t_i)] \\ &= \sum_i \{g(P_i)\,dx[\alpha(t_{i+1}) - \alpha(t_i)] + h(P_i)\,dy[\alpha(t_{i+1}) - \alpha(t_i)]\} \\ &= \sum_i \{g(\alpha(t_i))[X(t_{i+1}) - X(t_i)] + h(\alpha(t_i))[Y(t_{i+1}) - Y(t_i)]\}.\end{aligned} \tag{7.1}$$

Вспомним, что $\alpha^*\omega = f\,dt$, где

$$f(t) = g(\alpha(t))X'(t) + h(\alpha(t))Y'(t).$$

Покажем, что (при соответствующих предположениях относительно α, f и g) приближенное выражение (7.1) стремится к

$$\int_a^b f\,dt, \tag{7.2}$$

когда длина отрезков разбиения пути стремится к нулю. Итак, мы хотим сравнить (7.1) и (7.2) для данного разбиения и показать, что их разность стремится к нулю, когда $\max_i(t_{i+1}-t_i)$ стремится к нулю. Можно написать

$$X(t_{i+1}) - X(t_i) = \int_{t_i}^{t_{i+1}} X'(s)\,ds$$

и

$$Y(t_{i+1}) - Y(t_i) = \int_{t_i}^{t_{i+1}} Y'(s)\,ds.$$

Тогда сумма (7.1) имеет вид

$$\sum_i \left\{g(\alpha(t_i)) \int_{t_i}^{t_{i+1}} X'(s)\,ds + h(\alpha(t_i)) \int_{t_i}^{t_{i+1}} Y'(s)\,ds\right\},$$

а (7.2) можно написать в виде

$$\begin{aligned}\sum_i \int_{t_i}^{t_{i+1}} f(s)\,ds \\ &= \sum_i \left\{\int_{t_i}^{t_{i+1}} g(\alpha(s))X'(s)\,ds + \int_{t_i}^{t_{i+1}} h(\alpha(s))Y'(s)\,ds\right\}.\end{aligned}$$

Разница между этими выражениями в том, что для (7.1) мы имеем $g(\alpha(t_i))$ или $h(\alpha(t_i))$, стоящими перед знаком интеграла в каждом слагаемом, а в (7.2) функции $g(\alpha(s))$ и $h(\alpha(s))$ стоят под знаком интеграла. Интуитивно ясно, что для непрерывных функций g и h и гладком отображении α эта разница пренебрежимо мала при достаточно мелком разбиении. Сделаем ряд предположений, чтобы получить точную оценку разницы между (7.1) и (7.2). (Можно было бы сделать и более слабые предположения, но тогда потребовалось бы более сложное доказательство.)

(i) Предположим, что g и h равномерно[3] непрерывны, т. е. что для любого $\varepsilon > 0$ существует такое $\eta > 0$, что из неравенства $\|P - Q\| < \eta$ следует $|g(P) - g(Q)| < \varepsilon$ и $|h(P) - h(Q)| < \varepsilon$. Это — предположение относительно ω. По теореме Лагранжа о среднем значении оно будет выполняться (при $\eta N = \varepsilon$), если производные dg и dh удовлетворяют неравенствам $\|dg\| < N$ и $\|dh\| < N$ во всех точках.

(ii) Допустим, что существует такая постоянная M, что

$$|X'(t)| < M \quad \text{и} \quad |Y'(t)| < M$$

для всех значений t. Это — предположение относительно пути α.

По теореме о среднем значении можно найти такое $\delta > 0$, что для любых t' и t'' из условия

$$|t' - t''| < \delta$$

вытекает неравенство

$$|\alpha(t') - \alpha(t'')| < \eta.$$

Разделим весь путь на отрезки так, чтобы для всех i выполнялось неравенство $|t_{i+1} - t_i| < \delta$. Согласно предположению (i), мы будем иметь

$$|g(\alpha(s)) - g(\alpha(t_i))| < \varepsilon \quad \text{для} \quad t_i \le s \le t_{i+1}.$$

Аналогичную оценку можно получить для h. Тогда

$$\left| g(\alpha(t_i)) \int_{t_i}^{t_{i+1}} X'(t)\,dt - \int_{t_i}^{t_{i+1}} g(\alpha(s)) X'(s)\,ds \right|$$

[3]Известно, что непрерывная на компакте функция всегда равномерно непрерывна. — *Прим. ред.*

$$\leq \int_{t_i}^{t_{i+1}} |g(\alpha(t_i)) - g(\alpha(s))| \cdot |X'(s)|\, ds$$
$$\leq \varepsilon M |t_{i+1} - t_i|.$$

Просуммировав эти неравенства по всем отрезкам (с учетом аналогичной оценки для членов с h), можно увидеть, что разность между (7.1) и (7.2) не больше, чем

$$\sum_i \varepsilon M|t_{i+1} - t_i| + \varepsilon M|t_{i+1} - t_i| = 2\varepsilon M|b - a|.$$

Выбирая длину разбиения с достаточно малым δ, можно получить эту оценку со сколь угодно малой величиной ε. Тем самым доказано равенство нашего криволинейного интеграла и определенного интеграла (7.2).

Поскольку $f\, dt = \alpha^*\omega$, можно написать

$$\int_a^b f\, dt = \int_a^b \alpha^*\omega,$$

где правая часть равенства равна левой по определению. В результате получаем:

$$\int_\Gamma \omega = \int_a^b \alpha^*\omega.$$

В этом равенстве левая часть имеет очевидный интуитивный смысл, а правая удобна для вычисления.

Пример. В качестве применения этого результата вычислим интеграл I от

$$\omega = xy\, dx + x^2\, dy$$

вдоль двух различных путей, соединяющих точку $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ с $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Первый путь Γ_1 идет вдоль параболы $y = 1 - x^2$, а путь Γ_2 — вдоль прямолинейного отрезка, соединяющего эти точки (см. рис. 7.16).

Для параметризации Γ_1 введем функцию

$$\alpha : t \mapsto \begin{pmatrix} t \\ 1 - t^2 \end{pmatrix}, \quad 0 \leq t \leq 1,$$

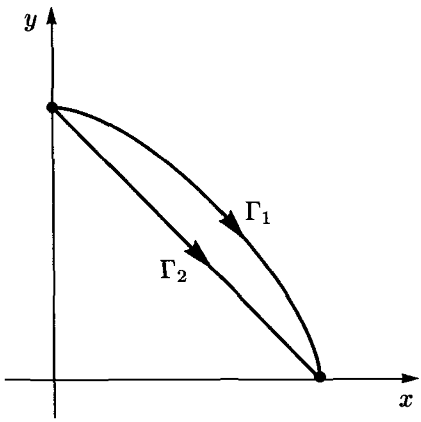


Рис. 7.16

т. е.

$$\alpha^* x = t, \quad \alpha^* y = 1 - t^2,$$

откуда следует, что

$$\alpha^*\, dx = dt, \quad \alpha^*\, dy = -2t\, dt.$$

Направление движения, задающего ориентацию на кривой, соответствует увеличению параметра t: начало кривой — это точка $\alpha(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, а ее конец — точка $\alpha(1) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

Тогда

$$\alpha^* \omega = t(1 - t^2)\, dt + t^2(-2t\, dt) = (t - 3t^3)\, dt$$

и интеграл от ω по этой кривой равен

$$\int_{\Gamma_1} \omega = \int_0^1 (t - 3t^3)\, dt = \left[\frac{1}{2}t^2 - \frac{3}{4}t^4\right]_0^1 = -\frac{1}{4}.$$

Для параметризации кривой Γ_2 введем функцию

$$\beta : t \mapsto \begin{pmatrix} t \\ 1 - t \end{pmatrix},$$

т. е.

$$\beta^* x = t, \quad \beta^* y = 1 - t.$$

Следовательно,

$$\beta^* dx = dt, \quad \beta^* dy = -dt$$

и

$$\beta^*\omega = t(1-t)\,dt + t^2(-dt) = (t - 2t^2)\,dt.$$

Тогда для этого пути получаем

$$\int_{\Gamma_2} \omega = \int_0^1 (t - 2t^2)\,dt = \left[\frac{1}{2}t^2 - \frac{2}{3}t^3\right]_0^1 = -\frac{1}{6}.$$

7.3. Точные дифференциальные формы

В предыдущем примере величина криволинейного интеграла зависит от формы пути, по которому он вычисляется, при фиксированном положении начальной и конечной точки. В общем случае это правильно, но есть один важный частный случай дифференциальных форм, для которых величина криволинейного интеграла полностью определяется начальной и конечной точками кривой Γ. Допустим, что дифференциальная форма ω имеет вид $\omega = df$. В этом случае

$$\int_\Gamma df = \int_a^b \alpha^*(df) = \int_a^b d(\alpha^* f).$$

Согласно фундаментальной теореме интегрального исчисления (формула Ньютона–Лейбница)

$$\int_a^b d(\alpha^* f) = \alpha^* f(b) - \alpha^* f(a).$$

Однако, $\alpha^* f(b) = f(\alpha(b)) = f(P_b)$, где P_b — конечная точка пути Γ, и аналогично $\alpha^* f(a) = f(P_a)$. Отсюда следует вывод, что если начало и конец пути Γ, т.е. точки P_a и P_b, фиксированы, то интеграл

$$\int_\Gamma df = f(P_b) - f(P_a)$$

определяется этими точками и не зависит от выбора пути Γ.

Итак, этот результат в комбинации с предыдущими вычислениями показывает, что не любая дифференциальная форма имеет вид $\omega = df$. Действительно, легко написать необходимое для этого условие: допустим, что

$$\omega = G\,dx + H\,dy = df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy.$$

Поскольку смешанные производные равны[4], т. е. $\partial^2 f/\partial x \partial y = \partial^2 f/\partial y \partial x$, мы имеем

$$\frac{\partial G}{\partial y} = \frac{\partial H}{\partial x}.$$

В вышеприведенном примере $G = xy$ и $H = x^2$, поэтому

$$\frac{\partial G}{\partial y} = x \neq \frac{\partial H}{\partial x} = 2x.$$

Дифференциальная форма ω, которую можно представить в виде $\omega = df$, называется *точной.*

Сейчас мы докажем, что *локально* условие

$$\frac{\partial G}{\partial y} = \frac{\partial H}{\partial x}$$

достаточно, чтобы дифференциальная форма $\omega = G\,dx + H\,dy$ имела вид $\omega = df$ для некоторой функции f, определенной с точностью до константы.

Сначала зафиксируем точку P_a, и пусть $f(P_a) = 0$. Теперь определим функцию f по формуле $f(P) = \int_\Gamma \omega$, где Γ — специально выбранная траектория, соединяющая точки P_a и P. Конечно, к f можно добавить любую константу без изменения ее дифференциала df. Выбор точки P_a определяет эту константу.

Давайте проделаем эту процедуру в декартовых координатах. Допустим, что

$$\omega = G(x, y)\,dx + H(x, y)\,dy.$$

[4]При дополнительном условии непрерывности этих производных. — *Прим. ред.*

Для простоты пусть P_a совпадает с началом координат. Следовательно, $f\left(\begin{pmatrix}0\\0\end{pmatrix}\right) = 0$. В качестве Γ выберем путь, соединяющий начало координат и точку $\begin{pmatrix}x\\y\end{pmatrix}$ прямолинейным отрезком (рис. 7.17), который легко параметризовать:

$$\alpha : t \mapsto \begin{pmatrix}xt\\yt\end{pmatrix} \quad 0 \le t \le 1.$$

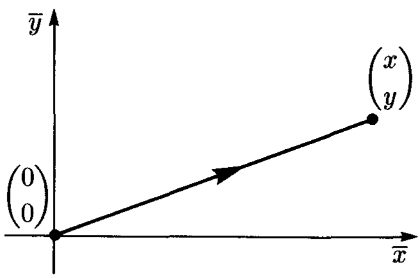


Рис. 7.17

Поскольку буквы x и y обозначают у нас координаты *конечной* точки пути Γ, то для обозначения переменных интегрирования используем буквы $\bar{x}$ и $\bar{y}$. Таким образом,

$$\alpha^*\bar{x} = xt, \qquad \alpha^*\bar{y} = yt,$$
$$\alpha^*\,d\bar{x} = x\,dt, \qquad \alpha^*\,d\bar{y} = y\,dt$$

и

$$\alpha^*\omega = G(xt, yt)x\,dt + H(xt, yt)y\,dt.$$

Тогда

$$f(P) = \int_0^1 \alpha^*\omega,$$

и, следовательно, формула

$$f(x, y) = \int_0^1 [xG(xt, yt) + yH(xt, yt)]\,dt \tag{7.3}$$

позволяет построить функцию f по дифференциальной форме ω. Заметим, что это можно сделать только при условии, что функции G и H определены на всем пути Γ.

До сих пор мы не делали никаких предположений относительно формы ω, кроме того, что она должна быть определена вдоль траектории интегрирования. Поэтому, вообще говоря, мы не ожидаем, что $df = \omega$. Поэтому сейчас мы должны сделать ряд предположений. Давайте вычислим $\partial f/\partial x$. В определении функции f продифференцируем по x под знаком интеграла[5]:

$$\frac{\partial f}{\partial x} = \int_0^1 G(xt, yt)\,dt + \int_0^1 \left(xt\frac{\partial G}{\partial x}(xt, yr) + ty\frac{\partial H}{\partial x}(xt, yt) \right) dt.$$

Теперь воспользуемся тождеством

$$\frac{d}{dt}(tG(xt, yt)) = G(xt, yt) + tx\frac{\partial G}{\partial x}(xt, yt) + ty\frac{\partial H}{\partial y}(xt, yt).$$

Из него следует:

$$\begin{aligned} G(x, y) &= [tG(xt, yt)]_0^1 = \int_0^1 \frac{d}{dt}(tG(xt, yt))\,dt \\ &= \int_0^1 \left(G(xt, yt) + tx\frac{\partial G}{\partial x}(xt, yt) + ty\frac{\partial G}{\partial y}(xt, yt) \right) dt. \end{aligned}$$

Подставив это выражение в $\partial f/\partial x$, получим

$$\frac{\partial f}{\partial x} = G(x, y) + \int_0^1 ty\left(\frac{\partial H}{\partial x}(xt, yt) - \frac{\partial G}{\partial y}(xt, yt) \right) dt.$$

Предположим, что вдоль всего пути

$$\frac{\partial G}{\partial y} = \frac{\partial H}{\partial x}.$$

Тогда $\dfrac{\partial f}{\partial x} = G$. Аналогично $\dfrac{\partial f}{\partial y} = H$. Поэтому $\omega = df$, т.е. дифференциальная форма ω является *точной*.

[5]Для этого достаточна непрерывная дифференцируемость подинтегральной функции. — *Прим. ред.*

Рассмотрим пример восстановления функции по ее дифференциалу в явном виде. Пусть

$$\omega = 2xy^3\,dx + 3x^2y^2\,dy.$$

Поскольку ω определена везде и

$$\frac{\partial}{\partial y}(2xy^3) = 6xy^2 = \frac{\partial}{\partial x}(3x^2y^2),$$

то $\omega = df$. А теперь вычислим f:

$$f(x,y) = \int_0^1 [xG(xt,yt) + yH(xt,yt)]\,dt,$$

$$f(x,y) = \int_0^1 (2x^2y^3 + 3x^2y^3)t^4\,dt = x^2y^3.$$

Действительно,

$$d(x^2y^3) = 2xy^3\,dx + 3x^2y^2\,dy.$$

7.4. Замкнутые дифференциальные формы

В предыдущем параграфе мы показали, что если ω определена в области R, имеющей форму звезды, все точки которой можно соединить с некоторой точкой O, лежащей внутри R, отрезками прямых, также лежащими в R (рис. 7.18), то условие $\partial G/\partial y = \partial H/\partial x$ достаточно для того, чтобы форма ω была точной. Если области определения G и H не удовлетворяют условию звездности, то такой вывод может оказаться неверным.

Например,

$$\omega = -\frac{y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy$$

удовлетворяет условию $\partial G/\partial y = \partial H/\partial x$, но ω не определена в точке $x = 0$, $y = 0$. В этом случае в плоскости с «проколотой

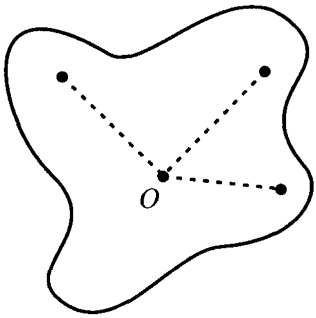


Рис. 7.18

точкой» $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ не существует функции f, для которой $\omega = df$. Действительно, если ω — точная форма, то интеграл от нее по замкнутому пути должен быть равен нулю. Но легко показать, что интеграл от нашей формы по окружности Γ единичного радиуса с центром в начале координат не равен нулю. Возьмем

$$\alpha^* x = \cos t, \qquad \alpha^* y = \sin t,$$
$$\alpha^* dx = -\sin t\, dt, \quad \alpha^* dy = \cos t\, dt,$$

так что

$$\alpha^* \omega = \sin^2 t\, dt + \cos^2 t\, dt = dt.$$

Тогда

$$\int_\Gamma \omega = \int_0^{2\pi} dt = 2\pi.$$

Дифференциальная форма $\omega = G\,dx + H\,dy$, определенная в некоторой области пространства $\mathbb{R}^2$, называется *замкнутой*, если $\partial G/\partial y = \partial H/\partial x$. Если область имеет форму звезды, то мы доказали, что замкнутая форма является точной. В общем случае это не так.

В последующих главах мы вернемся к рассмотрению этих идей, имеющих важнейшее значение для теории электромагнитного поля.

7.5. Перенос дифференциальной формы и интегрирование

Поскольку определение линейного интеграла $\int_\Gamma \omega$ не зависит от выбора параметра вдоль пути Γ, очевидно, что численное значение интеграла не зависит от параметризации. Полезно продемонстрировать эту независимость в явном виде. Предположим, что путь Γ параметризован двумя способами: с помощью переменных s и t.

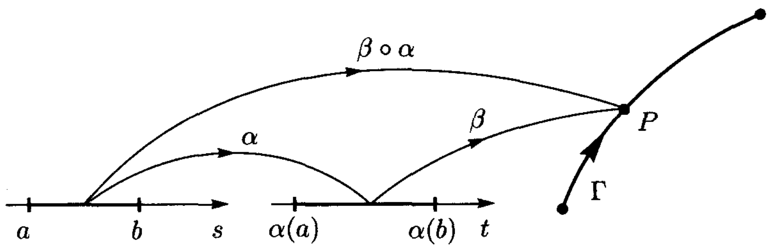


Рис. 7.19

Тогда существует взаимно-однозначное отображение α отрезка s-прямой на отрезок t-прямой, как это показано на рис. 7.19. Следовательно, параметризацию пути Γ можно записать в виде

$$P = \beta(t) \quad \text{и} \quad P = \beta(\alpha(s)) = \beta \circ \alpha(s).$$

Используя параметр s, цепное правило и формулу для замены переменных в обычных определенных интегралах, вычислим

$$\int_\Gamma \omega = \int_a^b (\beta \circ \alpha)^* \omega = \int_a^b \alpha^*(\beta^* \omega) = \int_{\alpha(a)}^{\alpha(b)} \beta^* \omega.$$

Это в точности совпадает с тем, что мы получили бы, вычисляя интеграл с помощью параметра t. С вычислительной точки зрения это значит, что использование разных параметризаций эквивалентно замене переменных в интеграле.

Криволинейный интеграл, заданный в одной плоскости, можно преобразовать в равный ему интеграл в другой плоскости, как это показано на рис. 7.20. Символ β обозначает взаимно-однозначное отображение пути Γ на плоскости $\mathbb{A}$ в путь $\beta(\Gamma)$ на плоскости $\mathbb{B}$. Выбрав дифференциальную форму ω на плоскости $\mathbb{B}$, мы определяем ее перенос согласно формуле

$$\beta^*\omega(P)[\mathbf{v}] = \omega(\beta(P))[d\beta_P[\mathbf{v}]].$$

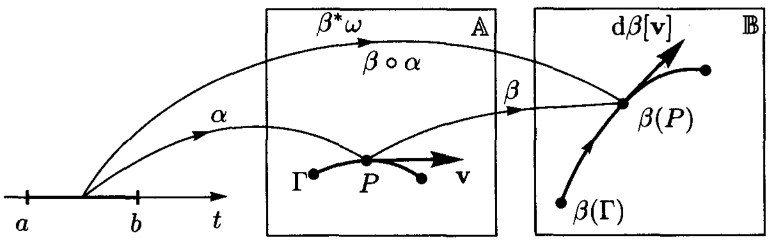


Рис. 7.20

Мы утверждаем, что

$$\int_{\beta(\Gamma)} \omega = \int_{\Gamma} \beta^*\omega.$$

Действительно, по определению криволинейного интеграла

$$\begin{aligned}\int_{\Gamma} \beta^*\omega &= \lim \sum_{i=0}^{N-1} \beta^*\omega(P_i)[\mathbf{v}] \\ &= \lim \sum_{i=0}^{N-1} \omega(\beta(P_i))[d\beta_{P_i}[\mathbf{v}_i]].\end{aligned}$$

Заметим, что при $N \to \infty$ векторы $d\beta[\mathbf{v}_i]$ располагаются вдоль пути $\beta(\Gamma)$ все более близко. Следовательно, последняя сумма в пределе $N \to \infty$ становится интегралом $\int_{\beta(\Gamma)} \omega$. Действительно, если мы параметризуем Γ с помощью отображения α, то согласно цепному правилу получим

$$\int_{\Gamma} \beta^*\omega = \int_a^b \alpha^*(\beta^*\omega) = \int_a^b (\beta \circ \alpha)^*\omega.$$

Однако, $\beta \circ \alpha$ есть параметризация пути $\beta(\Gamma)$, поэтому последний интеграл равен $\int_{\beta(\Gamma)} \omega$, что и требовалось доказать.

Пример. Смысл последнего результата состоит в том, что для вычисления интеграла можно вводить любую удобную систему координат. Например, если мы хотим вычислить интеграл от $\omega = x\,dy - y\,dx$ по полуокружности единичного радиуса от точки $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ до точки $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$ на плоскости xy (рис. 7.21), эту полуокружность можно заменить на направленный отрезок Γ в *полярной системе координат* с помощью отображения $\beta : \begin{pmatrix} r \\ \theta \end{pmatrix} \to \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}$. Тогда

$$\beta^*\omega = (r\cos\theta)\,d(r\sin\theta) - (r\sin\theta)\,d(r\cos\theta) = r^2\,d\theta.$$

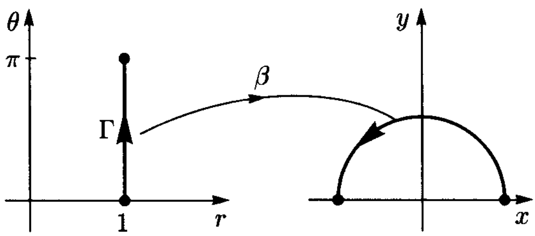


Рис. 7.21

Для отрезка Γ значение $r = 1$ и поэтому

$$\int_\Gamma \beta^*\omega = \int_0^\pi d\theta = \pi.$$

Ясно, что вычисление $\int_{\beta(\Gamma)} \omega$ с помощью параметризации $t \to \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$ дает то же самое значение интеграла.

7.6. Абсолютный криволинейный интеграл. Длина кривой

До сих пор мы рассматривали направленные криволинейные интегралы, вычисляемые вдоль *ориентированных* путей на плоскости, где скалярное произведение может быть не определено.

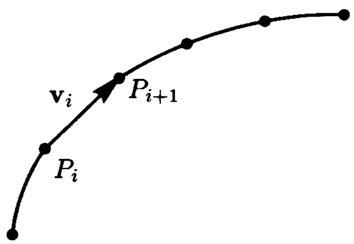


Рис. 7.22

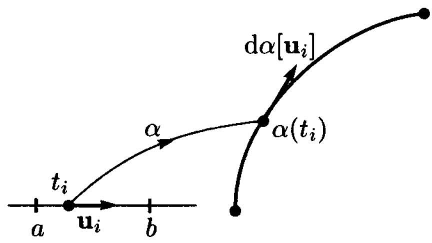


Рис. 7.23

Если скалярное произведение определить, то становится возможным определение *абсолютного* криволинейного интеграла[6] от *функции* f по пути Γ. Чтобы определить такой интеграл $\int_\Gamma f\,ds$, разобьем путь Γ на короткие отрезки с векторами $\mathbf{v}_i = \overline{P_iP_{i+1}}$ (рис. 7.22), после этого вычислим *длину* каждого отрезка по формуле для скалярного произведения $s_i = \|\mathbf{v}_i\| = \sqrt{(\mathbf{v}_i, \mathbf{v}_i)}$. Тогда интеграл снова определяется как предел суммы

$$\int_\Gamma f\,ds = \lim_{N\to\infty} \sum_{i=0}^{N-1} f(P_i)s_i.$$

Очевидно, что в этом случае ориентация пути Γ не имеет значения, так как длина вектора $\mathbf{v}_i$ равна длине вектора $(-\mathbf{v}_i)$.

Для вычисления абсолютного криволинейного интеграла удобно, как и выше, параметризовать путь Γ. Пусть

$$\alpha^* x = X(t), \qquad \alpha^* y = Y(t),$$

откуда

$$\alpha^*\,df = X'(t)\,dt, \qquad \alpha^*\,dy = Y'(t)\,dt.$$

Тогда вектор $\mathbf{u}_i = \overline{t_it_{i+1}}$ отображается в вектор

$$d\alpha[\mathbf{u}_i] = \begin{pmatrix} X'(t) \\ Y'(t) \end{pmatrix} dt[\mathbf{u}_i].$$

Если мы имеем дело с обычным евклидовым скалярным произведением, то длина этого вектора равна

$$\|d\alpha[\mathbf{u}_i]\| = \sqrt{X'(t_i)^2 + Y'(t_i)^2}\,dt[\mathbf{u}_i].$$

[6] В русскоязычной литературе такие интегралы называют *криволинейными интегралами первого рода*. — *Прим. ред.*

По определению

$$f(P_i) = f(\alpha(t_i)) = \alpha^* f(t_i),$$

что позволяет[7] написать

$$\int f\, ds = \lim_{N\to\infty} \sum_{i=0}^{N-1} \alpha^* f(t_i)\sqrt{X'(t_i)^2 + Y'(t_i)^2}\,(t_{i+1} - t_i). \qquad (7.4)$$

Это приводит нас к формуле вычисления абсолютного криволинейного интеграла с помощью параметризации:

$$\int_\Gamma f\, ds = \int_a^b \alpha^* f(t)\sqrt{X'(t_i)^2 + Y'(t_i)^2}\, dt.$$

Если $f(P) = 1$, то интеграл $\int ds$ определяет длину кривой Γ. В более общем случае, если $f(P)$ задает линейную плотность массы тонкой проволоки, имеющей форму кривой Γ, то $\int f\, ds$ представляет полную массу проволоки.

Есть интересный с точки зрения применения в физике пример абсолютного линейного интеграла, построенного с помощью неевклидова скалярного произведения. Это вычисление *собственного времени*, связанного с мировой линией движущейся частицы (рис. 7.24), когда прошедшее время измеряется по часам, движущимся вместе с частицей. В этом случае значение времени t вдоль мировой линии всегда растет. Его можно использовать в качестве параметра и написать

$$\alpha^* t = t, \quad \alpha^*(x) = X(t),$$

откуда

$$\alpha^*\, dt = dt, \quad \alpha^*(dx) = X'(t)\, dt.$$

Тогда *лоренцева* длина вектора

$$d\alpha[\mathbf{u}_i] = \begin{pmatrix} 1 \\ X'(t) \end{pmatrix} dt[\mathbf{u}_i]$$

[7]На самом деле длина s_i вектора $\mathbf{v}_i$ не обязана равняться $\|d\alpha[\mathbf{u}_i]\|$, и правая часть следующего равенства (7.5) не совпадает с правой частью (7.4). Правда, при непрерывности $f(t)$ и частных производных $X'(t)$, $Y'(t)$ пределы этих правых частей все же равны, но доказательство этого факта требует кропотливых оценок. — *Прим. ред.*

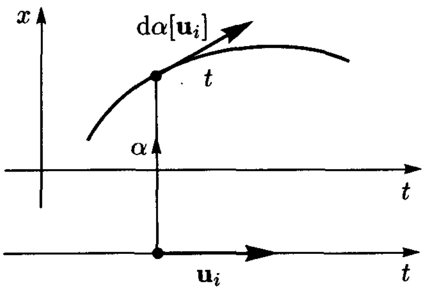


Рис. 7.24

равна $\sqrt{1 - X'(t)^2}\, dt[\mathbf{u}_i]$. Поскольку $X'(t)$ есть скорость частицы v, то получаем интеграл

$$\int_\Gamma \sqrt{1 - v^2}\, dt,$$

определяющий собственное время движущейся частицы.

Резюме

A. Криволинейные интегралы

Вы должны уметь объяснить смысл линейного интеграла $\int_\Gamma \omega$, знать его свойства и уметь пользоваться ими.

Вы должны знать процедуру вычисления линейного интеграла с помощью переноса дифференциальной формы, уметь вводить соответствующую параметризацию для заданных траекторий и вычислять криволинейные интегралы.

B. Дифференциалы и дифференциальные формы

Для заданной дифференциальной формы ω, определенной на плоскости, вы должны уметь выяснять, можно ли представить ее как дифференциал функции f и, если такая функция существует, уметь найти ее.

Задачи

7.1. Пусть $\omega = (y \cos xy + e^x)\, dx + (x \cos xy + 2y)\, dy$.

(a) Вычислите $\int_\Gamma \omega$ вдоль участка параболы $y = x^2$ от точки $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ до $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Используйте параметризацию ϕ, задаваемую переносом

$$\begin{pmatrix} \phi^* x \\ \phi^* y \end{pmatrix} = \begin{pmatrix} t \\ t^2 \end{pmatrix}.$$

(b) Вычислите $\int_\Gamma \omega$, если Γ — прямая линия, соединяющая начало координат с точкой $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. Вычислите тот же интеграл, если путь Γ состоит из отрезка $0 \le x \le \alpha$ вдоль оси x и отрезка $x = \alpha$, $0 \le y \le \beta$.

(c) Найдите такую функцию $f(x, y)$, чтобы выполнялось условие $\omega = df$.

7.2. Пусть снова $\omega = (y \cos xy + e^x)\,dx + (x \cos xy + 2y)\,dy$.

(a) Вычислите $\int_\Gamma \omega$ вдоль участка параболы Γ, определенного формулами

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} t \\ t^2 \end{pmatrix} \quad \text{для} \quad 0 \le t \le 1.$$

(b) Найдите такую функцию $f(x, y)$, чтобы $\omega = df$.

7.3. Пусть $\omega = y\,dx - x\,dy$.

(a) Вычислите $\int_\Gamma \omega$ вдоль полуокружности Γ от точки $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$ до точки $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Путь Γ определен формулами

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -\cos t \\ \sin t \end{pmatrix} \quad \text{для} \quad 0 \le t \le \pi.$$

(b) Покажите явно, что выбрав другую кривую, соединяющую те же точки $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$ и $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, можно получить другое значение интеграла от той же формы.

7.4. Пусть $\omega = (15x^2y^2 - 3y)\,dx + (10x^3y - 3x)\,dy$. Вычислите $\int_\Gamma \omega$, где Γ — путь от $(-1, 0)$ до $(1, 0)$ вдоль полуокружности $x^2 + y^2 = 1$, $y \ge 0$.

7.5. (a) Вычислите $\int_\Gamma \omega$, где $\omega = dx + 2x\,dy$, Γ — участок параболы $x = 1 - y^2$ между $y = -1$ и $y = +1$, как это показано на рис. 7.25.

(b) Найдите такие константу a и функцию f, для которых выполняется равенство $e^{ay}\omega = df$.

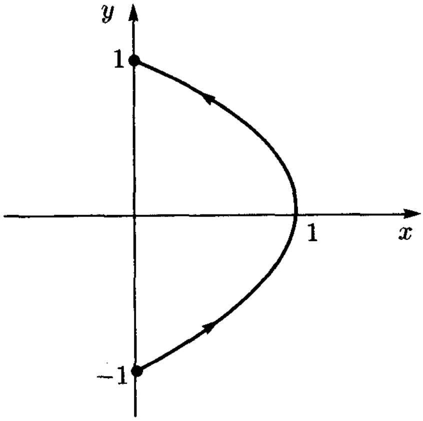


Рис. 7.25

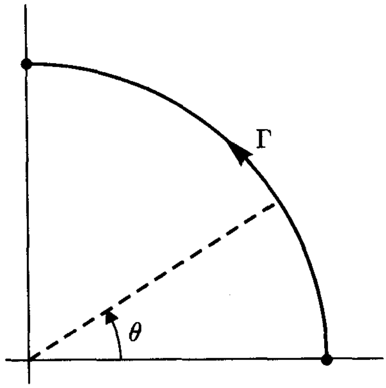


Рис. 7.26

7.6. (a) Пусть $\omega = (x^2 - 2xy)\,dx + (y^2 - 2xy)\,dy$. Вычислите $\int_\Gamma \omega$, где путь Γ проходит вдоль участка параболы $y = x^2$ между точками $(-1, 1)$ и $(1, 1)$.

(b) Пусть $\sigma = 30x^2y^5\,dx + 40x^3y^4\,dy$. Найдите целое число n и функцию f, для которых выполняется равенство $df = (xy)^n\sigma$.

7.7. Пусть $\omega = 10y^2\,dx + 4xy\,dy$.

(a) Вычислите $\int_\Gamma \omega$, где путь Γ проходит по дуге единичного радиуса между точками $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ и $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. (Указание: введите параметр θ — см. рис. 7.26.)

(b) Найдите такое целое число n и функцию f, чтобы выполнялось равенство $x^n\omega = df$.

7.8. (a) Покажите, что дифференциальная форма

$$\omega = 3xy\,dx + 2x\,dy$$

не является дифференциалом никакой функции f.

(b) Найдите уравнение семейства кривых, зависящих от одного параметра и обладающих свойством $\omega(\mathbf{v}) = 0$ для любого вектора $\mathbf{v}$, касательного к какой-либо из этих кривых. (Указание: если $y = F(x)$ и $\mathbf{v} = \begin{pmatrix} 1 \\ F'(x) \end{pmatrix}$, то вы можете написать дифференциальное уравнение для $F(x)$.)

(c) Найдите функции $f(x,y)$ и $g(x,y)$, для которых выполняется равенство $df = g\omega$. (Указание: функция f должна быть постоянной вдоль кривых, полученных в предыдущем упражнении.)

7.9. (a) Нарисуйте полуэллипс, заданный уравнением в полярных координатах

$$r = \frac{9}{5 - 4\cos\theta} \quad \text{для } 0 \le \theta \le \pi.$$

Вспомнив, что $x = r\cos\theta$ и $y = r\sin\theta$, покажите, что этот полуэллипс есть участок кривой

$$\frac{(x-4)^2}{25} + \frac{y^2}{9} = 1.$$

(b) Дифференциальную форму

$$\omega = \frac{xy\,dy - y^2\,dx}{x^2 + y^2}$$

выразите в полярных координатах через r, θ, dr и $d\theta$.

(c) Вычислите $\int_\Gamma \omega$, где путь Γ — полуэллипс из первого пункта этого упражнения. Воспользуйтесь полярными координатами, где θ — удобный параметр.

(d) Вычислите $\int_\Gamma \omega$ в координатах x и y. Удобную параметризацию определяет отображение

$$t \mapsto \begin{pmatrix} 4 + 5\cos t \\ 3\cos t \end{pmatrix}, \quad 0 \le t \le \pi.$$

7.10. (a) Предположим, что u и v — криволинейные координаты в области D на плоскости. Соответствующий якобиан

$$\operatorname{Det}\begin{pmatrix} \partial u/\partial x & \partial u/\partial y \\ \partial v/\partial x & \partial v/\partial y \end{pmatrix}$$

нигде в области D не обращается в нуль. Пусть ω — гладкая дифференциальная форма, определенная в области D, а Γ — кривая в этой области. Покажите, что $\int_\Gamma \omega$ имеет одинаковое значение, если ω и Γ выражать через переменные x, y или через u и v. (Предыдущая задача является примером этого утверждения.)

(b) Пусть путь Γ определен в полярных координатах согласно формулам: $\rho = F(\theta)$, при этом $F(\theta) > 0$ и $F(2\pi) = F(0)$. Покажите, что площадь, заключенная внутри этого замкнутого пути, равна $\int_\Gamma \omega$, где $\omega = \frac{1}{2}\rho^2\, d\theta$. (Указание: Форму ω выразите в декартовых координатах.)

7.11. Состояние газа, заключенного в цилиндр, можно представить точкой на плоскости, где декартовыми координатами являются P (давление) и V (объем). Количество тепла, поглощенного газом в некотором процессе, заданном траекторией Γ, равно $Q = \int_\Gamma \omega$, где

$$\omega = \frac{5}{2} P\, dV + \frac{3}{2} V\, dP.$$

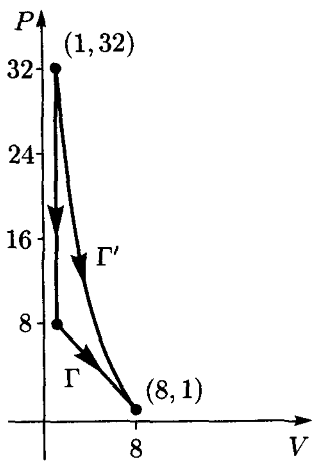


Рис. 7.27

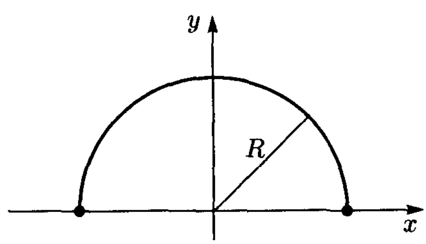


Рис. 7.28

(a) Вычислите $\int_\Gamma \omega$, где путь Γ проходит вдоль ломаной линии, изображенной на рис. 7.27. Линия последовательно соединяет точки $V = 1$, $P = 32$, затем $V = 1$, $P = 8$ и $V = 8$, $P = 1$.

(b) Вычислите интеграл $\int_{\Gamma'} \omega$ вдоль пути Γ', проходящем вдоль кривой $PV^{5/3} = 32$ от точки $V = 1,\ P = 32$ к точке $V = 8$, $P = 1$.

(c) Найдите такую функцию S, чтобы выполнялось равенство $dS = \omega/PV$.

7.12. (a) Имеется однородная проволока с массой M, согнутая в полуокружность с радиусом R, как это показано на рис. 7.28. Найдите y-ю координату центра масс этой проволоки и вычислите ее момент инерции относительно оси x.

(b) Решите ту же задачу в случае, когда линейная плотность проволоки пропорциональна y, но ее полная масса по-прежнему равна M.

7.13. Запишите в форме интеграла длину кривой $y = (x - 2)^3$, соединяющей точки $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ и $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ на евклидовой плоскости.

7.14. Рассмотрите дифференциальную форму

$$\omega = 5y\,dx + 3x\,dy.$$

(a) Найдите такие целые числа m и n, для которых выполняется равенство $d(x^m y^n \omega) = 0$.

(b) Для этих целых чисел найдите такую функцию f, чтобы выполнялось равенство $df = x^m y^n \omega$.

(c) Отобразим плоскость uv на плоскость xy так, чтобы

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \mapsto \begin{pmatrix} \alpha^2 + \beta \\ \beta^2 + 2\alpha \end{pmatrix}.$$

Перенесите форму ω на плоскость uv.

(d) Вычислите $\int \omega$ вдоль путей Γ_1 и Γ_2, соединяющих точки $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ и $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Путь Γ_1 состоит из двух прямых отрезков и проходит через точку $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, а путь Γ_2 проходит по двум прямым отрезкам через точку $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

(e) Вычислите абсолютный криволинейный интеграл $\int_{\Gamma_1} \|\omega\|\, ds$. (В ответе один член можно оставить в виде обычного определенного интеграла.)

Глава 8

Двойные интегралы

Глава 8 продолжает изучение интегрального исчисления. Она посвящена внешним 2-формам и двумерным интегралам, которые им соответствуют. Вводится определение внешней производной и отмечается ее инвариантность относительно переноса. Доказывается двумерный вариант теоремы Стокса, известный как теорема Грина. Исследуются поверхностные интегралы в трехмерном пространстве.

8.1. Внешняя производная

Мы уже знаем, что дифференциал функции f дает наилучшее линейное приближение к приращению значения функции при перемещении ее аргумента от точки $\mathbf{P}$ к точке $\mathbf{P} + \mathbf{v}$. Это можно записать в форме

$$df(\mathbf{P})[\mathbf{v}] = f(\mathbf{P} + \mathbf{v}) - f(\mathbf{P}) + o(\mathbf{v}),$$

где при $\mathbf{v} \to \mathbf{0}$ поправочный член $o(\mathbf{v})$ стремится к нулю быстрее, чем длина вектора $\mathbf{v}$. Можно представлять себе, что df является линейной функцией, значение которой на конце вектора $\mathbf{v}$ определяется значениями самой функции f на концах отрезка, заданного вектором $\mathbf{v}$ «в пределе», когда $\mathbf{v}$ становится очень малым.

Воспользовавшись аналогичным подходом, можно из 1-формы τ построить *2-форму* $d\tau$, которая будет называться *внешней производной* от τ. Эта 2-форма в каждой точке $\mathbf{P}$ будет *билинейной* функцией двух векторов $\mathbf{v}$ и $\mathbf{w}$. Именно, выбрав точку $\mathbf{P}$, *упорядоченную* пару векторов $\mathbf{v}$ и $\mathbf{w}$ и 1-форму, мы можем получить число, проинтегрировав τ по параллелограмму, построенному на векторах $\mathbf{v}$ и $\mathbf{w}$, двигаясь сначала «вперед» от $\mathbf{P}$ вдоль первого вектора $\mathbf{v}$, а в конце пути «назад» вдоль второго вектора $\mathbf{w}$ (рис. 8.1).

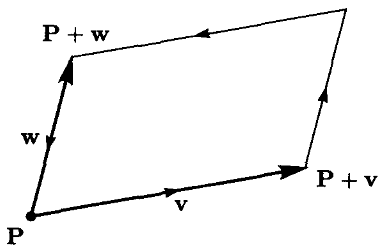


Рис. 8.1

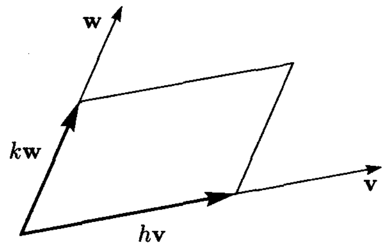


Рис. 8.2

Если векторы $\mathbf{v}$ и $\mathbf{w}$ достаточно малы и поведение формы τ в окрестности точки $\mathbf{P}$ достаточно хорошее, то можно ожидать, что значение этого интеграла будет приблизительно *билинейно*, т. е. оно будет зависеть приблизительно *линейно* от $\mathbf{v}$ (при фиксированном векторе $\mathbf{w}$) и линейно от $\mathbf{w}$ (при фиксированном $\mathbf{v}$). Обозначим $\mathbf{P}(h,k)$ параллелограмм, заданный векторами $h\mathbf{v}$ и $k\mathbf{w}$ (рис. 8.2). Хотелось бы определить $d\tau(\mathbf{P})$ формулой

$$\int_{\mathbf{P}(h,k)} \tau = hk\, d\tau(\mathbf{P})[\mathbf{v}, \mathbf{w}] + \text{остаток}, \tag{8.1}$$

где (мы надеемся) остаточный член стремится к нулю быстрее, чем h при $h \to 0$ (значение k фиксировано) и быстрее, чем k при $k \to 0$ (значение h фиксировано)[1].

Если 2-форма $d\tau(\mathbf{P})$ существует, то она единственна. Доказательство практически совпадает с доказательством утверждения, что дифференциал функции единственен. Предположим, что

[1]Обычно требуется более сильное условие: дробь остаток$/hk$ должна стремиться к нулю при одновременном стремлении к нулю обоих чисел h, k (см. раздел 8.8). — *Прим. ред.*

уравнение (8.1) справедливо для двух разных билинейных функций $d\tau$ и $\overline{d\tau}$. Пусть σ обозначает их разность. Тогда мы имеем

$$0 = hk\sigma[\mathbf{v}, \mathbf{w}] + \text{остаток}.$$

Разделим это равенство на hk и устремим h к нулю. Тогда получаем

$$0 = \sigma[\mathbf{v}, \mathbf{w}] + \frac{1}{k}\lim_{h\to 0}(\text{остаток}/h).$$

Но остаток стремится к нулю быстрее h, следовательно, $\sigma[\mathbf{v}, \mathbf{w}] = 0$. Это значит, что $\overline{d\tau}$ не может отличаться от $d\tau$.

А теперь займемся проблемой вычисления $d\tau(\mathbf{P})$ и доказательством существования 2-формы. Для простоты сначала предположим, что τ имеет вид $f\,dx$, где f — дважды дифференцируемая функция в окрестности точки $\mathbf{P}$. В этом выражении dx является формой, которая каждому касательному вектору приписывает его x-ю компоненту. В частности, $dx[h\mathbf{v}] = h\,dx[\mathbf{v}] = hv_x$, если $\mathbf{v} = \begin{pmatrix} v_x \\ v_y \end{pmatrix}$. Вклад в интеграл, получаемый при движении (рис. 8.3) вдоль одной стороны параллелограмма, скажем, от точки $\mathbf{P}$ к точке $\mathbf{P} + h\mathbf{v}$, получаемый с помощью параметризации $t \to \mathbf{P} + th\mathbf{v}$, равен

$$\begin{aligned}\int_0^1 f(\mathbf{P} + th\mathbf{v})x'(t)\,dt &= \int_0^1 f(\mathbf{P} + th\mathbf{v})\,dx[h\mathbf{v}]\,dt \\ &= h\,dx[\mathbf{v}]\int_0^1 f(\mathbf{P} + th\mathbf{v})\,dt.\end{aligned}$$

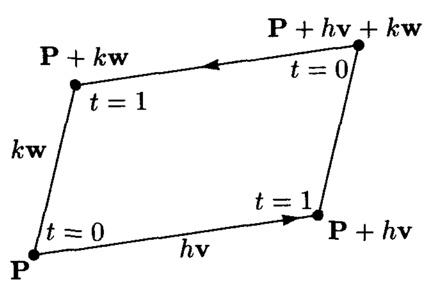


Рис. 8.3

Вклад противоположной стороны от точки $\mathbf{P} + h\mathbf{v} + k\mathbf{w}$ к точке $\mathbf{P} + k\mathbf{w}$ получается аналогично. Он равен

$$-h\,dx[\mathbf{v}] \int_0^1 f(\mathbf{P} + th\mathbf{v} + k\mathbf{w})\,dt.$$

Складывая оба члена, мы получаем

$$-h\,dx[\mathbf{v}] \int_0^1 [f(\mathbf{P} + th\mathbf{v} + k\mathbf{w}) - f(\mathbf{P} + th\mathbf{v})]\,dt.$$

Поскольку мы предположили, что f — дважды дифференцируемая функция, то можно воспользоваться формулой Тейлора

$$f(\mathbf{P} + th\mathbf{v} + k\mathbf{w}) - f(\mathbf{P} + k\mathbf{w}) = df_{(\mathbf{P}+th\mathbf{v})}[k\mathbf{w}] + O(k^2)$$

и записать суммарный вклад от двух сторон в виде

$$-h\,dx[\mathbf{v}] \int_0^1 df_{(\mathbf{P}+th\mathbf{v})}[k\mathbf{w}]\,dt + O(hk^2).$$

Вычисление интеграла вдоль других двух сторон дает выражение

$$+k\,dx[\mathbf{w}] \int_0^1 df_{(\mathbf{P}+tk\mathbf{w})}[h\mathbf{v}]\,dt + O(kh^2).$$

Собирая вместе все члены, получаем

$$\begin{aligned}\int_{\mathbf{P}(h,k)} \tau = &-hk \int_0^1 df_{(\mathbf{P}+th\mathbf{v})}[\mathbf{w}]\,dt \cdot dx[\mathbf{v}] \\ &+ hk \int_0^1 df_{(\mathbf{P}+tk\mathbf{w})}[\mathbf{v}]\,dt \cdot dx[\mathbf{w}] + O(kh^2) + O(hk^2).\end{aligned}$$

Здесь $df_{(\mathbf{P}+th\mathbf{v})}$ является вектором-строкой

$$\left(\frac{\partial f}{\partial x}(\mathbf{P} + th\mathbf{v}),\ \frac{\partial f}{\partial y}(\mathbf{P} + th\mathbf{v})\right).$$

Согласно нашему предположению о функции f, ее частные производные дифференцируемы. Следовательно, по теореме о среднем значении

$$\frac{\partial f}{\partial x}(\mathbf{P} + th\mathbf{v}) = \frac{\partial f}{\partial x}(\mathbf{P}) + O(h),$$

откуда

$$df_{(\mathbf{P}+th\mathbf{v})}[\mathbf{w}] = df_{\mathbf{P}}[\mathbf{w}] + O(h).$$

Интегрирование этого выражения дает

$$\int_0^1 df_{(\mathbf{P}+th\mathbf{v})}[\mathbf{w}]\,dt = df_{\mathbf{P}}[\mathbf{w}] + O(h).$$

Подставим это в наш интеграл по контуру параллелограмма и получим

$$\int_{\mathbf{P}(h,k)} \tau = hk(df_{\mathbf{P}}[\mathbf{v}]\,dx[\mathbf{w}] - df_{\mathbf{P}}[\mathbf{w}]\,dx[\mathbf{v}]) + O(h^2k) + O(hk^2).$$

Таким образом, мы можем получить выражение (8.1), если положим

$$d\tau_{\mathbf{P}}[\mathbf{v},\mathbf{w}] = df_{\mathbf{P}}[\mathbf{v}]\,dx[\mathbf{w}] - df_{\mathbf{P}}[\mathbf{w}]\,dx[\mathbf{v}].$$

Из этого выражения видно, что $d\tau$ является антисимметричной функцией своих аргументов, т. е. $d\tau[\mathbf{w},\mathbf{v}] = -d\tau[\mathbf{v},\mathbf{w}]$.

2-форма $d\tau$ может быть записана короче, если ввести понятие внешнего произведения двух 1-форм. По определению, это внешнее произведение равно

$$(\sigma \wedge \lambda)[\mathbf{v},\mathbf{w}] = \sigma[\mathbf{v}]\lambda[\mathbf{w}] - \sigma[\mathbf{w}]\lambda[\mathbf{v}],$$

где σ и λ являются 1-формами, а $\mathbf{v}$ и $\mathbf{w}$ — векторами. Тогда можно написать

$$d\tau_{\mathbf{P}}[\mathbf{v},\mathbf{w}] = (df_{\mathbf{P}} \wedge dx)[\mathbf{v},\mathbf{w}],$$

или еще короче

$$d\tau = df \wedge dx.$$

Из определения внешнего произведения ясно, что

$$\lambda \wedge \sigma = -\sigma \wedge \lambda,$$

т. е. оно антисимметрично. В частности, $\sigma \wedge \sigma = -\sigma \wedge \sigma = 0$, т. е. произведение 1-формы на себя равно нулю.

Кроме того, очевидно, что

$$(\sigma + \omega) \wedge \lambda = (\sigma \wedge \lambda) + (\omega \wedge \lambda),$$

т. е. внешнее произведение дистрибутивно по отношению к операции сложения.

Теперь рассмотрим 1-форму общего вида: $f\,dx + g\,dy$. Те же самые соображения, примененные к $g\,dy$, дают

$$d(g\,dy) = dg \wedge dy.$$

Поскольку интеграл от ω линеен по ω, мы получаем

$$d(f\,dx + g\,dy) = df \wedge dx + dg \wedge dy.$$

Это выражение можно доказать непосредственно, исходя из определения d. Теперь выразим df и dg через dx и dy:

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy; \qquad dg = \frac{\partial g}{\partial x}\,dx + \frac{\partial g}{\partial y}\,dy.$$

Поскольку $dx \wedge dx = 0$ и $dy \wedge dy = 0$, мы имеем

$$d\tau = \frac{\partial f}{\partial y}\,dy \wedge dx + \frac{\partial g}{\partial x}\,dx \wedge dy.$$

И наконец, поскольку $dy \wedge dx = -dx \wedge dy$, получаем

$$\boxed{d\tau = \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dx \wedge dy.}$$

В качестве примера возьмем 1-форму

$$\tau = x^2y^2\,dx + x^3y\,dy.$$

Тогда

$$\begin{aligned} d\tau &= d(x^2y^2) \wedge dx + d(x^3y) \wedge dy \\ &= (2xy^2\,dx + 2x^2y\,dy) \wedge dx + (3x^2y\,dx + x^3\,dy) \wedge dy \\ &= 2x^2y\,dy \wedge dx + 3x^2y\,dx \wedge dy = x^2y\,dx \wedge dy. \end{aligned}$$

В частном случае, когда 1-форма τ точная, т. е. $\tau = d\phi$, мы получаем, что $d\tau = 0$. Это видно из определения $d\tau$: поскольку $d\tau[\mathbf{v}, \mathbf{w}]$ является наилучшим линейным приближением к интегралу по контуру параллелограмма $\int \tau$, и поскольку интеграл от полного дифференциала по замкнутому контуру равен нулю, то $d(d\phi) = 0$. Этот результат можно установить и прямым вычислением:

$$\tau = d\phi = \frac{\partial \phi}{\partial x}\,dx + \frac{\partial \phi}{\partial y}\,dy,$$
$$d\tau = \frac{\partial^2 \phi}{\partial y \partial x}\,dy \wedge dx + \frac{\partial^2 \phi}{\partial x \partial y}\,dx \wedge dy = 0,$$

потому что смешанные частные производные второго порядка равны. Заметим, что попутно мы получили короткий вариант определения замкнутой формы: форма τ является замкнутой тогда и только тогда, когда

$$d\tau = 0.$$

Итак, мы показали, что если f дифференцируема и x — координатная функция, то $d(f\,dx) = df \wedge dx$. Используем этот результат для доказательства более общей формулы

$$d(f\tau) = df \wedge \tau + f\,d\tau,$$

где f — дифференцируемая функция, а τ — дифференцируемая 1-форма. Записав $\tau = g\,dx + h\,dy$, мы получим $f\tau = (fg)\,dx + (fh)\,dy$ и, следовательно,

$$d(f\tau) = d(fg) \wedge dx + d(fh) \wedge dy.$$

Поэтому

$$d(f\tau) = g\,df \wedge dx + f\,dg \wedge dx + h\,df \wedge dy + f\,dh \wedge dy$$

и мы видим, что

$$d(f\tau) = df \wedge (g\,dx + h\,dy) + f(dg \wedge dx + dh \wedge dy),$$

т. е.

$$\boxed{d(f\tau) = df \wedge \tau + f\,d\tau.}$$

8.2. 2-формы

Поскольку общий вид 1-формы на плоскости — это выражение $f\,dx + g\,dy$, то самый общий вид произведения двух 1-форм будет некоторой функцией $F(x,y)$, умноженной на $dx \wedge dy$. Назовем такое выражение *2-формой* и будем писать

$$\sigma = F(P)\,dx \wedge dy = F(x,y)\,dx \wedge dy.$$

Значение 2-формы в точке P, т.е. $F(P)\,dx \wedge dy$, мы будем рассматривать как правило, которое каждой паре векторов приписывает определенное число.

Чтобы лучше понять смысл «постоянной» 2-формы $dx \wedge dy$, вычислим ее сначала для пары единичных векторов $(\mathbf{e}_x, \mathbf{e}_y)$. По определению

$$dx \wedge dy[\mathbf{e}_x, \mathbf{e}_y] = dx[\mathbf{e}_x]\,dy[\mathbf{e}_y] - dx[\mathbf{e}_y]\,dy[\mathbf{e}_x] = 1 \cdot 1 - 0 \cdot 0 = 1.$$

Более общая формула:

$$dx \wedge dy[h\mathbf{e}_x, k\mathbf{e}_y] = hk.$$

Очевидно, что это — площадь (с некоторым знаком) прямоугольника, построенного на векторах $h\mathbf{e}_x$ и $k\mathbf{e}_y$, если площадь прямоугольника, построенного на векторах $\mathbf{e}_x$ и $\mathbf{e}_y$, принимается за единицу (рис. 8.4).

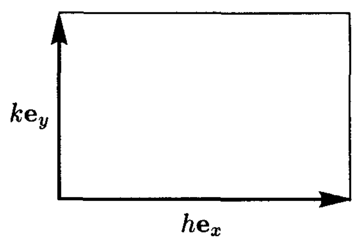


Рис. 8.4

Мы можем вычислить $dx \wedge dy$ для любой упорядоченной пары векторов $(\mathbf{v}, \mathbf{w})$. Запишем $\mathbf{v} = a\mathbf{e}_x + c\mathbf{e}_y$, $\mathbf{w} = b\mathbf{e}_x + d\mathbf{e}_y$. Если ввести матрицу A

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \mathbf{v} = A\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{w} = A\begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

то

$$dx \wedge dy[\mathbf{v}, \mathbf{w}] = dx[\mathbf{v}]\, dy[\mathbf{w}] - dx[\mathbf{w}]\, dy[\mathbf{v}]$$
$$= ad - bc = \operatorname{Det} A.$$

В общем случае 2-форма σ является функцией трех переменных: P, $\mathbf{v}$ и $\mathbf{w}$. Для фиксированной точки P она билинейна и антисимметрична относительно векторов $\mathbf{v}$ и $\mathbf{w}$. Если $\sigma = F\, dx \wedge dy$, то

$$\sigma(P)(\mathbf{v}, \mathbf{w}) = F(P) \operatorname{Det} A,$$

где $\mathbf{v}$, $\mathbf{w}$ и A определены выше. Итак, каждому параллелограмму с вершиной в точке P 2-форма $\sigma(P)$ приписывает площадь с определенным знаком — ориентированную площадь.

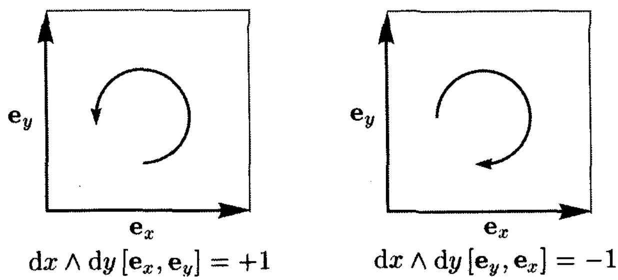


Рис. 8.5

Очень важно помнить, что величина $dx \wedge dy$ для данного параллелограмма зависит от его ориентации, которая определяется порядком векторов, задающих его. Для ориентированного прямоугольника, соответствующего векторам $[\mathbf{e}_x, \mathbf{e}_y]$, $dx \wedge dy = 1$, а $dy \wedge dx = -1$. Для того же прямоугольника, но с противоположной ориентацией, соответствующей порядку $[\mathbf{e}_y, \mathbf{e}_x]$, $dx \wedge dy = -1$ и $dy \wedge dx = +1$ (см. рис. 8.5). В общем случае, чтобы получить 2-форму τ для параллелограмма, построенного на векторах $\mathbf{v}$ и $\mathbf{w}$, надо задать его ориентацию, т. е. выбрать, какой вектор будет «первым», а после этого вычислять либо $\tau[\mathbf{v}, \mathbf{w}]$, либо $\tau[\mathbf{w}, \mathbf{v}]$.

8.3. Интегрирование 2-форм

Итак, 2-форма τ приписывает число каждому малому ориентированному параллелограмму (паре векторов), точно так же, как 1-форма приписывает число каждому малому направленному отрезку линии (вектору). Поэтому мы можем интегрировать 2-формы по области R подобно тому, как мы интегрировали 1-формы по путям.

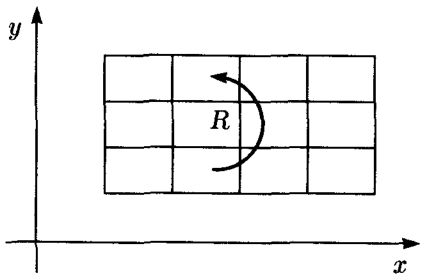


Рис. 8.6

Приступим к построению такого интеграла. Возьмем прямоугольник R, ориентированный так, как показано на рис. 8.6, разбиваем его на $N_x N_y$ маленьких прямоугольников и составляем римановы суммы

$$S_{N_x N_y} = \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} \sigma(P_{ij}) \left[\frac{b-a}{N_x}\mathbf{e}_x, \frac{d-c}{N_y}\mathbf{e}_y\right].$$

Если

$$\sigma = F(x,y)\, dx \wedge dy,$$

то

$$S_{N_x N_y} = \frac{b-a}{N_x} \cdot \frac{d-c}{N_y} \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} F(x_i, y_j),$$

где

$$x_i = a + \frac{i}{N_x}(b-a), \qquad y_j = c + \frac{j}{N_y}(d-c).$$

Теперь определим интеграл от 2-формы τ по ориентированной области R согласно формуле

$$\int_R \tau = \lim_{\substack{N_x\to\infty \\ N_y\to\infty}} S_{N_xN_y},$$

при условии, что предел не зависит от характера измельчения разбиения[2].

Двойной интеграл от $F(x,y)\,dx\wedge dy$ по области R можно вычислять как «повторный» интеграл. Чтобы вычислить выражение

$$I = \lim_{\substack{N_x\to\infty \\ N_y\to\infty}} \frac{b-a}{N_x}\cdot\frac{d-c}{N_y}\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} F(x_i,y_j),$$

сначала просуммируем по j для каждого фиксированного значения i и прежде, чем суммировать по i, устремим $N_y\to\infty$. По определению обычного риманова интеграла от функции одной переменной

$$\lim_{N_y\to\infty}\frac{d-c}{N_y}\sum_{j=0}^{N_y-1} F(x_i,y_j) = \int_c^d F(x_i,y)\,dy.$$

Следовательно,

$$I = \lim_{N_x\to\infty}\frac{b-a}{N_x}\sum_{i=0}^{N_x-1}\int_c^d F(x_i,y)\,dy.$$

Рассматривая предел этой римановой суммы как интеграл, мы можем написать

$$I = \int_a^b\left(\int_c^d F(x,y)\,dy\right)dx,$$

[2]Авторы не останавливаются на вопросе о том, для каких функций $F(x,y)$ написанный предел существует. На самом деле для этого достаточно, чтобы $F(x,y)$ была непрерывной на прямоугольнике R. То же самое относится к рассматриваемому ниже вопросу о сведении двойного интеграла к повторному. — *Прим. ред.*

т. е. наш двойной интеграл I мы выразим через «повторный» интеграл, который можно вычислять методами интегрального исчисления функций одной переменной. Точно так же мы можем сначала просуммировать по i, а потом по j. В результате получим

$$I = \int_c^d \left(\int_a^b F(x,y)\,dx \right) dy.$$

При вычислении интеграла от 2-формы τ по ориентированной области R мы должны думать об ориентации прямоугольника. Если R ориентирован так, что x будет «первой» координатой, а y — «второй», как это показано на рис. 8.7(a), мы должны писать $\tau = F(x,y)\,dx \wedge dy$ и вычислять повторный интеграл

$$\int_a^b \left(\int_c^d F(x,y)\,dy \right) dx \quad \text{или} \quad \int_c^d \left(\int_a^b F(x,y)\,dx \right) dy.$$

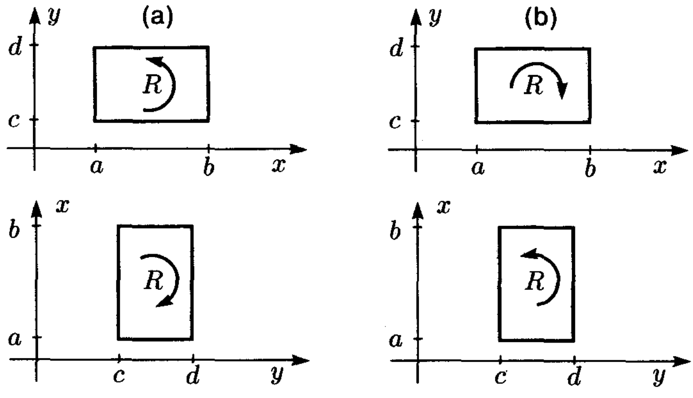


Рис. 8.7. (a) Координата x — первая. (b) Координата y — первая.

Если R ориентирован так, что первой координатой будет y, как это показано на рис. 8.7(b), мы можем написать

$$\tau = G(y,x)\,dy \wedge dx \qquad (\text{где } G(y,x) = -F(x,y))$$

и вычислять повторный интеграл

$$\int_a^b \left(\int_c^d G(y,x)\,dy \right) dx \quad \text{или} \quad \int_c^d \left(\int_a^b G(y,x)\,dx \right) dy.$$

Изменение ориентации прямоугольника R приводит к изменению *знака* интеграла.

В случае криволинейных интегралов понятие ориентации (и почему изменяется знак при изменении ориентации) интуитивно ясно. (Например, для случая силового поля криволинейный интеграл дает работу, выполняемую при движении вдоль пути. Прибор регистрирует разность энергий $E_B - E_A$. При изменении направления движения прибор будет показывать изменение энергии $E_A - E_B$.) Важно иметь аналогичный интуитивный пример для наших двумерных интегралов. Вот такой пример. Ориентация плоскости изменяется в зависимости от того, смотрим мы на нее сверху или снизу. Предположим, что в трехмерном пространстве есть плоскость xy при $z = 0$. Мы смотрим на эту плоскость сверху и видим вращение по часовой стрелке. Но если смотреть на эту же плоскость снизу, то то же самое вращение мы будем видеть как вращение против часовой стрелки (рис. 8.8). Можно сказать, что выбор ориентации плоскости в трехмерном пространстве тесно связан с выбором «стороны» этой плоскости. Допустим, что через поверхность течет какое-то вещество. Например, поверхность — это клеточная мембрана, и нам интересно знать, какое количество ионов проходят через нее.

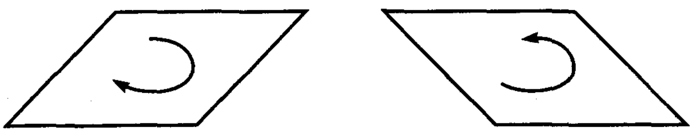


Рис. 8.8

Конечно, используя слово «через», мы должны сделать выбор: переход в каком направлении мы считаем со знаком «+», а в каком — со знаком «−». Итак, для измерения полного потока через поверхность мы должны задать ее ориентацию. Изменение ориентации приводит к изменению знака потока и, соответственно, знака нашего интеграла.

Абсолютные двойные интегралы. Часто необходимо вычислять *абсолютный* двойной интеграл, который вычисляется по области, не имеющей ориентации[3]. Если, например, σ обозначает плотность плоской пластины (массу на единицу площади), имеющей форму прямоугольника R, то масса всей пластины определяется двойным интегралом

$$M = \int_R \sigma \, dA.$$

Очевидно, что M должна быть положительным числом, и ориентация прямоугольника R не имеет значения (рис. 8.9). В этом интеграле dA можно считать функцией, которая любому малому параллелограмму приписывает его обычную геометрическую площадь, т. е. абсолютное значение его ориентированной площади. В качестве координат выберем x и y. Тогда $dA = dx\,dy$ или $dA = dy\,dx$, порядок координат не важен.

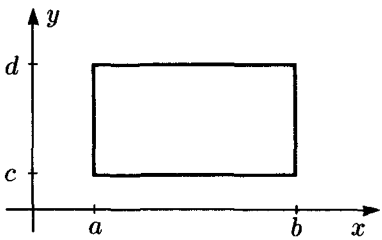


Рис. 8.9

Абсолютный интеграл $\int_R F(x,y)\,dx\,dy$ вычисляется как повторный интеграл

$$\int_a^b \left(\int_c^d F(x,y)\,dy \right) dx,$$

или как

$$\int_c^d \left(\int_a^b F(x,y)\,dx \right) dy.$$

Важно отметить, что существуют два разных типа геометрических объектов: выражения типа $\sigma\,dA$, которые можно называть

[3] В русскоязычной литературе такие интегралы называют интегралами *первого рода*. Интегралы, рассматривавшиеся выше, в русскоязычной литературе называют интегралами *второго рода*. — *Прим. ред.*

плотностями, при этом интегрирование по области R не зависит от ее ориентации; и 2-формы $\tau = F\,dx \wedge dy$, интеграл от которых зависит от ориентации. У них совершенно различный физический смысл. Позднее мы увидим, что они по-разному ведут себя при замене переменных или переносе.

Двойные интегралы по произвольной области как повторные интегралы. Двойные интегралы можно вычислять как повторные для любой области, ограниченной прямыми $x = \mathrm{const}$ и графиками функций, которые не пересекаются. Например, на рис. 8.10 изображена область, ограниченная слева прямой $x = a$, справа прямой $x = b$, сверху линией $y = \phi(x)$ и снизу линией $y = \psi(x)$.

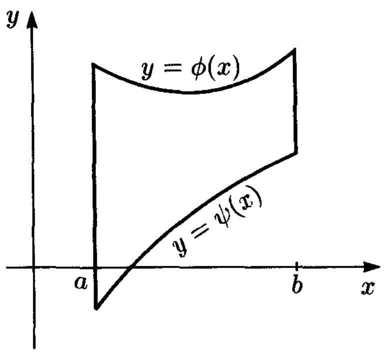


Рис. 8.10

В этом случае справедлива формула[4]:

$$\int_R F(x,y)\,dx\,dy = \int_a^b \left(\int_{\psi(x)}^{\phi(x)} F(x,y) \right) dy\,dx.$$

В качестве примера вычислим интеграл $I = \int_R 2xy\,dx\,dy$ по четверти круга, ограниченной прямыми $x = 0$, $x = a$, $y = 0$ и дугой

[4]Отсутствующее у авторов определение левой части этой формулы можно дать так: дополним область R до координатного прямоугольника $\hat{R}$ и доопределим функцию $F(x,y)$ нулем вне R, получим функцию $\hat{F}$ на $\hat{R}$. Тогда можно свести дело к предыдущему понятию, определив $\int_R F\,dx\,dy$ как $\int_{\hat{R}} \hat{F}\,dx\,dy$. Для следующего ниже равенства достаточны непрерывность и ограниченность на R функции F и непрерывность $\psi(x)$ и $\phi(x)$. — *Прим. ред.*

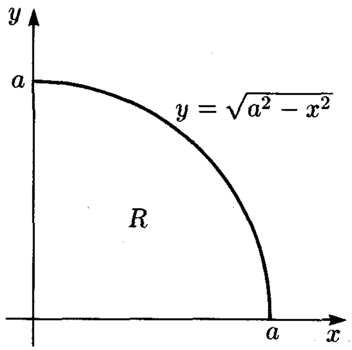


Рис. 8.11

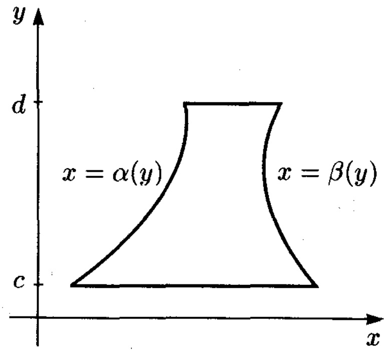


Рис. 8.12

$y = \sqrt{a^2 - x^2}$ (рис. 8.11). Интегрируя сначала по y, а потом по x, получаем

$$I = \int_0^a \left(\int_0^{\sqrt{(a^2 - x^2)}} 2xy\, dy \right) dx$$
$$= \int_0^a [xy^2]_0^{\sqrt{(a^2-x^2)}}\, dx = \int_0^a x(a^2 - x^2)\, dx = \frac{1}{4}a^3.$$

Иногда двойной интеграл легче вычислять, если изменить порядок интегрирования и сначала выполнить интегрирование по x. Например, интеграл $I = \int_S F(x,y)\, dx\, dy$ по области S, изображенной на рис. 8.12 (который совсем не просто получить, если сначала проинтегрировать по y), может быть вычислен так:.

$$I = \int_c^d \left(\int_{\alpha(y)}^{\beta(y)} F(x,y)\, dx \right) dy.$$

Даже для областей, не имеющих форму прямоугольника, двойной интеграл часто можно вычислять как повторный интеграл при обоих порядках интегрирования. Предположим, например, что нам надо вычислить $I = \int_R y\, dx\, dy$ по области R, заключенной между параболой $y = x^2$ и прямой $y = x$ (рис. 8.13).

Сначала проинтегрируем по y, а потом по x:

$$I = \int_0^1 \left(\int_{x^2}^x y\, dy \right) dx = \int_0^1 \left[\frac{1}{2} y^2 \right]_{x^2}^x dx = \frac{1}{2} \int_0^1 (x^2 - x^4)\, dx = \frac{1}{15}.$$

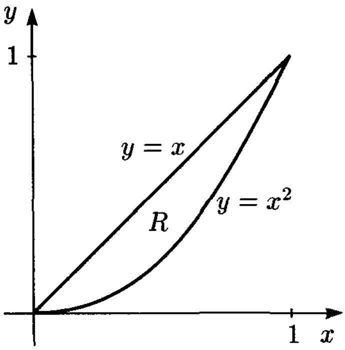


Рис. 8.13

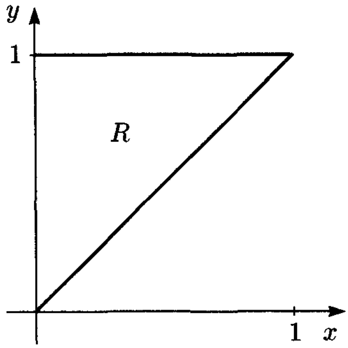


Рис. 8.14

Эту же область можно заключить между прямой $x = y$ и параболой $x = \sqrt{y}$, тогда сначала интегрируем по x, затем по y:

$$I = \int_0^1 \left(\int_y^{\sqrt{y}} dx \right) y\, dy = \int_0^1 (\sqrt{y} - y) y\, dy = \int_0^1 (y^{3/2} - y^2)\, dy = \frac{1}{15}.$$

Иногда, чтобы изменить порядок интегрирования, имеет смысл рассматривать повторный интеграл как двойной. Например, интеграл

$$I = \int_0^1 \left(\int_x^1 e^{-y^2}\, dy \right) dx$$

довольно сложно вычислять в таком виде. Однако, его можно записать как двойной интеграл

$$I = \int_R e^{-y^2}\, dx\, dy,$$

где область R заключена между прямыми $x = 0$, $y = 1$ и $y = x$ (рис. 8.14).

В этом двойном интеграле проще сначала проинтегрировать по x, а потом по y:

$$I = \int_0^1 e^{-y^2} \left(\int_0^y dx \right) dy.$$

Тогда

$$I = \int_0^1 y e^{-y^2}\, dy = \int_0^1 \frac{1}{2} e^{-u}\, du = \frac{1}{2}(1 - e^{-1}).$$

Заметим, что в этом примере исходный интеграл можно вычислить по-другому. Введем первообразную функции e^{-y^2}

$$G(y) = \int_0^y e^{-t^2}\,dt,$$

так что

$$G'(y) = e^{-y^2}.$$

Тогда

$$I = \int_0^1 dx \int_x^1 e^{-y^2}\,dy = \int_0^1 (G(1) - G(x))\,dx.$$

Интегрирование по частям дает

$$I = [(G(1) - G(x))x]_0^1 - \int_0^1 x(-G'(x))\,dx.$$

Первый член в правой части равенства обращается в нуль на обоих пределах. Поскольку $G'(x) = e^{-x^2}$, мы получаем

$$I = \int_0^1 xe^{-x^2} dx = \frac{1}{2}(1 - e^{-1}),$$

что совпадает с результатом, полученным ранее.

Иногда для вычисления интеграла по x и y необходимо разделить область интегрирования на части.

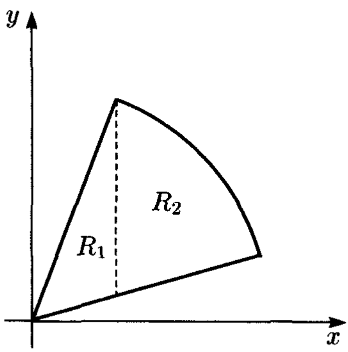


Рис. 8.15

Например, чтобы вычислить $\int F(x,y)\,dx\,dy$ по сектору R, изображенному на рис. 8.15, разделим его на области R_1 и R_2. Тогда можно написать

$$\int_R F(x,y)\,dx\,dy = \int_{R_1} F(x,y)\,dx\,dy + \int_{R_2} F(x,y)\,dx\,dy$$

и каждый интеграл свести к повторному. Существует более естественный способ вычисления этого интеграла — ввести полярные координаты. Важную проблему замены переменных мы обсудим в разделе 8.5.

8.4. Ориентация

Мы уже знаем, что знак криволинейного интеграла зависит от ориентации пути, а знак интеграла от 2-формы зависит от *ориентации плоскости*. Надеемся, что у читателя появилось интуитивное ощущение того, что означает ориентация. В то же время мы понимаем, что необходимо дать точное математическое определение, что и будет сделано в этом разделе. Прежде чем углубиться в абстрактные математические определения, давайте рассмотрим пример. Интуитивно ясно, что на плоскости имеются две возможные ориентации (рис. 8.16).

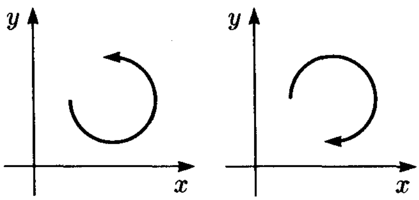


Рис. 8.16

Мы не можем дать внутренние характеристики этих ориентаций, но точно знаем, что они *разные* и что их только две. То же самое можно сказать о прямой (рис. 8.17) или о трехмерном пространстве, где существуют правая и левая системы координат (рис. 8.18).

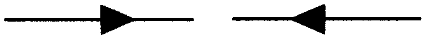


Рис. 8.17

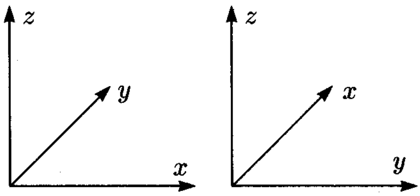


Рис. 8.18

Сначала рассмотрим плоскость, где нам известно (раздел 1.9), что невырожденная матрица A сохраняет ориентацию или изменяет ее на противоположную в зависимости от знака $\operatorname{Det} A$. Это дает нам ключ для общего определения[5].

Пусть V — абстрактное двумерное векторное пространство. Как мы уже видели в конце главы 1, задание базиса в пространстве V равносильно заданию изоморфизма $L : V \to \mathbb{R}^2$. Если L и L' — два таких базиса, то

$$L' = BL,$$

где B — невырожденная матрица 2×2 (рис. 8.19).

$$\begin{tikzcd} & V \arrow[dl, "L"'] \arrow[dr, "L'"] & \\ \mathbb{R}^2 \arrow[rr, "B"] & & \mathbb{R}^2 \end{tikzcd}$$

Рис. 8.19

Назовем базисы L и L' *подобными*, если $\operatorname{Det} B > 0$, и *противоположными*, если $\operatorname{Det} B < 0$. Утверждается, что множество всех базисов в пространстве V распадается на два подмножества, которые обозначим $\mathcal{F}_1$ и $\mathcal{F}_2$. Все базисы подмножества $\mathcal{F}_1$ взаимно

[5]Позднее мы увидим, что это определение работает в n-мерном векторном пространстве при любом n.

подобны, так же, как и все базисы подмножества $\mathcal{F}_2$. Каждый базис из $\mathcal{F}_1$ противоположен каждому базису из $\mathcal{F}_2$. Действительно, выберем базис L_1. Пусть подмножество $\mathcal{F}_1$ содержит базисы вида

$$BL_1, \qquad \operatorname{Det} B > 0,$$

а подмножество $\mathcal{F}_2$ — базисы вида

$$BL_1, \qquad \operatorname{Det} B < 0.$$

Любой базис принадлежит одному из этих подмножеств. Если два базиса L и L' принадлежат $\mathcal{F}_1$, то

$$L = BL_1, \quad L' = B'L_1, \quad \operatorname{Det} B > 0, \quad \operatorname{Det} B' > 0,$$

следовательно,

$$L' = B'B^{-1}L \quad \text{и} \quad \operatorname{Det} B'B^{-1} = \operatorname{Det} B'(\operatorname{Det} B)^{-1} > 0,$$

а это значит, что L и L' подобны. Если в этих выражениях $\operatorname{Det} B < 0$ и $\operatorname{Det} B' < 0$, то базисы L и L' подобны. В случае, когда один определитель положителен, а другой отрицателен, базисы L и L' противоположны. Итак, если $\mathcal{F}$ обозначает множество *всех* базисов в пространстве V, то мы имеем

$$\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2, \quad \mathcal{F}_1 \cap \mathcal{F}_2 = \varnothing.$$

Выбор ориентации в пространстве V означает выбор одного из этих подмножеств. Другими словами, по определению ориентация в пространстве V — это такое подмножество базисов V, что любые два базиса из подмножества подобны, и любой базис, подобный базису из подмножества, тоже принадлежит этому подмножеству. Отметим, что выбор базиса L в пространстве V *определяет* ориентацию пространства, т. е. множество всех базисов, подобных L. Итак, пусть мы выбрали ориентацию пространства V. Тогда базис L' называется *хорошим* или *положительным*, если он принадлежит выбранному подмножеству базисов, и *плохим* или *отрицательным*, если нет. Таким образом, раз мы выбрали ориентацию, то любой базис становится либо хорошим, либо плохим. (Конечно, другой выбор ориентации изменяет эту классификацию на противоположную.)

$$\begin{tikzcd} V \arrow[r, "A"] & W \arrow[d, "M"] \\ & \mathbb{R}^2 \end{tikzcd}$$

Рис. 8.20

$$\begin{tikzcd} V \arrow[r, "A"] \arrow[d, "L"'] & W \arrow[d, "M"] \\ \mathbb{R}^2 \arrow[r, "C"'] & \mathbb{R}^2 \end{tikzcd}$$

Рис. 8.21

Пусть W — другое векторное пространство и пусть $A : V \to W$ — линейный изоморфизм, т. е. существует обратное линейное отображение $A^{-1} : W \to V$ (см. рис. 8.20). Предположим, что мы выбрали ориентацию $\mathcal{O}_V$ в пространстве V и ориентацию $\mathcal{O}_W$ в пространстве W. Пусть $M \in \mathcal{O}_W$ — хороший базис в пространстве W. Тогда $M : W \to \mathbb{R}^2$, следовательно, $M \circ A$ — изоморфизм, т. е. задает базис в пространстве V. Итак, существуют две возможности — базис $M \circ A$ либо хороший, либо плохой. Давайте посмотрим на эту альтернативу по-другому. Пусть L — хороший базис в пространстве V. Обозначим

$$C = \mathrm{Mat}_{L,M}(A) = MAL^{-1}$$

Матрица $C = MAL^{-1}$ — это матрица перехода от базиса MA к базису L (рис. 8.21); другими словами,

$$MA = CL.$$

Итак, базис MA будет хорошим тогда и только тогда, когда $\mathrm{Det}\, C > 0$. Если базис L заменить на $L' = B_1 L$ ($\mathrm{Det}\, B_1 > 0$) и M заменить на $M' = B_2 M$ ($\mathrm{Det}\, B_2 > 0$), то

$$C' = M'AL'^{-1} = B_2 MAL^{-1} B_1^{-1} = B_2 C B_1^{-1}.$$

Тогда

$$\mathrm{Det}\, C' = \mathrm{Det}\, B_2 \, \mathrm{Det}\, C \, \mathrm{Det}\, B_1^{-1}$$

имеет тот же знак, что и $\mathrm{Det}\, C$. Таким образом, знак $\mathrm{Det}\, C$ не зависит от частного выбора $L \in \mathcal{O}_V$ и $M \in \mathcal{O}_W$.

Если $\mathrm{Det}\, C > 0$, мы будем говорить, что отображение A *сохраняет ориентацию* (или положительно). Если же $\mathrm{Det}\, C < 0$, мы говорим, что A *изменяет ориентацию* (или отрицательно). Предположим, что $V \xrightarrow{A} W$ и $W \xrightarrow{A'} Z$ — два линейных изоморфизма, и мы выбрали ориентацию во всех трех пространствах. Тогда легко проверить, что

- если A и A' сохраняют ориентацию, то $A' \circ A$ тоже сохраняет ориентацию;
- если оба изоморфизма A и A' меняют ориентацию, то $A' \circ A$ ее сохраняет;
- если один изоморфизм сохраняет ориентацию, а другой изменяет, то $A' \circ A$ изменяет ориентацию.

Пусть $\phi : V \to W$ — дифференцируемое отображение. Тогда в каждой точке $\mathbf{p} \in V$ дифференциал $d\phi_{\mathbf{p}} : V \to W$ — линейное отображение. Мы говорим, что ϕ сохраняет ориентацию, если в каждой точке дифференциал $d\phi_{\mathbf{p}}$ является линейным отображением, сохраняющим ориентацию. (В частности, мы предполагаем, что $d\phi_{\mathbf{p}}$ — линейный изоморфизм для каждой точки $\mathbf{p}$.)

До сих пор мы предполагали, что пространство V — двумерное. Но это не имеет значения. Например, если V — трехмерное пространство, то все определения работают. Просто надо знать, что матрица 3×3 обратима тогда и только тогда, когда ее определитель не равен нулю, что любые три линейно независимые вектора в трехмерном пространстве V образуют базис, т. е. определяют изоморфизм с $\mathbb{R}^3$. Тогда все рассуждения, приведенные в конце главы 1, применимы и здесь. Такая же ситуация в n-мерном пространстве — требуется определение размерности n и определителя матрицы $n \times n$. Эти вопросы мы будем обсуждать в главах 10 и 11. (В пространстве одного измерения базисом является ненулевой вектор, матрица 1×1 — это просто число a, которое рассматривается как определитель этой матрицы.)

8.5. Перенос и интегрирование 2-форм

Чтобы упростить вычисление двойного интеграла, часто вводятся новые, более удобные координаты. Например, область W, изображенная в координатах x, y на рис. 8.22, становится прямоугольником R в системе координат r, θ. Переход от одной системы координат к другой осуществляется в этом случае с помощью известного преобразования α, которое дается формулами $\alpha^* x = r\cos\theta$ и $\alpha^* y = r\sin\theta$.

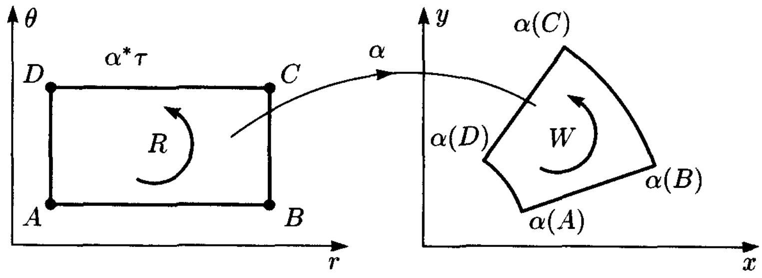


Рис. 8.22

Мы можем использовать это преобразование и свести двойной интеграл по ориентированной области $\int_W$ к интегралу от 2-формы в плоскости r, θ. Определим перенос 2-формы $\alpha^*\tau$ так, что если $W = \alpha(R)$, то

$$\int_W \tau = \int_R \alpha^*\tau.$$

Для определения переноса 2-формы мы расширим определение переноса от 1-формы очевидным образом. Для 1-формы λ мы определяли перенос $\alpha^*\lambda$ согласно формуле

$$\alpha^*\lambda(\mathbf{p})[\mathbf{v}] = \lambda(\alpha(\mathbf{p}))[d\alpha[\mathbf{v}]].$$

А теперь определим перенос 2-формы, применяя $d\alpha$ к обоим векторам аргумента τ (рис. 8.23). Таким образом, по определению

$$\alpha^*\tau(\mathbf{p})[\mathbf{v}, \mathbf{w}] = \tau(\alpha(\mathbf{p}))[d\alpha[\mathbf{v}], d\alpha[\mathbf{w}]].$$

Это определение гарантирует, что, как будет доказано, формула $\int_R \alpha^*\tau = \int_{\alpha(R)} \tau$ справедлива для любой прямоугольной области R, при условии, конечно, что α дифференцируемо, сохраняет ориентацию и оба интеграла существуют. Для доказательства

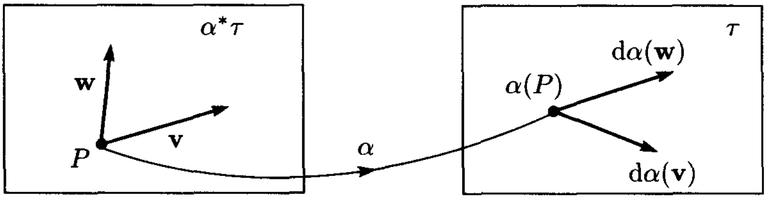


Рис. 8.23

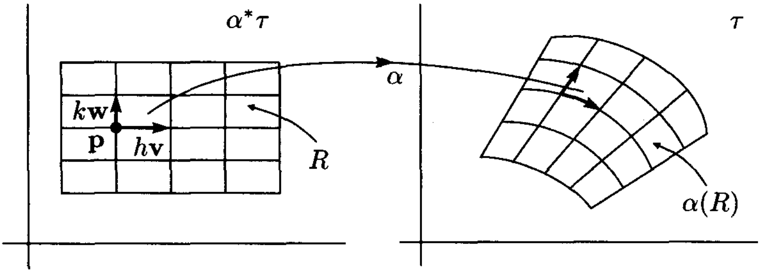


Рис. 8.24

мы будем аппроксимировать интеграл $\int_R \alpha^*\tau$ римановой суммой по большому числу малых прямоугольников (рис. 8.24).

Вклад в эту сумму прямоугольника с вершиной в точке $\mathbf{p}$ равен $\alpha^*\tau(\mathbf{p})[h\mathbf{v}, k\mathbf{w}]$. По нашему определению переноса 2-формы это равно $\tau(\alpha(\mathbf{p}))[d\alpha[h\mathbf{v}], d\alpha[k\mathbf{w}]]$, т. е. значению формы τ на параллелограмме, который является наилучшим линейным приближением к образу прямоугольника, построенного на векторах $h\mathbf{v}$ и $k\mathbf{w}$ при преобразовании α.

Конечно, образ прямоугольника при преобразовании α не является точным параллелограммом (рис. 8.25) и величина

$$\tau(\alpha(\mathbf{p}))[d\alpha(h\mathbf{v}), d\alpha(k\mathbf{w})]$$

не будет точно равна интегралу от τ по образу прямоугольника.

Таким образом, у нас возникнут погрешности двух типов: мы заменяем образ прямоугольника на параллелограмм (обозначим его P), так что

(i) $\int_{\alpha(\text{прямоуг.})}$ заменяем на $\int_P$,

и

(ii) $\int_P$ заменяем на $\tau(\alpha(\mathbf{p}))[d\alpha(h\mathbf{v}), d\alpha(k\mathbf{w})]$.

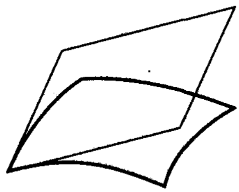


Рис. 8.25

Если τ непрерывна, то погрешность типа (ii) будет $o(hk)$. Если τ имеет равномерно ограниченные первые производные во всей области, ограниченные, скажем, числом κ, то

$$|\tau(\mathbf{q}) - \tau(\alpha(\mathbf{p}))| \leq \kappa(h^2 + k^2)^{1/2} \quad \text{для любого } \mathbf{q} \text{ в } P.$$

Тогда погрешность типа (ii) будет не больше, чем

$$\kappa(h^2 + k^2)^{1/2} hk.$$

Погрешность типа (i) можно оценить, исходя из формулы Тейлора, заменив, например, искривленный образ каждой стороны на аппроксимирующую прямую линию. Предполагая, что первая и вторая производные α ограничены в области R, можно получить, что погрешность оценивается сверху суммой слагаемых вида h^2k и hk^2, умноженных на некоторую константу. Отсюда можно получить оценку

$$\alpha^*(\tau(\mathbf{p}_i))[h\mathbf{v}, h\mathbf{w}] = \int_{\alpha(\text{прямоуг.})} \tau + \text{погрешность},$$

где

$$|\text{погрешность}| < C(h^2 + k^2)^{1/2} \times (\text{площадь } R).$$

Суммируя по всем прямоугольникам, получаем

$$\sum \alpha^*(\tau(\mathbf{p}_i))[h\mathbf{v}, h\mathbf{w}] = \int_{\alpha(R)} \tau + \text{погрешность},$$

где

$$|\text{погрешность}| < C(h^2 + k^2)^{1/2} \times (\text{площадь } R).$$

Если мы будем неограниченно измельчать разбиение, то сумма слева стремится к $\int_R \alpha^*\tau$, а погрешность справа стремится к нулю. Отсюда следует, что

$$\boxed{\int_R \alpha^*\tau = \int_{\alpha(R)} \tau.}$$

Более аккуратное доказательство этого важного результата, не требующее таких строгих ограничений на α и τ, справедливое для всех n, читатель может найти в книге Loomis and Sternberg, *Advanced Calculus*, раздел 8.11. Мы вообще рекомендуем для изучения теории интегрирования главу 8 этой книги (ее можно читать независимо от других).

Теперь необходимо разработать процедуру вычисления переноса от 2-форм. Поскольку любая 2-форма на плоскости выражается в виде произведения двух 1-форм, сначала вычисляем $\alpha^*(\lambda \wedge \sigma)$, где λ и σ обозначают 1-формы. По определению

$$\begin{aligned}\alpha^*(\lambda \wedge \sigma)[\mathbf{v}, \mathbf{w}] &= (\lambda \wedge \sigma)[d\alpha[\mathbf{v}], d\alpha[\mathbf{w}]] \\ &= \lambda[d\alpha[\mathbf{v}]]\sigma[d\alpha[\mathbf{w}]] - \lambda[d\alpha[\mathbf{w}]]\sigma[d\alpha[\mathbf{v}]].\end{aligned}$$

С другой стороны,

$$(\alpha^*\lambda \wedge \alpha^*\sigma)[\mathbf{v}, \mathbf{w}] = \alpha^*\lambda[\mathbf{v}]\alpha^*\sigma[\mathbf{w}] - \alpha^*\lambda[\mathbf{w}]\alpha^*\sigma[\mathbf{v}].$$

Согласно определению переноса для 1-формы, $\alpha^*\lambda[\mathbf{v}] = \lambda[d\alpha[\mathbf{v}]]$. Отсюда следует, что

$$\boxed{\alpha^*(\lambda \wedge \sigma) = \alpha^*\lambda \wedge \alpha^*\sigma,}$$

т. е. операция переноса коммутирует с внешним произведением.

Поскольку самый общий вид 2-формы на плоскости xy — это

$$\tau = F\,dx \wedge dy,$$

мы немедленно получаем, что

$$\alpha^*\tau = (\alpha^*F)\,d(\alpha^*x) \wedge d(\alpha^*y).$$

Если, например, $\alpha^*x = r\cos\theta$, $\alpha^*y = r\sin\theta$, то

$$\begin{aligned} d(\alpha^*x) \wedge d(\alpha^*y) &= (\cos\theta\, dr - r\sin\theta\, d\theta) \wedge (\sin\theta\, dr + r\cos\theta\, d\theta) \\ &= r\, dr \wedge d\theta. \end{aligned}$$

Итак, мы установили, что формула перехода от декартовых координат к полярным имеет вид

$$\int_R (\alpha^*F)(r,\theta) r\, dr \wedge d\theta = \int_{\alpha(R)} F(x,y)\, dx \wedge dy.$$

2-форма $r\, dr \wedge d\theta$ каждому маленькому параллелограмму на плоскости r, θ ставит в соответствие не его ориентированную площадь (что делает $dr \wedge d\theta$), а ориентированную площадь его *образа* в плоскости xy, получившегося в результате преобразования α.

Теперь мы можем получить формулу замены переменных для ориентированных двойных интегралов в общем случае. Пусть R — ориентированная область в плоскости uv, которую дифференцируемое преобразование α переводит в ориентированную область $\alpha(R)$ на плоскости xy (рис. 8.26).

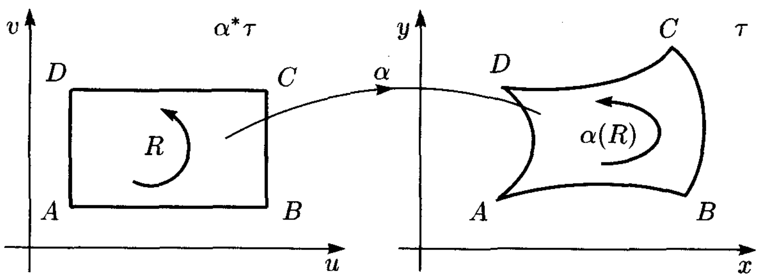


Рис. 8.26

Преобразование α задается с помощью переноса для координатных функций x и y:

$$\alpha^*x = X(u,v), \quad \alpha^*y = Y(u,v).$$

Тогда

$$\begin{aligned} d(\alpha^*x) &= \frac{\partial(\alpha^*x)}{\partial u}\, du + \frac{\partial(\alpha^*x)}{\partial v}\, dv = \frac{\partial X}{\partial u}\, du + \frac{\partial X}{\partial v}\, dv, \\ d(\alpha^*y) &= \frac{\partial(\alpha^*y)}{\partial u}\, du + \frac{\partial(\alpha^*y)}{\partial v}\, dv = \frac{\partial Y}{\partial u}\, du + \frac{\partial Y}{\partial v}\, dv. \end{aligned}$$

2-форма τ на плоскости xy записывается в виде $\tau = F(x,y)\,dx \wedge dy$. Перенос этой 2-формы равен

$$\begin{aligned}\alpha^*\tau &= \alpha^* F\, d(\alpha^* x) \wedge d(\alpha^* y) \\ &= \alpha^* F \left[\frac{\partial(\alpha^* x)}{\partial u}\frac{\partial(\alpha^* y)}{\partial v} - \frac{\partial(\alpha^* x)}{\partial v}\frac{\partial(\alpha^* y)}{\partial u}\right] du \wedge dv,\end{aligned}$$

или

$$\alpha^*\tau = F(X(u,v), Y(u,v)) \left[\frac{\partial X}{\partial u}\frac{\partial Y}{\partial v} - \frac{\partial X}{\partial v}\frac{\partial Y}{\partial u}\right] du \wedge dv.$$

Множитель в квадратных скобках является детерминантом матрицы Якоби (*якобиана*) J, который выражает $d\alpha$ в выбранных координатах,

$$J = \begin{pmatrix} \partial(\alpha^* x)/\partial u & \partial(\alpha^* x)/\partial v \\ \partial(\alpha^* y)/\partial u & \partial(\alpha^* y)/\partial v \end{pmatrix} = \begin{pmatrix} \partial X/\partial u & \partial X/\partial v \\ \partial Y/\partial u & \partial Y/\partial v \end{pmatrix}.$$

Следовательно, мы можем написать

$$\int_{\alpha(R)} F\, dx \wedge dy = \int_R \alpha^* F \operatorname{Det} J\, du \wedge dv.$$

Это равенство совершенно естественно. Поскольку $\operatorname{Det} J$ является множителем, «преобразующим площадь» для линейного преобразования $d\alpha$, то $\operatorname{Det} J\, du \wedge dv$ каждой малой области на плоскости uv ставит в соответствие ориентированную площадь его образа в плоскости xy. Если порядок u и v определяется ориентацией R, а порядок x и y — ориентацией $\alpha(R)$, причем α сохраняет ориентацию, то якобиан J будет иметь положительный определитель. Изменение порядка u и v (или порядка x и y) соответствует изменению ориентации на плоскости uv (на плоскости xy). При этом столбцы или строки J меняются местами, что вызывает изменение знака $\operatorname{Det} J$.

В качестве примера рассмотрим вычисление площади ориентированной области W, ограниченной осью x, прямой $y = mx$, гиперболой $x^2 - y^2 = 1$ и гиперболой $x^2 - y^2 = 4$ (рис. 8.27, 8.28). Чтобы это сделать, напишем $W = \alpha(R)$, где $\alpha^* x = u \cosh v$ и

$\alpha^* y = u \sinh v$. Отображение α переводит ориентированный прямоугольник R, заданный неравенствами

$$1 \le u \le 2, \quad 0 \le v \le \operatorname{arctanh} m,$$

в область W. Например, вертикальный отрезок прямой $u = 2$ отображается в кусок гиперболы $x^2 - y^2 = 4$.

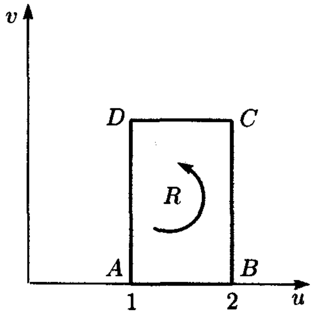


Рис. 8.27

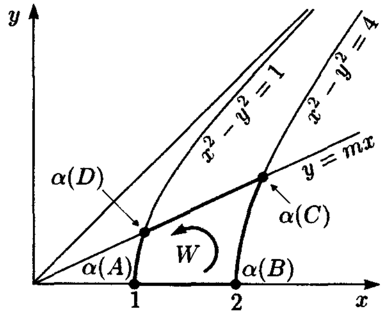


Рис. 8.28

Поскольку в ориентации области W первой координатой является x, ее площадь $A = \int_W dx \wedge dy$. Выполнив перенос, получаем $A = \int_R \alpha^*(dx \wedge dy) = \int_R \operatorname{Det} J \, du \wedge dv$. Здесь $\alpha^*(dx \wedge dy) = d(u \cosh v) \wedge d(u \sinh v) = (\cosh v \, du + \sinh v \, dv) \wedge (\sinh v \, du + u \cosh v \, dv) = u \cosh^2 v \, du \wedge dv + u \sinh^2 v \, dv \wedge du = u \, du \wedge dv$, следовательно,

$$A = \int_R u \, du \wedge dv = \int_1^2 u \, du \int_0^{\tanh^{-1} m} dv = \frac{3}{2} \tanh^{-1} m.$$

С другой стороны, мы можем вычислить

$$\operatorname{Det} J = \operatorname{Det} \begin{pmatrix} \cosh v & u \sinh v \\ \sinh v & u \cosh v \end{pmatrix} = u$$

и получить тот же самый интеграл $A = \int_R u \, du \wedge dv$.

Мы видим, что секрет полезности преобразования координат состоит в том, что границы области W становятся образами сторон прямоугольника на плоскости uv. Если, например, W — треугольник, заключенный между осями координат и прямой

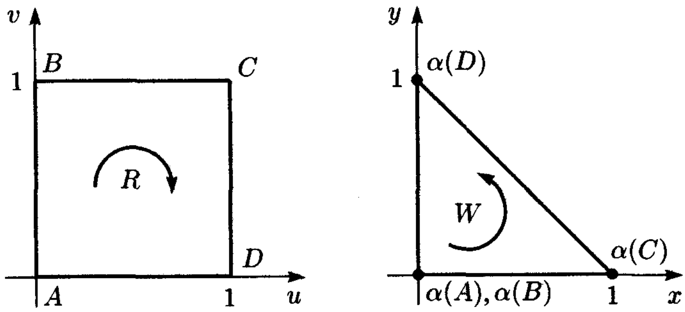


Рис. 8.29

$x + y = 1$, то полезным преобразованием координат будет такое, которое переводит отрезки прямых $u = \text{const}$ с фиксированным интервалом по v в отрезки $x + y = \text{const}$ между осями координат. Такое преобразование координат описывается формулами: $\alpha^* x = uv$, $\alpha^* y = u(1 - v)$, которые обладают свойством $\alpha^*(x + y) = u$. Читателю рекомендуется проверить, что α преобразует квадрат единичной площади R на плоскости uv в область W. Если в ориентации R сделать v первой координатой, то ориентация $\alpha(R)$ будет согласована с ориентацией W. Итак, если v — первая координата, то якобиан

$$J = \begin{pmatrix} \partial(\alpha^* x)/\partial v & \partial(\alpha^* x)/\partial u \\ \partial(\alpha^* y)/\partial v & \partial(\alpha^* y)/\partial u \end{pmatrix} = \begin{pmatrix} u & v \\ -u & 1 - v \end{pmatrix}$$

имеет положительный определитель, равный u, что подтверждает предыдущее утверждение. (Если первой координатой считать u, то $\operatorname{Det} J = -u$.)

Используем это преобразование для вычисления интеграла

$$I = \int_W \frac{e^{-(x+y)}}{\sqrt{xy}}\, dx \wedge dy,$$

который довольно сложно получить как повторный интеграл по x и y. Перейдем к новым переменным. Получаем

$$\alpha^* \left(\frac{e^{-(x+y)}}{\sqrt{xy}} \right) = \frac{e^{-u}}{\sqrt{u^2 v(1 - v)}} = \frac{e^{-u}}{u} \frac{1}{\sqrt{v(1 - v)}}$$

и $\alpha^*(dx \wedge dy) = (u\,dv + v\,du) \wedge (-u\,dv + (1-v)\,du) = u\,dv \wedge du$. Тогда на плоскости uv имеем

$$I = \int_R \frac{e^{-u}}{\sqrt{v(1-v)}}\, dv \wedge du.$$

Порядок дифференциалов $dv \wedge du$ под знаком интеграла связан с тем, что в ориентации R первой координатой является v. Окончательно получаем

$$I = \int_0^1 e^{-u}\, du \int_0^1 \frac{dv}{\sqrt{v(1-v)}} = (1 - e^{-1})\pi.$$

(При вычислении второго интеграла мы воспользовались формулой

$$\int \frac{dv}{\sqrt{v(1-v)}} = -\arcsin(-2v+1).)$$

Наконец, перейдем к задаче замены переменных для вычисления *абсолютного* двойного интеграла $I = \int_W F\,dx\,dy$. Для этого сначала преобразуем этот интеграл в ориентированный двойной интеграл $\int_W F\,dx \wedge dy$, соответствующий ориентации W, для которой первой координатой является x.

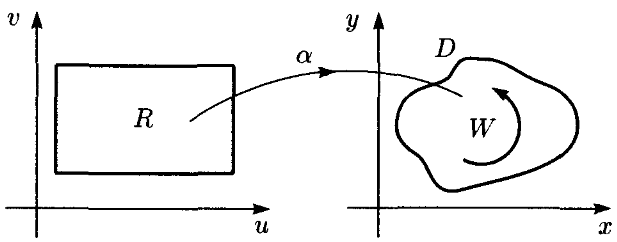


Рис. 8.30

Теперь напишем W как $\alpha(R)$ и зададим ориентацию R (рис. 8.30). Если $\operatorname{Det} J\left(\dfrac{x,y}{u,v}\right)$ положителен, то выберем на R ориентацию, при которой первой координатой является u. Тогда

$$I = \int\limits_{\left(\substack{R \\ u \text{ первая}}\right)} \alpha^* F(\operatorname{Det} J)\, du \wedge dv = \int\limits_{\left(\substack{R \\ \text{неориент.}}\right)} \alpha^* F(\operatorname{Det} J)\, du\, dv.$$

Если $\operatorname{Det} J\left(\dfrac{x,y}{u,v}\right)$ отрицателен, то зададим на R ориентацию, при которой первой координатой является v. Тогда

$$I = \int\limits_{\left(\substack{R\\ v\ \text{первая}}\right)} \alpha^* F(\operatorname{Det} J)\, dv \wedge du = -\int\limits_{\left(\substack{cR\\ \text{неориент.}}\right)} \alpha^* F(\operatorname{Det} J)\, du\, dv.$$

В обоих случаях можно брать *абсолютное* значение $\operatorname{Det} J$:

$$I = \int_R \alpha^* F \left|\operatorname{Det} J\left(\frac{x,y}{u,v}\right)\right| du\, dv.$$

При этом не возникает вопроса об ориентации или о порядке переменных интегрирования. Перестановка x и y или u и v не влияет на абсолютное значение $\operatorname{Det} J$.

В главе 15 мы обсудим интегрирование форм более высоких степеней.

8.6. 2-формы в трехмерном пространстве

В предыдущем параграфе мы определили понятие переноса для 2-форм. При этом правила вычисления очень простые: если ω_1 и ω_2 — *линейные* дифференциальные формы, то

$$\phi^*(\omega_1 \wedge \omega_2) = \phi^*\omega_1 \wedge \phi^*\omega_2.$$

Если f — функция и τ — 2-форма, то

$$\phi^*(f\tau) = \phi^*(f)\phi^*\tau.$$

Если τ_1 и τ_2 — 2-формы, то

$$\phi^*(\tau_1 + \tau_2) = \phi^*\tau_1 + \phi^*\tau_2.$$

Таким образом, при работе с дифференциальными формами сохраняются все обычные свойства алгебраических операций. Кроме того, мы определили линейный оператор d, преобразующий 1-формы в 2-формы:

$$d(f\,dg) = df \wedge dg,$$

или в более общем виде:

$$d(f\omega) = df \wedge \omega + f\,d\omega,$$

удовлетворяющий условию

$$d(\omega_1 + \omega_2) = d\omega_1 + d\omega_2.$$

Перенос ϕ^* коммутирует с оператором d в том смысле, что

$$\phi^*(df) = d\phi^* f, \quad \text{где } f \text{ — функция}$$

и

$$\phi^* d\omega = d\phi^*\omega, \quad \text{где } \omega \text{ — 1-форма.}$$

2-формой в трехмерном пространстве (в декартовых координатах x, y, z) мы по определению будем называть выражение вида

$$a\,dx \wedge dy + b\,dx \wedge dz + c\,dy \wedge dz,$$

где a, b и c — функции. Если дана 1-форма

$$\omega = A\,dx + B\,dy + C\,dz,$$

то правила для внешней производной d и для внешнего умножения дают

$$\begin{aligned}
d\omega &= dA \wedge dx + dB \wedge dy + dC \wedge dz \\
&= \left(\frac{\partial A}{\partial x}\,dx + \frac{\partial A}{\partial y}\,dy + \frac{\partial A}{\partial z}\,dz\right) \wedge dx \\
&\quad + \left(\frac{\partial B}{\partial x}\,dx + \frac{\partial B}{\partial y}\,dy + \frac{\partial B}{\partial z}\,dz\right) \wedge dy \\
&\quad + \left(\frac{\partial C}{\partial x}\,dx + \frac{\partial C}{\partial y}\,dy + \frac{\partial C}{\partial z}\,dz\right) \wedge dz \\
&= \left(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right) dx \wedge dy + \left(\frac{\partial C}{\partial x} - \frac{\partial A}{\partial z}\right) dx \wedge dz \\
&\quad + \left(\frac{\partial C}{\partial y} - \frac{\partial B}{\partial z}\right) dy \wedge dz.
\end{aligned}$$

Если $\phi : \mathbb{R}^2 \to \mathbb{R}^3$ и τ — 2-форма в пространстве $\mathbb{R}^3$, то $\phi^*\tau$ будет 2-формой в пространстве $\mathbb{R}^2$.

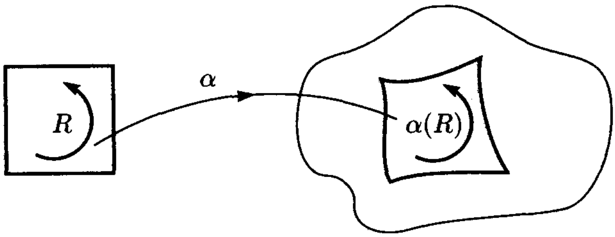


Рис. 8.31

Если R — некоторая область в $\mathbb{R}^2$ и мы выбрали в ней ориентацию, то можно написать $\int_R \phi^*\tau$, который рассматривается как интеграл от τ по ориентированной поверхности $\alpha(R)$.

8.7. Различие между 2-формами и плотностями

Вернемся в двумерное пространство. Мы помним, что перенос $\alpha^*\tau$ определен для любой 2-формы τ. Другими словами, если α отображает плоскость uv в плоскость xy, то

$$\alpha^*(f\,dx \wedge dy) = (\alpha^* f) \cdot (\operatorname{Det} J)\,du \wedge dv, \tag{8.2}$$

где якобиан перехода J определяется матрицей

$$J = \begin{pmatrix} \partial\alpha_1/\partial u & \partial\alpha_1/\partial v \\ \partial\alpha_2/\partial u & \partial\alpha_2/\partial v \end{pmatrix}, \qquad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}.$$

Эта формула верна для любого дифференцируемого отображения α. Кроме того, мы доказали, что если α — взаимнооднозначное отображение, *сохраняющее ориентацию*, то

$$\int_R \alpha^*(\tau) = \int_{\alpha(R)} \tau. \tag{8.3}$$

Это верно для интегралов от 2-форм, когда ориентация имеет значение. Предположим, однако, что мы хотим вычислить абсолютный интеграл. Тогда выбор ориентации не имеет значения, но формула для замены переменных или перенос для выражения

$f\,dA$ или $f\,dx\,dy$ не будет совпадать с (8.2). Давайте вернемся назад, к доказательству формулы замены переменных. Тогда интеграл $\int_{\alpha(\text{прямоуг.})}$ мы заменяли на

$$\tau(P)[d\alpha(h\mathbf{v}), d\alpha(k\mathbf{w})]$$
$$= f(P) \times \text{ориентированную площадь прямоугольника.}$$

Очевидно, что в случае вычисления абсолютного интеграла мы должны заменить ориентированную площадь на абсолютное значение этой площади. Таким образом, в формуле (8.2) надо написать

$$\alpha^*(f\,dx\,dy) = \alpha^* f \cdot |\operatorname{Det} J|\,du\,dv. \tag{8.4}$$

Читателю предлагается проверить в качестве поучительного упражнения, что если α — взаимно-однозначное отображение, то

$$\int_R \alpha^*(f\,dx\,dy) = \int_{\alpha(R)} f\,dx\,dy$$

независимо от ориентации.

Итак, 2-формы $\tau = f\,dx \wedge dy$ и плотности $f\,dx\,dy$ являются совершенно разными объектами — при замене переменных они преобразуются по-разному. Например, плотность может быть положительной и отрицательной (плотность электрического заряда): если $f \geq 0$, то при замене переменных $f\,dx\,dy$ становится $\alpha^*(f)|\operatorname{Det} J|\,du\,dv$ и $\alpha^* f \cdot |\operatorname{Det} J|$ по-прежнему положительна (если $\operatorname{Det} J \neq 0$, что выполняется при дифференцируемости α^{-1}). Однако, не имеет смысла задавать вопрос о знаке 2-формы τ, потому что $\operatorname{Det} J$, входящий в (8.2), может быть и положительным, и отрицательным. Плотности и 2-формы совпадают, только если мы выбрали ориентацию и делаем замену переменных, сохраняющую эту ориентацию (для них $\operatorname{Det} J > 0$).

8.8. Формула Грина на плоскости

Рассматривая криволинейные интегралы, мы уже встречались с обобщением фундаментальной теоремы интегрального исчисления — формулы Ньютона–Лейбница, а именно, с формулой

$$\int_\gamma df = f(B) - f(A),$$

где криволинейная трактория γ проходит от точки A к точке B (рис. 8.32). Эта теорема связывает интеграл от df по одномерной области (т. е. по пути γ) со значениями самой функции f в конечных точках траектории.

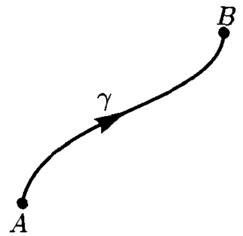


Рис. 8.32

Аналогичный результат, связывающий двумерную область и ее одномерные границы, известен как *формула Грина.* Эта теорема утверждает, что для любой дифференцируемой 1-формы τ и любой ориентированной области на плоскости R справедливо равенство

$$\int_R d\tau = \int_{\partial R} \tau.$$

Слева 2-форма $d\tau$ интегрируется по области R, а справа 1-форма τ интегрируется по пути ∂R, являющимся *границей* области R.[6]

[6]Для того, чтобы обе части этой формулы существовали, область R и ее граница ∂R должны удовлетворять некоторым требованиям, например, R ограничена, ∂R кусочно-непрерывно-дифференцируема. — *Прим. ред.*

Ориентация R определяет направление движения по ∂R. Например, если R — кольцевая область, ориентированная против часовой стрелки, как показано на рис. 8.33, то ∂R состоит из внешней окружности, по которой движение происходит против часовой стрелки, и внутренней окружности, проходимой по часовой стрелке.

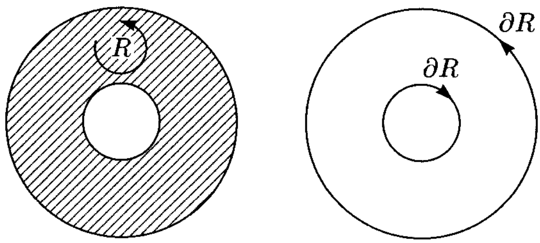


Рис. 8.33

Прежде чем формально доказывать теорему Грина, имеет смысл еще раз вспомнить определение оператора d, действующего на 1-форму. Это поможет понять, почему должна быть такая теорема. Мы определяли $d\tau$ как антисимметричную билинейную функцию пары векторов $\mathbf{v}, \mathbf{w}$, обладающую свойством

$$d\tau(P)[h\mathbf{v}, k\mathbf{w}] = \int\limits_{P(h,k)} \tau + \text{погрешность},$$

где $P(h, k)$ есть параллелограмм, построенный на векторах $h\mathbf{v}$ и $k\mathbf{w}$, причем погрешность стремится к нулю быстрее, чем произведение hk, т. е. быстрее площади параллелограмма.

Рассмотрим область, состоящую из N параллелограммов, каждый из которых построен на векторах $h\mathbf{v}$ и $k\mathbf{w}$ (рис. 8.34). Тогда мы имеем

$$\sum_{i=1}^{N} d\tau(P_i)[h\mathbf{v}, k\mathbf{w}] = \sum_{i=1}^{N} \int\limits_{P_i} \tau + \sum_{i=1}^{N} (\text{погрешность})_i.$$

В сумме криволинейных интегралов по контурам параллелограммов вклады от внутренних сторон, принадлежащих одновременно двум параллелограммам, взаимно уничтожаются, так как они

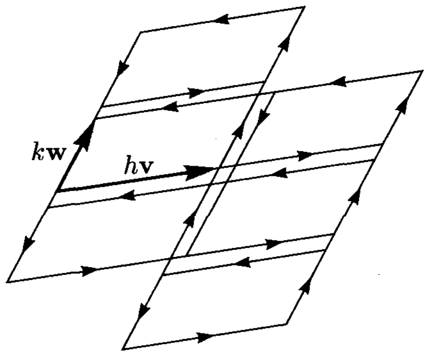


Рис. 8.34

проходятся в противоположных направлениях. Таким образом, остается один криволинейный интеграл по границе области R:

$$\sum_{i=1}^{N} d\tau(P_i)[h\mathbf{v}, k\mathbf{w}] = \int\limits_{\partial R} \tau + \sum_{i=1}^{N} (\text{погрешность})_i.$$

Если h и k стремятся к нулю, то сумма в левой части равенства стремится к $\int_R d\tau$. Поскольку число N пропорционально $1/hk$ и каждый поправочный член стремится к нулю быстрее, чем hk, сумма поправочных членов стремится к нулю при h и k, стремящихся к нулю. В результате мы получаем

$$\int\limits_R d\tau = \int\limits_{\partial R} \tau$$

для любой области, состоящей из параллелограммов.

Чтобы доказать теорему Грина более строго, рассмотрим сначала частный случай, когда $\tau = G(x, y)\, dy$, и вычислим криволинейный интеграл $\int_{\partial R} \tau$ по контуру прямоугольника, стороны которого параллельны осям x и y (рис. 8.35).

Вклад в интеграл дает только сторона $x = a$, проходимая от $y = 0$ до $y = b$, и сторона $x = 0$, проходимая от $y = b$ к $y = 0$.

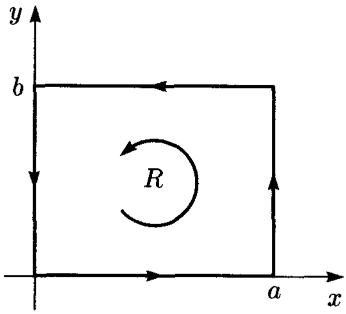


Рис. 8.35

Можно записать:

$$\int\limits_{\partial R} \tau = \int_0^b G(a, y)\, dy + \int_b^0 G(0, y)\, dy$$
$$= \int_0^b (G(a, y) - G(0, y))\, dy.$$

Согласно фундаментальной теореме интегрального исчисления последнее подинтегральное выражение можно записать так:

$$G(a, y) - G(0, y) = \int_0^a \frac{\partial G}{\partial x}(x, y)\, dx.$$

Поэтому криволинейный интеграл $\int_{\partial R} \tau$ можно записать как повторный интеграл

$$\int_0^b \left(\int_0^a \frac{\partial G}{\partial x}(x, y)\, dx \right) dy,$$

который равен ориентированному двойному интегралу

$$\int_R \frac{\partial G}{\partial x}(x, y)\, dx \wedge dy.$$

Аналогичные соображения для 1-формы $F(x,y)\,dx$ дают

$$\int\limits_{\partial R} F(x,y)\,dx = \int_0^a \left(\int_0^b \frac{\partial F}{\partial y}(x,y)\right) dx \wedge dy$$
$$= -\int_0^a (F(x,b) - F(x,0))\,dx = -\int\limits_R \frac{\partial F}{\partial y}\,dx \wedge dy.$$

Сложив это с предыдущим результатом, получаем

$$\int\limits_{\partial R} F(x,y)\,dx + G(x,y)\,dy = \int\limits_R \frac{\partial G}{\partial x}(x,y)\,dx - \frac{\partial F}{\partial y}(x,y)\,dx \wedge dy.$$

Поскольку для дифференциальной 1-формы

$$\tau = F(x,y)\,dx + G(x,y)\,dy$$

внешняя производная равна

$$d\tau = -\frac{\partial F}{\partial y}(x,y)\,dx \wedge dy + \frac{\partial G}{\partial x}(x,y)\,dx \wedge dy,$$

то последнее равенство с интегралами означает справедливость формулы Грина для координатного прямоугольника R:

$$\int\limits_R d\tau = \int\limits_{\partial R} \tau.$$

А теперь докажем теорему Грина в более общем случае, когда область на плоскости является образом прямоугольника при гладком биективном отображении α. Стратегия уже знакома: мы перенесем интегралы $\int_{\partial[\alpha(R)]} \tau$ и $\int_{\alpha(R)} d\tau$ на плоскость st, где рассматриваемая область интегрирования является прямоугольником (рис. 8.36).

Мы будем пользоваться условием, что образом границы области R является граница $\alpha(R)$. Следовательно,

$$\int\limits_{\partial[\alpha(R)]} \tau = \int\limits_{\partial R} \alpha^* \tau.$$

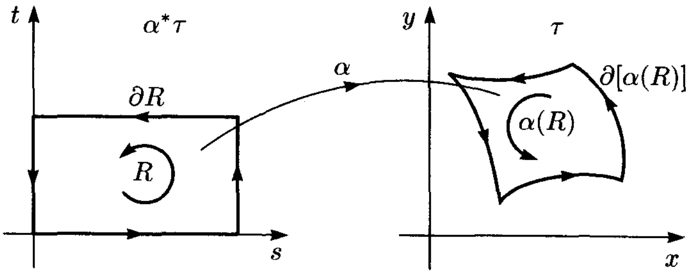


Рис. 8.36

Однако, для прямоугольной области R мы уже доказали, что

$$\int\limits_{\partial R} \alpha^*\tau = \int\limits_{R} d(\alpha^*\tau).$$

Более того, согласно определению переноса, мы имеем

$$\int\limits_{\alpha(R)} d\tau = \int\limits_{R} \alpha^*(d\tau).$$

Следовательно, для доказательства равенства $\int_{\partial[\alpha(R)]} \tau = \int_{\alpha(R)} d\tau$ достаточно показать, что

$$\alpha^*(d\tau) = d(\alpha^*\tau).$$

Это легко сделать прямым вычислением. Пусть $\tau = f\,dx + g\,dy$. Используя равенства $\alpha^*(dx) = d(\alpha^*x)$, $\alpha^*(dy) = d(\alpha^*y)$, получаем, что $\alpha^*\tau = (\alpha^*f)\,d(\alpha^*x) + (\alpha^*g)\,d(\alpha^*y)$. Далее, из формулы $d(f\,dh) = df \wedge dh$ имеем

$$d(\alpha^*\tau) = d(\alpha^*f) \wedge d(\alpha^*x) + d(\alpha^*g) \wedge d(\alpha^*y).$$

С другой стороны, мы знаем, что

$$d\tau = df \wedge dx + dg \wedge dy.$$

Правило $\alpha^*(\sigma \wedge \omega) = \alpha^*\sigma \wedge \alpha^*\omega$ дает нам

$$\alpha^* d\tau = \alpha^*(df) \wedge \alpha^*(dx) + \alpha^*(dg) \wedge \alpha^*(dy),$$

следовательно,

$$\alpha^* d\tau = d(\alpha^* f) \wedge d(\alpha^* x) + d(\alpha^* g) \wedge d(\alpha^* y).$$

Из сравнения этого выражения с (8.2) получаем то, что нам требовалось:

$$d(\alpha^* \tau) = \alpha^*(d\tau).$$

Таким образом, мы имеем

$$\int_{\alpha(\partial R)} \tau = \int_{\partial R} \alpha^* \tau = \int_R d(\alpha^* \tau) = \int_R \alpha^*(d\tau) = \int_{\alpha(R)} d\tau,$$

т. е. формула Грина доказана для любой области, которая является образом прямоугольника.

Добавим еще доказательство этой формулы для случая, когда область является многоугольником. На плоскости любой многоугольник можно разложить на треугольники. Действительно, любой многоугольник разбивается на выпуклые многоугольники (рис. 8.37). А любой выпуклый многоугольник разлагается на треугольники, если взять точку внутри него и соединить эту точку со всеми вершинами (рис. 8.38).

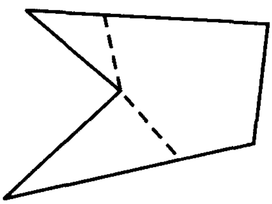


Рис. 8.37

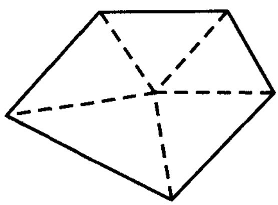


Рис. 8.38

Если бы мы знали формулу Грина для треугольников, то она была бы нам известна для всех многоугольников, так как интегралы, вычисленные по внутренним границам треугольников, взаимно уничтожаются (рис. 8.39).

Итак, мы должны доказать теорему для треугольников. Аффинное преобразование переводит любой треугольник в треугольник. Следовательно, инвариантность интегралов при гладком (в частности, аффинном) преобразовании означает, что теорему достаточно доказать для какого-либо одного треугольника.

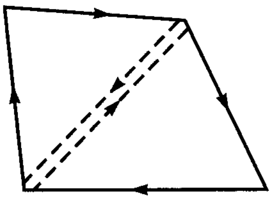


Рис. 8.39

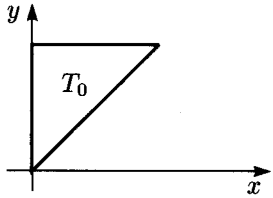


Рис. 8.40

Итак, на плоскости xy рассмотрим треугольник (рис. 8.40)

$$T_0 = \{0 \le x \le y \le 1\}.$$

Мы можем рассмотреть четырехугольник

$$T_\varepsilon\{0 \le x \le y,\ \varepsilon \le y \le 1\},$$

где $0 \le \varepsilon < 1$. Очевидно, что если мы докажем теорему Грина для T_ε, она будет верна в предельном случае для треугольника T_0, поскольку криволинейные интегралы и интегралы по площади внутри малой области при $\varepsilon \to 0$ стремятся к нулю. Область T_ε для $\varepsilon > 0$ есть образ прямоугольника $\{0 \le u \le 1,\ \varepsilon \le v \le 1\}$ на плоскости uv, полученный в результате отображения

$$x = uv$$
$$y = v.$$

Это отображение имеет гладкое обратное

$$u = x/y,$$
$$v = y,$$

если $y \ge \varepsilon > 0$. Следовательно, мы свели теорему к случаю, который уже доказан, т. е. к образу прямоугольника (рис. 8.41).

В качестве иллюстрации применения формулы Грина рассмотрим интеграл от 1-формы

$$\tau = x^2\,dy$$

по замкнутому пути ∂R (рис. 8.42), образованному отрезком прямой от точки $\begin{pmatrix}0\\0\end{pmatrix}$ к точке $\begin{pmatrix}1\\0\end{pmatrix}$, параболой $y = 1 - x^2$ от точки $\begin{pmatrix}1\\0\end{pmatrix}$ к точке $\begin{pmatrix}0\\1\end{pmatrix}$ и отрезком от точки $\begin{pmatrix}1\\0\end{pmatrix}$ к точке $\begin{pmatrix}0\\0\end{pmatrix}$.

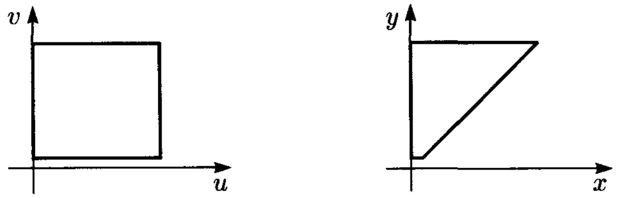


Рис. 8.41

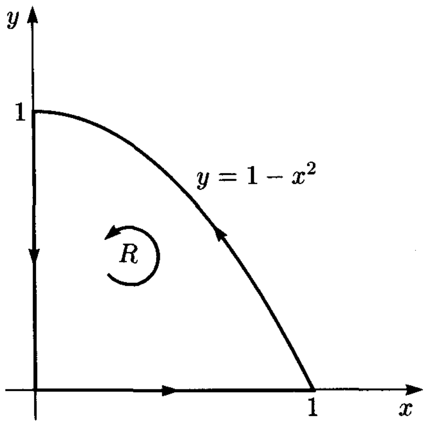


Рис. 8.42

Прямолинейные отрезки, как нетрудно проверить, не дают вклада в интеграл. Параметризуя параболу

$$\beta^* x = 1 - t, \quad \beta^* y = 1 - (1-t)^2 = 2t - t^2, \quad 0 < t < 1,$$

получаем

$$\beta^* \tau = (1-t)^2(2-2t)\,dt = 2(1-t)^3\,dt,$$

и

$$\int_{\partial R} \tau = \int_0^1 2(1-t)^3\,dt = \frac{1}{2}.$$

По формуле Грина интеграл от $d\tau = 2x\,dx \wedge dy$ по области R должен давать то же самое. Вычисляем этот интеграл как повторный и получаем

$$\int_0^1 2x\,dx \int_0^{1-x^2} dy = \int_0^1 2(1-x^2)x\,dx = \frac{1}{2},$$

что и следовало ожидать.

Формулу Грина можно использовать для вычисления площадей через криволинейные интегралы. Например, если $\tau = x\,dy$, то $d\tau = dx \wedge dy$ и

$$\int_{\partial R} \tau = \int_R dx \wedge dy = \text{площади } R,$$

предполагая, что R ориентирована против часовой стрелки. Конечно, любая 1-форма $\tau' = x\,dy + df$, где f — произвольная дифференцируемая функция, обладает этим свойством, так как

$$d\tau' = d(x\,dy) + d(df) = dx \wedge dy = d\tau.$$

Например, выберем $f = -\frac{1}{2}xy$. Тогда получаем

$$\tau' = x\,dy - \frac{1}{2}x\,dy - \frac{1}{2}y\,dx = \frac{1}{2}(x\,dy - y\,dx).$$

Вводя полярные координаты согласно формулам[7]

$$\alpha^* x = r\cos\theta, \quad \alpha^* y = r\sin\theta,$$

мы можем получить

$$\alpha^*\tau' = \frac{1}{2}r^2\,d\theta,$$

откуда следует хорошо известная формула

$$A = \int_0^{2\pi} \frac{1}{2}r^2\,d\theta$$

для площади, ограниченной замкнутой кривой, заданной в полярных координатах.

[7]В практических вычислениях символ α^* часто опускают и пишут просто $x = r\cos\theta$, $y = r\sin\theta$.

В заключение выпишем основные формулы этой главы:

$$
\begin{aligned}
\lambda \wedge \sigma &= -\sigma \wedge \lambda, \\
(\omega_1 + \omega_2) \wedge \sigma &= \omega_1 \wedge \sigma + \omega_2 \wedge \sigma, \\
d(f\omega) &= df \wedge \omega + f\, d\omega, \\
\alpha^*(\lambda \wedge \sigma) &= \alpha^*\lambda \wedge \alpha^*\sigma, \\
\alpha^*(\tau_1 + \tau_2) &= \alpha^*\tau_1 + \alpha^*\tau_2, \\
\alpha^*\, d\tau &= d\alpha^*\tau, \\
\int_R \alpha^*\tau &= \int_{\alpha(R)} \tau \quad \text{для } \alpha\text{, сохраняющего ориентацию,} \\
\int_{\partial R} \tau &= \int_R d\tau.
\end{aligned}
$$

Резюме

A. 2-формы

Для заданной на плоскости дифференциальной 1-формы τ надо уметь давать определение ее внешней производной $d\tau$ и вычислять $d\tau$ в координатах x, y.

Вы должны знать определение и уметь вычислять интеграл от 2-формы по ориентированному прямоугольнику на плоскости.

B. Двойные интегралы

Вы должны уметь вычислять двойные интегралы по заданным на плоскости областям, сводя их к повторным интегралам с соответствующими пределами интегрирования; уметь изменять порядок интегрирования в повторном интеграле, преобразуя его к двойному интегралу.

Выбрав преобразование одной области на плоскости в другую, надо уметь вычислять интегралы по второй области с помощью переноса. Надо уметь находить такое преобразование, в результате которого упрощается вычисление двойного интеграла.

C. Формула Грина

Надо знать формулу Грина на плоскости и уметь применять ее.

Задачи

8.1. В трех случаях, перечисленных ниже, u и v являются функциями на плоскости, а x и y — аффинные координаты. Выразите $dx \wedge dy$ через $du \wedge dv$. Нарисуйте графики типичных кривых $u = \text{const}$ и $v = \text{const}$ в первой четверти плоскости $(x, y > 0)$ и попробуйте дать геометрическую интерпретацию соотношениям между $dx \wedge dy$ и $du \wedge dv$ на примере параллелограмма, стороны которого касательны к кривым $u = \text{const}$ и $v = \text{const}$.

$$
\begin{array}{lll}
\text{(a)} & x = u\cos v, & y = u\sin v. \\
\text{(b)} & x = u\cosh v, & y = u\sinh v. \\
\text{(c)} & x = u^2 - v^2, & y = 2uv.
\end{array}
$$

8.2. Вычислите интеграл $\iint_S x^2y^2\,dx\,dy$, где область S расположена в первой четверти и находится между гиперболами $xy = 1$, $xy = 2$ и прямыми линиями $y = x$, $y = 4x$.

8.3. (a) Изменяя порядок интегрирования, покажите, что

$$\int_0^a \left(\int_0^y e^{m(a-x)} f(x)\,dx\right) dy = \int_0^a (a-x)e^{m(a-x)} f(x)\,dx,$$

где a и m — постоянные и $a > 0$.

(b) Покажите, что

$$\int_0^x \left(\int_0^v \left[\int_0^u f(t)\,dt\right] du\right) dv = \frac{1}{2}\int_0^x (x-t)^2 f(t)\,dt.$$

Если делать это последовательно в два шага, то вам не придется рассматривать тройной интеграл!!

8.4. Вычислите повторный интеграл

$$I = \int_0^1 y\left(\int_0^{1-y^2} \frac{\sin \pi x}{x^4}\,dx\right) dy,$$

представив его как двойной по соответствующей области W. После этого вычислите интеграл как повторный, изменив порядок интегрирования. Нарисуйте график области W.

8.5. Рассмотрите отображение, определенное формулами

$$x = u + v, \quad y = v - u^2.$$

(a) Вычислите якобиан этого отображения в переменных u и v.

(b) Треугольник T лежит на плоскости uv и имеет вершины в точках: $(0,0)$, $(2,0)$, $(0,2)$. Нарисуйте образ этого треугольника S на плоскости xy.

(c) Вычислите площадь S как двойной интеграл по области S и как двойной интеграл по области T.

(d) Вычислите интеграл

$$\iint_S \frac{dx\,dy}{(x-y+1)^2}.$$

(a) Пусть S обозначает единичный квадрат на плоскости uv (рис. 8.43). Опишите и нарисуйте образ S при отображении

$$\phi : \begin{pmatrix} u \\ v \end{pmatrix} \mapsto \begin{pmatrix} uv \\ v(2-u^2) \end{pmatrix}.$$

На рисунке покажите $\phi(A)$, $\phi(B)$ и $\phi(C)$.

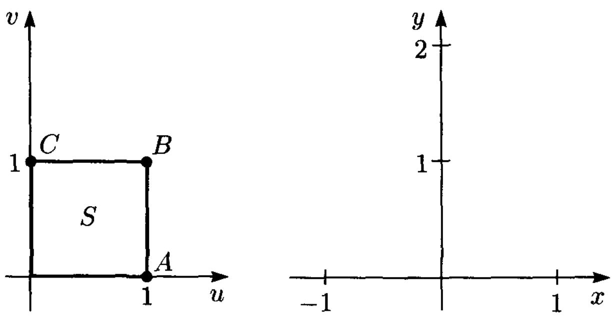


Рис. 8.43

(b) Вычислите интеграл $\iint y\,dx\,dy$ по этой области на плоскости xy.

(c) Вычислите этот же интеграл, интегрируя соответствующую функцию по квадрату S на плоскости uv.

(a) Вычислите интеграл

$$\int_W (x+2y)\,dx\,dy$$

для треугольной области W, изображенной на рис. 8.44.

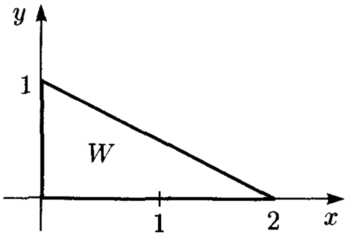


Рис. 8.44

(b) Вычислите этот же интеграл в переменных u и v, определенных формулами $x = 2uv$, $y = u - uv$.

8.8. Вычислите интеграл

$$\int_0^a \left(\int_0^x \sqrt{x^2 + y^2}\, dy \right) dx$$

сначала как повторный интеграл по y и x, а потом в полярных координатах.

8.9. (a) Вычислите повторный интеграл

$$\int_0^1 \left(\int_0^{1/v} u^5 v^9\, du \right) dv.$$

(b) Интерпретируйте этот интеграл как двойной по соответствующей области на плоскости uv. Нарисуйте эту область, выписав в явном виде ограничивающие кривые. (Не беспокойтесь здесь и в дальнейшем, что область не ограничена[8].)

(c) Полученный двойной интеграл еще раз перепишите как повторный с другим порядком интегрирования и вычислите его.

(d) Выполните в двойном интеграле замену переменных $u = x^2 y^{-3}$ и $v = x^{-1} y^2$. Получите новый двойной интеграл на плоскости xy. Нарисуйте область интегрирования для этого нового интеграла.

(e) Преобразуйте полученный двойной интеграл в повторный и вычислите его.

[8]Здесь появляется *несобственный* двойной интеграл. Определение этого понятия и изучение свойств несобственных интегралов составляет самостоятельный раздел математического анализа. — *Прим. ред.*

(f) Покажите, что x и y являются дифференцируемыми функциями в первой четверти плоскости uv.

8.10. (a) Вычислите интеграл

$$\iint_Q y\,dA,$$

где Q — первая четверть круга единичного радиуса. Этот интеграл преобразуйте в повторный по переменным x и y. (A обозначает обычную площадь на плоскости xy.)

(b) Введите полярные координаты r и θ обычным способом и преобразуйте данный интеграл в интеграл по соответствующей области на плоскости $r\theta$. Дайте описание этой области. (Под плоскостью $r\theta$ подразумевается новое пространство $\mathbb{R}^2$, где декартовыми координатами являются r и θ.)

(c) Преобразуйте полученный интеграл в повторный в переменных r, θ и вычислите его еще раз.

8.11. (a) Вычислите интеграл $I = \int_W 2y\,dx \wedge dy$ как сумму двух повторных интегралов на плоскости xy. Область W (рис. 8.45) ограничена прямыми $y = x/2$, $y = 2x$ и гиперболами $xy = 2$, $xy = 8$.

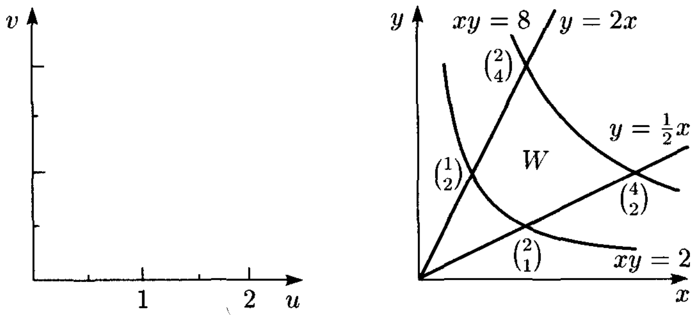


Рис. 8.45

(b) На плоскости uv найдите такой прямоугольник R, чтобы W была образом R после преобразования α, заданного формулами: $\alpha^*x = 2uv$, $\alpha^*y = u/v$.

(c) Вычислите $\alpha^*(2y\,dx \wedge dy)$.

(d) Вычислите интеграл I как интеграл по области R.

8.12. (a) Вычислите криволинейный интеграл $\int[(2y^2+3x)\,dx+2xy\,dy]$ по кривой γ, изображенной на рис. 8.46, которая состоит из прямолинейных отрезков $0 \le x \le 2$, $0 \le y \le 2$ и дуги окружности $x^2+y^2=4$ при $x \ge 0$ и $y \ge 0$.

(b) Постройте такой двойной интеграл по области, ограниченной кривой γ, чтобы он был равен криволинейному интегралу, вычисленному в предыдущем пункте. Вычислите этот двойной интеграл, преобразуя его к полярным координатам.

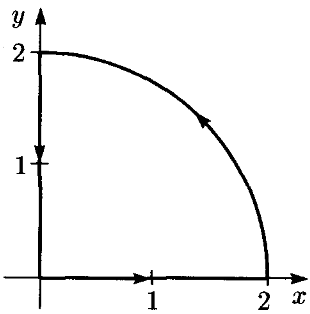


Рис. 8.46

(c) Найдите функцию $f(x)$, обладающую свойством

$$\int f(x)[(2y^2+3x)\,dx+2xy\,dy]=0,$$

когда интеграл вычисляется по *любой* замкнутой кривой на плоскости.

8.13. Один из способов замены переменных в ориентированном интеграле $I=\int_W f\,dx\wedge dy$, где $W=\phi(S)$, состоит в том, чтобы воспользоваться формулой Грина и преобразовать интеграл I в криволинейный интеграл по замкнутому пути ∂W. Затем преобразовать результат в криволинейный интеграл на плоскости uv, еще раз применить формулу Грина и выразить I как двойной интеграл на плоскости uv. Воспользовавшись этим подходом, получите формулу замены переменных для двойного интеграла.

8.14. Пусть u и v будут функциями на плоскости, у которых первые и вторые производные по x и y непрерывны. Пусть S — связная область на плоскости, границу которой обозначим ∂S. Покажите, что

$$\int_{\partial S}\left[\left(v\frac{\partial u}{\partial x}-u\frac{\partial v}{\partial x}\right)dx+\left(u\frac{\partial v}{\partial y}-v\frac{\partial u}{\partial y}\right)dy\right]$$
$$=2\iint_S\left(u\frac{\partial^2 v}{\partial x\partial y}-v\frac{\partial^2 v}{\partial x\partial y}\right)dx\,dy.$$

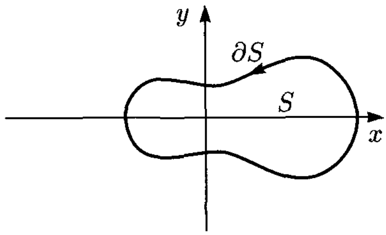


Рис. 8.47

5. (a) Тонкая пластина S (рис. 8.47) имеет постоянную плотность σ (здесь σ — масса на единицу площади). Момент инерции пластины относительно оси x равен

$$I_x=\sigma\int_S y^2\,dx\,dy.$$

Покажите, что $I_x=\sigma\int_{\partial S}xy^2\,dy$, где ∂S — граница пластины, проходимая против часовой стрелки.

(b) Момент инерции пластины относительно оси y равен

$$I_y=\sigma\int_S x^2\,dx\,dy.$$

Найдите две дифференциальные формы ω и τ, для которых справедливо равенство

$$I_y=\sigma\int_{\partial S}\omega=\sigma\int_{\partial S}\tau.$$

(c) Момент инерции пластины относительно оси z равен

$$I_z=\sigma\int_S(x^2+y^2)\,dx\,dy=I_x+I_y.$$

Найдите такую дифференциальную форму Ω, для которой выполняется равенство $I_z = \sigma \int_{\partial S} \Omega$. Форму Ω запишите в полярных координатах r и θ.

Глава 9

Гауссова оптика

В главе 9 дан пример использования результатов первых восьми глав в физической теории — оптике. Она состоит целиком из приложений, и ее пропуск при чтении никак не отразится на понимании последующего материала.

9.1. Оптические теории

В истории физики часто случается, что старая теория вытесняется новой, но при этом старая остается справедливой, становясь либо приближением к новой, которое пригодно при определенных ограничениях, либо частным случаем новой теории. Так, ньютоновская механика может рассматриваться как приближение к релятивистской механике, справедливое при скоростях, малых по сравнению со скоростью света. Аналогично, ньютоновская механика может рассматриваться как приближение к квантовой механике, когда изучаемые объекты достаточно велики. Законы Кеплера, описывающие движение планет, являются частным случаем законов Ньютона, когда сила взаимодействия между телами обратно пропорциональна квадрату расстояния между ними. Законы Кеплера можно также считать приближением к уравнениям движения Ньютона, когда мы пренебрегаем влиянием одной планеты на движение других.

распространения луча. Это явление называется поляризацией света, и его можно наблюдать, например, используя специальные поляроидные фильтры.

Геометрическая оптика является приближением к волновой оптике, в котором пренебрегают волновой природой света. Это можно делать, когда размеры апертур велики по сравнению с длиной волны света и когда мы не интересуемся тем, что происходит в фокусах и в окрестности тени.

Линейная оптика является приближением к геометрической оптике, когда исследуются малые углы. В рамках линейной оптики делаются приближения: $\sin\theta \doteq \theta$, $\tan\theta \doteq \theta$, $\cos\theta \doteq 1$ и т. д., т. е. отбрасываются все выражения, квадратичные (и более высоких порядков) по углам. Например, в геометрической оптике справедлив закон Снеллиуса, согласно которому, если свет проходит область с коэффициентом преломления n (относительно вакуума) и попадает в область с коэффициентом преломления n', то $n \sin i = n' \sin i'$, где i и i' — углы, которые составляет направление распространения луча с нормалью к поверхности, разделяющей эти области (рис. 9.3). В линейной оптике этот закон упрощается: $ni = n'i'$, потому что углы i и i' малы. (Этот приближенный закон был известен еще Птолемею.)

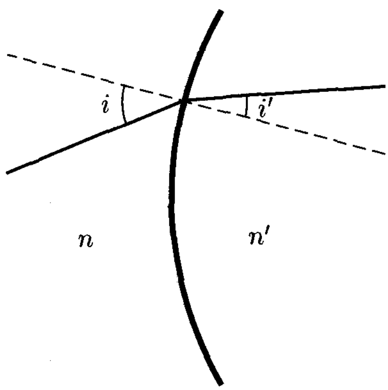


Рис. 9.3

Различие между геометрической оптикой и ее линейным приближением называется (геометрической) аберрацией. Например,

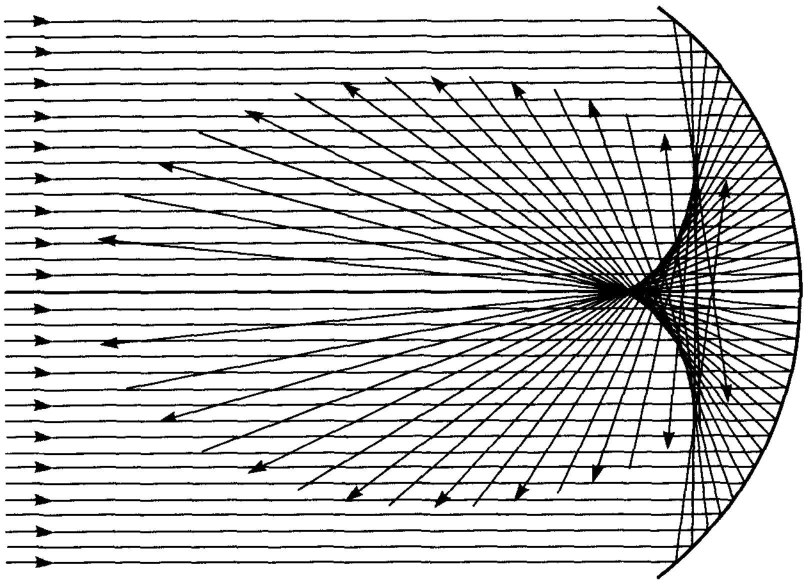


Рис. 9.4. Сферическая аберрация

пусть пучок параллельных лучей падает на сферическое зеркало (рис. 9.4). Тщательное наблюдение отраженных лучей показывает, что не все они пересекаются в одной точке. Только лучи, распространяющиеся вблизи диаметра, пересекаются вблизи фокуса[1] зеркала. В рамках линейной оптики мы ограничиваемся рассмотрением лучей, которые близки к диаметру, так что можно считать, что они пересекаются в фокусе. (Это отклонение от фокуса сферического зеркала называется сферической аберрацией.)

Гауссова оптика является частным случаем линейной оптики, когда предполагается, что все рассматриваемые поверхности обладают вращательной симметрией относительно некоторой центральной оси. Это очень важный частный случай, потому что все линзы и большинство полированных зеркал обладают этим свойством. Наше обсуждение оптических теорий изображено на схеме:

[1] «Фокус» в данном случае расположен в середине радиуса сферы, параллельного направлению лучей. — *Прим. ред.*

Гауссова оптика

частный случай при наличии вращательной симметрии

Линейная оптика

приближение, пренебрегающее аберрацией

Геометрическая оптика

приближение коротких волн, пренебрегающее дифракцией, интерференцией и поляризацией

Волновая оптика

частный случай, когда не рассматривают излучение света

Электродинамика Максвелла

приближение, пренебрегающее квантовыми эффектами

Квантовая электродинамика

9.2. Матричные методы

В рамках гауссовой оптики мы изучаем траекторию светового луча, когда он проходит через различные преломляющие поверхности оптической системы (или отражается от отражающих поверхностей). Введем систему координат, в которой ось z (направленная слева направо) совпадает с оптической осью (осью симметрии системы). Мы ограничимся исследованием коаксиальных лучей — лучей, лежащих в одной плоскости с оптической осью[2].

В силу вращательной симметрии очевидно, что можно ограничиться изучением лучей, лежащих в одной выбранной плоскости. Траектория луча, проходящего через систему преломляющих поверхностей, будет представлять собой последовательность прямых линий. Наша задача состоит в том, чтобы связать прямолинейный луч, выходящий из системы, с прямолинейным лучом, падающим на нее. Для этого мы должны получить метод описания прямых линий. Выберем фиксированное значение $z = z_0$ и плоскость, перпендикулярную оптической оси, которую будем называть плоскостью отсчета. Тогда прямая линия определяется двумя параметрами: высотой q над осью z и углом θ, определяющим наклон линии к оптической оси (рис. 9.5). Угол θ будем измерять в радианах и считать положительным, если вращение против часовой стрелки совмещает положительное направление оси z с направлением луча. На каждой стадии вычислений следует выбирать удобную плоскость отсчета.

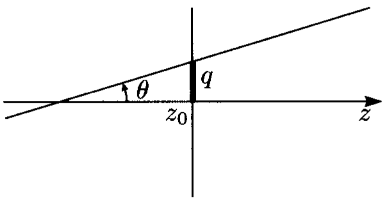


Рис. 9.5

Например, если свет падает на оптическую систему слева и выходит направо (рис. 9.6), мы должны выбрать одну плоскость отсчета z_1 слева от системы линз и вторую плоскость отсчета z_2 справа. Луч, входящий в систему по прямой линии с параметрами q_1 и θ_1 при $z = z_1$, выходит из системы по прямой линии с параметрами q_2, θ_2 при $z = z_2$. Наша задача состоит в том, чтобы для любой системы линз найти связь между (q_2, θ_2) и (q_1, θ_1).

[2]Хотя это и является упрощающим предположением, на самом деле можно доказать, что линейность означает изучение произвольных лучей, потому что их можно свести к коаксиальным лучам, проектируя на два взаимно перпендикулярных направления.

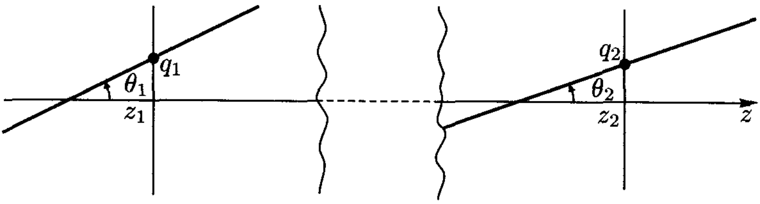


Рис. 9.6

Теперь надо сделать очень простой, но чрезвычайно важный шаг — главный в геометрической оптике и механике.

Переменную θ заменим на $p = n\theta$, где n обозначает показатель преломления среды в плоскости отсчета. (В механике это соответствует замене скорости на импульс.)

Тогда световой луч может быть описан вектором $\begin{pmatrix} q \\ p \end{pmatrix}$, и мы должны найти вектор $\begin{pmatrix} q_2 \\ p_2 \end{pmatrix}$ как функцию $\begin{pmatrix} q_1 \\ p_1 \end{pmatrix}$. Мы пренебрегаем квадратичными членами и членами более высоких порядков, следовательно, в нашем приближении $\begin{pmatrix} q_2 \\ p_2 \end{pmatrix}$ будет линейной функцией от $\begin{pmatrix} q_1 \\ p_1 \end{pmatrix}$, т. е.

$$\boxed{\begin{pmatrix} q_2 \\ p_2 \end{pmatrix} = M_{21} \begin{pmatrix} q_1 \\ p_1 \end{pmatrix},}$$

где M_{21} — некоторая матрица. Главным результатом выбора p вместо переменной θ является

$$\boxed{\operatorname{Det} M_{21} = 1.}$$

Другими словами, изучение гауссовой оптики сводится к изучению группы вещественных матриц 2×2, имеющих определитель, равный единице, т. е. группы $S1(2, \mathbb{R})$. Чтобы доказать это утверждение, заметим, что если луч последовательно проходит три плоскости отсчета z_1, z_2 и z_3, то по определению матрица перехода

$$\boxed{M_{31} = M_{32} M_{21}.}$$

Таким образом, если оптическая система состоит из двух или более блоков, мы должны только проверить, что $\operatorname{Det} M = 1$ в отдельности для каждого блока. Для упрощения изложения предположим, что в системе нет зеркал.

Основные процессы в оптике

Прохождение света через систему линз можно рассматривать как совокупность основных оптических процессов.

(a) Распространение (трансляция): луч движется по прямой между двумя плоскостями отсчета, лежащими в одной и той же среде. Для описания этого процесса расстояние между плоскостями обозначим t, показатель преломления среды n. Очевидно, что при таком движении светового луча угол θ, и, следовательно, p, не изменяются: $(p_2 = p_1)$, и $q_2 = q_1 + (t/n)p_1$ (рис. 9.7).

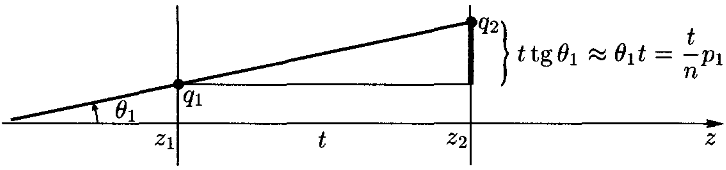


Рис. 9.7

Введем приведенное расстояние $T = t/n$. Тогда

$$\begin{pmatrix} q_2 \\ p_2 \end{pmatrix} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} q_1 \\ p_1 \end{pmatrix}, \qquad \operatorname{Det} \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} = 1.$$

(b) Преломление на границе между двумя областями с разными показателями преломления. Показатели преломления двух сред обозначим n_1 и n_2. Естественно, надо учитывать кривизну поверхности раздела сред. Выбираются две плоскости отсчета, слева и справа вблизи от этой поверхности. На поверхности раздела величина q не изменяется, а вот угол и, следовательно, p изменяются согласно закону Снеллиуса (его линеаризированному варианту). В закон Снеллиуса входит наклон касательной к поверхности в точке, где происходит преломление. В нашем приближении мы пренебрегаем квадратичными членами в этом наклоне и членами третьего и более высокого порядка в уравнении поверхности. Таким образом, мы предполагаем, что линия пересечения этой поверхности с нашей плоскостью будет параболой:

$$z - z_1 = \frac{1}{2}kq^2.$$

Тогда производная z по q, т. е. $z'(q) = kq$, будет $\tan(\pi/2 - \psi)$, где ψ — угол, изображенный на рис. 9.8.

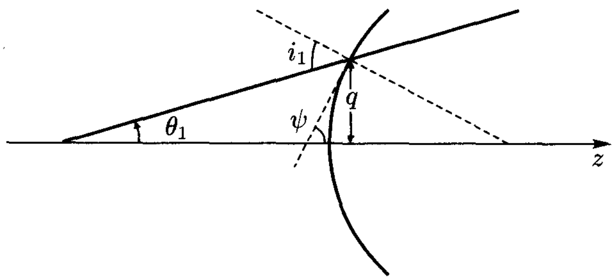


Рис. 9.8

Для малых углов θ, т. е. для малых значений q, угол ψ будет близок к $\pi/2$, и тогда $\tan(\pi/2 - \psi)$ можно заменить на $\pi/2 - \psi$, если мы готовы опустить члены более высокого порядка по q и p. Таким образом, в гауссовом приближении $\pi/2 - \psi = kq$. Угол между падающим лучом и касательной равен $(\pi/2 - i_1)$. Из того, что сумма внутренних углов треугольника равна π, следует, что $(\pi - \psi) + \theta_1 + (\pi/2 - i_1) = \pi$ или

$$i_1 = \theta_1 + kq.$$

Аналогично получаем

$$i_2 = \theta_2 + kq,$$

где $q = q_1 = q_2$ определяют точку на преломляющей поверхности, в которую падает луч. Первое уравнение умножим на n_1, а второе на n_2. Используя приближенную формулу для закона Снеллиуса $n_1 i_1 = n_2 i_2$, получаем

$$\begin{pmatrix} q_2 \\ p_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -P & 1 \end{pmatrix} \begin{pmatrix} q_1 \\ p_1 \end{pmatrix},$$

константа $P = (n_2 - n_1) \cdot k$ называется *оптической силой* преломляющей поверхности.

Сопряженные плоскости

Итак, движение луча в любой гауссовой оптической системе между двумя плоскостями отсчета соответствует матрице

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \quad \text{где} \quad AD - BC = 1.$$

Теперь можно составить словарь, переводящий свойства матрицы в оптические свойства.

Например, две плоскости z_1 и z_2 называются *сопряженными*, если для любого значения q_1 все лучи, выходящие из q_1, z_1, сходятся в точку q_2, z_2. Это значит, что значение q_2 не должно зависеть от p_1, т. е. что

$$B = 0.$$

Тонкие линзы

Заметим, что произведение двух матриц вида $\begin{pmatrix} 1 & 0 \\ -P & 1 \end{pmatrix}$ имеет тот же вид, т. е.

$$\begin{pmatrix} 1 & 0 \\ -P_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -P_2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -(P_1 + P_2) & 1 \end{pmatrix}.$$

Это позволяет написать уравнение для так называемых *тонких линз*, состоящих из преломляющих поверхностей с пренебрежимо малым расстоянием между ними (рис. 9.9)[3]. В этом случае плоскости отсчета z_1 и z_2 удобно брать совпадающими с плоскостью линз. При этом с плоскостью z_1 связывают лучи, падающие слева, а с плоскостью z_2 — лучи, выходящие из линз и распространяющиеся направо.

---

[3]На рис. 9.9 преломляющие поверхности изображены не параболами, как делалось выше, а дугами окружностей. Можно доказать, что уравнение окружности радиуса R с центром на оси z справа от точки z_1, в которой эта окружность пересекает ось z, с точностью до $o(q^2)$ имеет вид $z - z_1 = \dfrac{q^2}{2R}$. Поэтому константа k из предыдущего построения в этом случае равна $1/R$. Заметим, что число $k = 1/R$ называют *кривизной* окружности радиуса R. — *Прим. ред.*

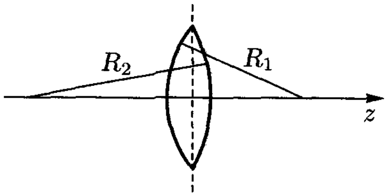


Рис. 9.9

Матрица для левой преломляющей поверхности равна

$$\begin{pmatrix} 1 & 0 \\ -\dfrac{n_2-n_1}{R_1} & 1 \end{pmatrix}.$$

Соответственно, матрица для правой преломляющей поверхности равна

$$\begin{pmatrix} 1 & 0 \\ -\dfrac{n_1-n_2}{R_2} & 1 \end{pmatrix}.$$

(Заметим, что на рис. 9.9 значение R_2 отрицательно.) Перемножив эти матрицы, мы получаем матрицу для тонкой линзы

$$\begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix}, \quad \text{где} \quad \frac{1}{f} = (n_2-n_1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right).$$

Число f называется *фокусным расстоянием линзы*. Мы будем предполагать, что линзы находятся в вакууме, так что $n_1 = 1$ и $n_2 > 1$. Если R_1 положительно, а R_2 отрицательно (двояковыпуклая линза), то фокусное расстояние f положительно, поскольку $n_2 - n_1 > 0$.

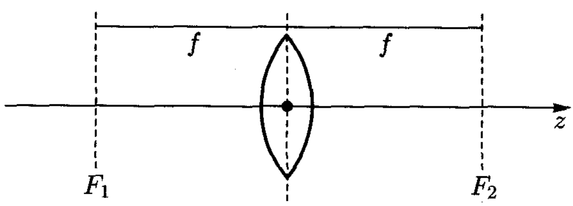


Рис. 9.10

Вычислим матрицу для тонкой линзы, задающую движение луча между плоскостью отсчета F_1, расположенной слева на расстоянии f от линзы, и плоскостью отсчета F_2, расположенной справа на расстоянии f от нее (рис. 9.10):

$$\begin{pmatrix} 1 & f \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix} \begin{pmatrix} 1 & f \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & f \\ -1/f & 0 \end{pmatrix}.$$

Плоскость F_1 называется *первой фокальной плоскостью*. Если луч, падающий на линзу, проходит через эту плоскость в точке $q_1 = 0$ под углом p_1, то выходящий луч имеет параметры

$$\begin{pmatrix} q_2 \\ p_2 \end{pmatrix} = \begin{pmatrix} 0 & f \\ -1/f & 0 \end{pmatrix} \begin{pmatrix} 0 \\ p_1 \end{pmatrix} = \begin{pmatrix} fp_1 \\ 0 \end{pmatrix},$$

т. е. выходящий луч параллелен оси z. Обратное утверждение: если падающий луч параллелен оси z, то выходящий луч имеет параметры

$$\begin{pmatrix} q_2 \\ p_2 \end{pmatrix} = \begin{pmatrix} 0 & f \\ -1/f & 0 \end{pmatrix} \begin{pmatrix} q_1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -q_1/f \end{pmatrix},$$

т. е. пересекает ось z во второй фокальной плоскости. В общем случае, можно видеть, что p_2 не зависит от p_1, следовательно, лучи, проходящие через фиксированную точку в первой фокальной плоскости, выходят из линзы параллельным пучком под одним и тем же углом к оси z. Далее, q_2 не зависит от q_1, т. е. все параллельные друг другу до линзы лучи проходят через одну точку во второй фокальной плоскости.

Рассмотрим простой пример использования матричного метода фокусировки. Пусть плоскость отсчета z_1 находится слева от тонкой линзы на расстоянии s_1, а плоскость отсчета z_2 находится справа от линзы на расстоянии s_2. Матрица, задающая движение луча между этими плоскостями, такова:

$$\begin{pmatrix} 1 & s_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix} \begin{pmatrix} 1 & s_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 - s_2/f & s_2 + s_1 - s_1 s_2/f \\ -1/f & 1 - s_1/f \end{pmatrix}.$$

Плоскости будут сопряженными, если верхний правый матричный элемент равен нулю. Мы получаем хорошо известное уравнение тонкой линзы $1/s_1 + 1/s_2 = 1/f$. Запишем его в виде

$$s_1 + s_2 - P s_1 s_2 = 0, \quad \text{где} \quad P = 1/f.$$

Уравнение можно решить относительно s_2 при условии, что $s_1 \neq 1/P$. Таким образом, кроме случая $s_1 = f$, любая плоскость имеет единственную сопряженную плоскость. Если же $s_1 = f$, т.е. z_1 совпадает с первой фокальной плоскостью, то все лучи, проходящие в этой плоскости через фиксированную точку q, выходят из линзы параллельным пучком. Значит, плоскость, сопряженная первой фокальной плоскости, «лежит в бесконечности». Аналогичные рассуждения можно провести для второй фокальной плоскости.

Для $s_1 \neq f$ и значения s_2, соответствующего сопряженной плоскости, увеличение дается уравнением

$$\frac{q_2}{q_1} = 1 - \frac{s_2}{f} = 1 - s_2\left(\frac{1}{s_1} + \frac{1}{s_2}\right) = -\frac{s_2}{s_1}.$$

Если s_1 и s_2 положительны (объект находится слева от линзы, а изображение справа), то увеличение отрицательно. Это значит, что изображение перевернуто.

Умножая матрицы, можно построить матрицу для любой комбинации тонких линз. Например, пусть есть две линзы с фокусными расстояниями f_1 и f_2, расположенными на расстоянии l друг от друга (рис. 9.11). В этом случае для плоскостей отсчета z_1 (первая линза) и z_2 (вторая линза) получается матрица

$$\begin{pmatrix} 1 & 0 \\ -1/f_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f_1 & 1 \end{pmatrix} = \begin{pmatrix} 1 - l/f_1 & l \\ l/f_1 f_2 - 1/f_2 - 1/f_1 & 1 - l/f_2 \end{pmatrix}.$$

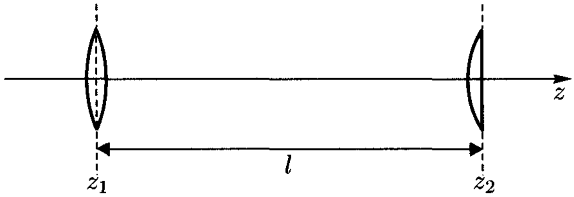


Рис. 9.11

Телескоп

Очень интересная ситуация возникает с парой линз, когда $l = f_1 + f_2$. В этом случае матрица имеет вид

$$\begin{pmatrix} A & B \\ 0 & D \end{pmatrix},$$

т. е. $C = 0$. Это значит, что $p_2 = Dp_1$, т. е. направление выходящего луча зависит только от направления падающего луча. Это условие выполняется в конструкции, которая называется *астрономическим телескопом*. Он состоит из линзы (*объектива*) с большим положительным фокусным расстоянием f_1 и *окуляра* с малым положительным фокусным расстоянием f_2, расположенных на расстоянии $f_1 + f_2$ друг от друга (рис. 9.12). Такой телескоп преобразует параллельные лучи, идущие от далекой звезды, в параллельные лучи, попадающие в глаз наблюдателя.

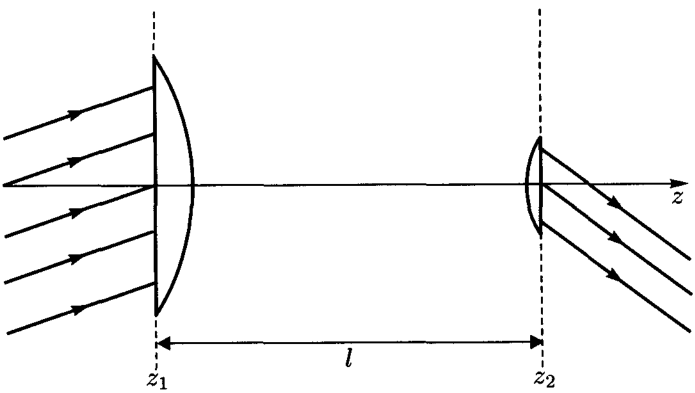


Рис. 9.12

Угловое увеличение телескопа равно отношению наклона *выходящего* луча к наклону *падающего луча*, что равняется

$$D = 1 - \frac{1}{f_2} = 1 - \frac{f_1 + f_2}{f_2} = -\frac{f_1}{f_2}.$$

Это увеличение отрицательно (образ перевернут), и его величина равна отношению фокусного расстояния объектива к фокусному расстоянию окуляра.

Общий случай

Здесь мы хотим показать, что *любая* матрица $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ с $\operatorname{Det} M = 1$ описывает некоторую оптическую систему.

Пусть сначала $C \neq 0$. В этом случае матрицу M можно представить в виде произведения трех оптических матриц:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ C & 1 \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, \tag{9.1}$$

Непосредственной проверкой можно убедиться, что это равенство выполняется при $s = \dfrac{1-A}{C}$, $t = \dfrac{1-D}{C}$, и только при этих s, t.

Если $C = 0$ (при этом $A \neq 0$), то, домножив матрицу M слева на $\begin{pmatrix} 1 & 0 \\ P & 1 \end{pmatrix}$ (берем $P \neq 0$), получим матрицу

$$M' = \begin{pmatrix} 1 & 0 \\ P & 1 \end{pmatrix} \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \begin{pmatrix} A & B \\ PA & PB + D \end{pmatrix},$$

у которой левый нижний элемент — не нуль, поэтому M' — оптическая матрица, а поскольку $M = \begin{pmatrix} 1 & 0 \\ -P & 1 \end{pmatrix} M'$, то M — тоже оптическая матрица.

Гауссово разложение

Из формулы (9.1) вытекает, что для любой матрицы $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ с $\operatorname{Det} M = 1$ и $C \neq 0$ существуют и единственны числа s, t такие, что

$$\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ C & 1 \end{pmatrix}.$$

Таким образом, для любой не телескопической оптической системы существуют такие две плоскости, что матрица, описывающая движение луча между ними, имеет вид $\begin{pmatrix} 1 & 0 \\ C & 1 \end{pmatrix}$. Эти плоскости сопряжены между собой, и соответствующее увеличение равно 1. Гаусс назвал их *главными плоскостями*. Если начать с оптической матрицы $\begin{pmatrix} 1 & 0 \\ C & 1 \end{pmatrix}$ между двумя главными плоскостями, то так же, как и в случае тонких линз, для любой плоскости можно найти сопряженную. Для этого достаточно написать $C = -P = -1/f$. Например, справа и слева от главных плоскостей находятся фокальные плоскости с фокусным расстоянием f:

$$\begin{pmatrix} 1 & f \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix} \begin{pmatrix} 1 & f \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & f \\ -1/f & 0 \end{pmatrix}.$$

Гаусс так интерпретировал этот результат.

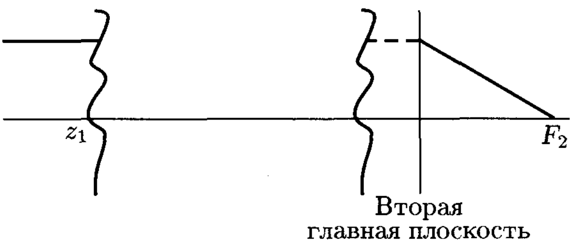


Рис. 9.13

Пусть луч $\begin{pmatrix} q \\ 0 \end{pmatrix}$, параллельный оси, входит в оптическую систему при $z = z_1$ (рис. 9.13). Этот луч приходит ко второй главной плоскости на той же высоте, но с наклоном $-q/f$ (вектор $\begin{pmatrix} q \\ -q/f \end{pmatrix}$) и фокусируется на оси во второй фокальной точке F_2. Аналогично, луч, выходящий из первой фокальной точки с любым наклоном в первой главной плоскости, преобразуется в луч, распространяющийся параллельно оси, и приходит в z_2 на той же высоте, какую он имел в первой главной плоскости.

Итак, мы видим, что произвольная не телескопическая оптическая система может быть описана тремя параметрами: положением двух главных плоскостей и фокусным расстоянием. (Мы уже знали, что должно быть три независимых параметра — три элемента матрицы 2×2, потому что четвертый матричный элемент определяется из условия, что определитель матрицы равен 1.)

Если мы знаем положение главных плоскостей, то положение фокальных плоскостей мы получаем из равенств

$$H_1 - F_1 = f, \qquad F_2 - H_2 = f.$$

Если фокальные плоскости рассматриваются как плоскости отсчета в оптической системе, то из определения фокальной плоскости мы знаем, что оптическая матрица должна иметь нуль в верхнем левом и в нижнем правом углах. Таким образом, матрица распространения луча между двумя фокальными плоскостями имеет вид

$$\begin{pmatrix} 0 & f \\ -1/f & 0 \end{pmatrix}.$$

Теперь возьмем (рис. 9.14) две другие плоскости y_1 и y_2, связанные с фокальными плоскостями соотношениями

$$F_1 - y_1 = n_1 x_1 \quad \text{и} \quad y_2 - F_2 = n_2 x_2.$$

Рис. 9.14

Матрица, соответствующая этим плоскостям, будет равна

$$\begin{pmatrix} 1 & x_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & f \\ -1/f & 0 \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -x_2/f & f - (x_1 x_2/f) \\ -1/f & -x_1/f \end{pmatrix}.$$

Отсюда видно, что плоскости y_1 и y_2 сопряжены тогда и только тогда, когда $x_1 x_2 = f^2$ (это уравнение называется *уравнением Ньютона*). Тогда увеличение дается формулой

$$m = -x_2/f = -f/x_1.$$

Например, пусть y_2 находится справа от F_2, так что x_2 положительно. Тогда, если фокусное расстояние f положительно, то увеличение m отрицательно, т. е. изображение оказывается перевернутым.

Давайте перечислим полученные результаты. Пусть $Sl(2, \mathbb{R})$ обозначает группу матриц 2×2 с определителем, равным 1. В этом разделе мы показали изоморфизм между $Sl(2, \mathbb{R})$ и системами гауссовой оптики. Каждая матрица соответствует оптической системе. Умножение матриц соответствует композиции оптических систем.

А теперь перейдем к изложению идей Гамильтона в самой простой форме.

9.3. Метод Гамильтона в гауссовой оптике

Предположим, что в нашей оптической системе z_1 и z_2 — не сопряженные плоскости. Это значит, что элемент B в оптической матрице не равен нулю. Тогда из уравнений

$$q_2 = Aq_1 + Bp_1,$$
$$p_2 = Cq_1 + Dp_1$$

мы можем выразить p_1 и p_2 через q_1 и q_2:

$$p_1 = (1/B)(q_2 - Aq_1),$$
$$p_2 = (1/B)(Dq_2 - q_1),$$

где мы воспользовались условием $AD - BC = 1$. Этот результат можно интерпретировать геометрически. Выбираем точку q_1 на плоскости z_1 и точку q_2 на плоскости z_2. Существует единственный луч, соединяющий эти точки. (Для сопряженных плоскостей это не так. Если q_2 — изображение точки q_1, то имеется бесконечное число лучей, соединяющих точки q_1 и q_2, т. е. все лучи, выходящие из точки q_1, попадают в точку q_2. Если q_2 не является изображением точки q_1, то лучей, соединяющих точки q_1 и q_2, не существует.) Пусть функция $W = W(q_1, q_2)$ имеет вид

$$W(q_1, q_2) = (1/2B)(Aq_1^2 + Dq_2^2 - 2q_1q_2) + K,$$

где K — постоянная. Тогда p_1 и p_2 выражаются в форме

$$p_1 = -(\partial W/\partial q_1) \quad \text{и} \quad p_2 = (\partial W/\partial q_2).$$

Гамильтон назвал эту функцию *точечной характеристикой* системы. В современной физической литературе иногда ее называют *эйконалом.* Предположим, что мы выбрали плоскости z_1, z_2 и z_3, причем ни одна пара плоскостей не является сопряженной. Пусть $z_1 < z_2 < z_3$ и z_2 не совпадает с преломляющей поверхностью. Обозначим W_{21} точечную характеристику системы $z_1 - z_2$, а W_{32} — системы $z_2 - z_3$. Утверждается, что точечная характеристика системы $z_1 - z_3$ (с точностью до произвольной аддитивной постоянной) дается формулой

$$W_{31}(q_1, q_3) = W_{21}(q_1, q_2) + W_{32}(q_2, q_3),$$

причем луч, распространяющийся из точки q_1 в точку q_3, проходит через точку $q_2 = q_2(q_1, q_3)$ на плоскости z_2.

Чтобы убедиться в справедливости этой формулы, заметим, что раз плоскость z_2 не совпадает с преломляющей поверхностью, то она не изменяет направление луча. Тогда

$$p_2 = (\partial W_{21}/\partial q_2)(q_1, q_2) = -(\partial W_{32}/\partial q_2)(q_2, q_3).$$

А теперь применим цепное правило и получим, что

$$\partial W_{31}/\partial q_1 = -p_1 \quad \text{и} \quad \partial W_{31}/\partial q_3 = p_3.$$

Итак, функция W определена с точностью до аддитивной постоянной. Гамильтон показал, что при соответствующем выборе этой постоянной можно сделать так, что $W(q_1, q_2)$ будет *оптической длиной луча, соединяющего точки* q_1 *и* q_2.

Понятие *оптической длины* определяется следующим образом. Для отрезка пути l в среде с постоянным коэффициентом преломления n оптическая длина пути равна nl. Пусть *путь* γ является ломаной линией, где каждый отрезок l_i лежит в среде с постоянным коэффициентом преломления n_i. Тогда оптическая длина пути γ равна

$$L(\gamma) = \sum n_i l_i.$$

Докажем теорему Гамильтона в рамках гауссовой оптики. В этом приближении в производных функции W сохраняются только члены, линейные по p и q. Таким образом, для вычисления оптической длины в функции W мы должны учитывать члены не выше квадратичных. Итак, световой луч, падающий на плоскость z_1, определяется параметрами $\begin{pmatrix} q_1 \\ p_1 \end{pmatrix}$, а луч, выходящий из плоскости z_2, — параметрами $\begin{pmatrix} q_2 \\ p_2 \end{pmatrix}$. Тогда оптическая длина пути

$$L(\gamma) = L_o + \frac{1}{2}(p_2q_2 - p_1q_1),$$

где L_o обозначает оптическую длину вдоль оси системы между плоскостями z_1 и z_2 ($p_1 = q_1 = 0 = p_2 = q_2$). Если мы докажем эту формулу, то предположив, что z_1 и z_2 не являются сопряженными плоскостями, значения p_1 и p_2 можно выразить через q_1 и q_2, т. е. подстановка $p_1 = (1/B)(q_2 - Aq_1)$ и $p_2 = (1/B)(Dq_2 - q_1)$ в $L(\gamma)$ даст выражение для W, если принять $K = L_o$.

Прежде чем доказывать формулу для $L(\gamma)$, заметим, что она правильно себя ведет для случая комбинированной системы: если мы взяли z_1, z_2 и z_3, то очевидно, что длина вдоль оси складывается:

$$\frac{1}{2}(p_2q_2 - p_1q_1) + \frac{1}{2}(p_3q_3 - p_2q_2) = \frac{1}{2}(p_3q_3 - p_1q_1).$$

Следовательно, мы должны доказать формулу в двух фундаментальных случаях.

(1) Если коэффициент преломления n постоянен, d — проекция луча на ось z, то в этом случае

$$\begin{aligned} L(\gamma) &= n(d^2 + (q_2 - q_1)^2)^{1/2} \\ &\doteq nd + \frac{1}{2}\frac{n}{d}(q_2 - q_1)^2 \\ &= nd + \frac{1}{2}\left[\frac{n}{d}(q_2 - q_1)\right](q_2 - q_1) \\ &= nd + \frac{1}{2}p(q_2 - q_1), \end{aligned}$$

где $p_2 = p_1 \doteq p = \dfrac{n}{d}(q_2 - q_1)$.

(2) Пусть теперь коэффициент преломления слева от преломляющей поверхности $z' - z = \frac{1}{2}kq^2$ равен n_1, а справа n_2 (рис. 9.15).

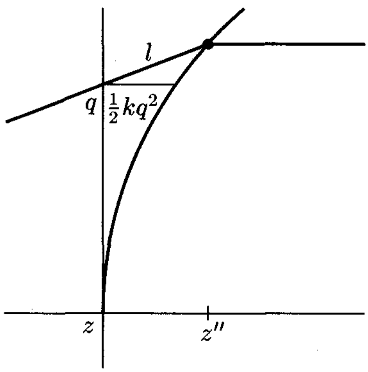


Рис. 9.15

Предположим, что слева от преломляющей поверхности выбрана точка z_3, а справа z_4. Если бы n_1 равнялось n_2, то оптическая длина между этими точками была бы равна $n_1 l_3 + n_2(l + l_4)$, где l_3 — длина отрезка слева от выбранной плоскости, а $l + l_4$ — длина отрезка справа, причем l_4 — отрезок справа от преломляющей поверхности. (Мы нарисовали картину для $k > 0$, но то же самое справедливо и для $k < 0$.) Если $n_2 \neq n_1$, то $n_2 l_4$ будет другим, но его можно вычислить по формуле случая (1) между точками z и z_4. Кроме того, влияние преломляющей поверхности приведет к замене $n_2 l$ на $n_1 l$ в выражении для длины пути, и вклад преломляющей поверхности будет равен

$$(n_1 - n_2)l.$$

Итак,

$$l = (z'' - z)\operatorname{cosec}\theta_1,$$

где z'' находится из системы уравнений

$$z'' - z = \frac{1}{2}kq''^2$$
$$q'' = (\tan\theta_1)(z'' - z) + q.$$

Очевидно, что с точностью до членов более высокого порядка мы можем взять $z'' = z' = \frac{1}{2}kq^2 + z$ и $\operatorname{cosec}\theta_1$ заменить на 1. Следовательно,

$$\begin{aligned}(n_1 - n_2)l &\doteq \frac{1}{2}k(n_1 - n_2)q^2 = -\frac{1}{2}pq^2 \\ &= \frac{1}{2}[k(n_1 - n_2)q]q \\ &= \frac{1}{2}(p_2 - p_1)q,\end{aligned}$$

поскольку на преломляющей поверхности $q_2 = q_1 = q$ и $p_2 = p_1 - Pq$, где $P = k(n_2 - n_1)$. Итак, наша формула доказана.

9.4. Принцип Ферма

Рассмотрим преломляющую поверхность с оптической силой $P = (n_2 - n_1)k$, расположенную в некотором z. Здесь P может обращаться в нуль. Пусть плоскость z_1 находится слева, а плоскость z_2 справа от z (рис. 9.16). Предположим, что коэффициенты преломления среды на участках от z_1 до z и от z до z_2 постоянны. Пусть q_1 — точка на плоскости z_1, q_2 — точка на плоскости z_2 и q — точка на плоскости z. Рассмотрим траекторию, состоящую из трех участков: луч от точки q_1 движется к точке q, второй

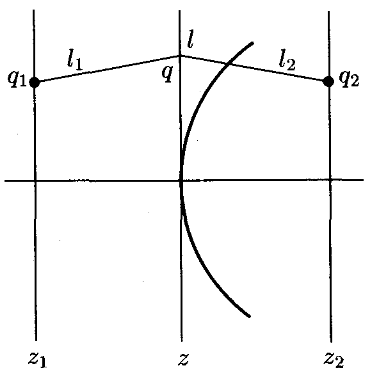


Рис. 9.16

участок — через преломляющую поверхность в точке q, и потом опять световой луч от q до точки q_2.

Поскольку q выбрана произвольно, то, вообще говоря, эта траектория не будет оптическим путем. Однако оптическая длина этой траектории

$$n_1 l_1 + nl + n_2 l_2$$

в гауссовом приближении состоит из трех членов:

$$L(q_1, q, q_2) = L_o + \frac{1}{2}(p_1(q - q_1) + p_2(q_2 - q) - Pq^2),$$

как было показано в предыдущем разделе. В этом выражении

$$p_1 = \frac{n_1}{d_1}(q - q_1) \quad \text{и} \quad p_2 = \frac{n_2}{d_2}(q_2 - q).$$

Следовательно, мы можем написать

$$L = L_o + \frac{1}{2}\left(\frac{d_1}{n_1}p_1^2 + \frac{d_2}{n_2}p_2^2 - Pq^2\right).$$

Предположим, что мы фиксировали точки q_1 и q_2. Найдем положение точки q, когда L достигает экстремума. Другими словами, мы хотим решить уравнение $\partial L/\partial q = 0$ для фиксированных значений q_1 и q_2. Вычисляя производную по q от L и учитывая, что $\partial p_1/\partial q = n_1/d_1$ и $\partial p_2/\partial q = -n_2/d_2$, получаем уравнение

$$p_1 - p_2 - Pq = 0.$$

Откуда имеем

$$p_2 = Pq + p_1.$$

Но это в точности то же самое соотношение между p_1 и p_2, которое определяется из матрицы преломления в точке z.

Итак, мы доказали следующий факт. Фиксируем точки q_1 и q_2. Рассмотрим траектории, соединяющие точки q_1 и q_2, состоящие из двух отрезков: от точки q_1 до q' и от q' до q_2. Из всех возможных траекторий световой луч распространяется вдоль той, где L достигает экстремума, т. е. выполняется условие

$$\frac{\partial L}{\partial q'} = 0.$$

Это и есть знаменитый принцип *наименьшего времени* Ферма (в гауссовом приближении). Давайте подставим

$$p_1 = \frac{n_1}{d_1}(q - q_1) \quad \text{и} \quad p_2 = \frac{n_2}{d_2}(q_2 - q)$$

в формулу для L. Получим еще одно выражение для L:

$$L = n_1 d_1 + n_2 d_2 + \frac{1}{2}\left[\frac{n_1}{d_1}(q - q_1)^2 + \frac{n_2}{d_2}(q_2 - q)^2 - Pq^2\right].$$

Коэффициент при q^2 равен $n_1/d_1 + n_2/d_2 - P$. Таким образом, оптический путь *минимален* при условии

$$\frac{n_1}{d_1} + \frac{n_2}{d_2} - P > 0$$

и *максимален* при условии

$$\frac{n_1}{d_1} + \frac{n_2}{d_2} - P < 0.$$

Если $P > 0$, то мы получаем минимум при малых значениях d_1 и d_2 и максимум при больших значениях d_1 и d_2. Ситуация становится неопределенной, когда

$$\frac{n_1}{d_1} + \frac{n_2}{d_2} = P,$$

т. е. мы не можем найти решение для q'. Это условие означает, что соответствующие плоскости являются сопряженными. Итак, мы получаем минимум, если плоскость, сопряженная к z_1, не лежит между z_1 и z_2, и максимум в противном случае. Тот факт, что L достигает минимума только до первой сопряженной точки, справедлив и для произвольной оптической системы, где он известен как *теорема коэффициентов Морзе.*

Рассмотрим интуитивно ясный пример этого явления. Пусть свет отражается от вогнутого сферического зеркала. Возьмем точку Q на оси z внутри сферы. Луч распространяется вдоль диаметра сферы и, отражаясь, попадает в точку Q. Очевидно, что расстояние от зеркала до точки Q будет локальным минимумом, если точка Q расположена между зеркалом и центром, в противном случае это расстояние будет локальным максимумом.

9.5. От гауссовой оптики к линейной

Что произойдет, если мы не будем предполагать наличие вращательной симметрии оптической системы, но по-прежнему будем учитывать только члены первого порядка по углам и расстояниям? Прежде всего для описания луча нам необходимо иметь четыре переменных: q_x и q_y — координаты точки, в которой луч пересекает плоскость, перпендикулярную оси z; два угла θ_x и θ_y, определяющие направление луча. Направление в трехмерном пространстве задается единичным вектором $\mathbf{v} = (v_x,\ v_y,\ v_z)$. Если направление этого вектора близко к положительному направлению оси z, то можно считать, что он имеет вид $\mathbf{v} = (\theta_x,\ \theta_y,\ v_z)$, где $v_z \doteq 1 - \dfrac{1}{2}(\theta_x^2 + \theta_y^2) \doteq 1$, при условии, что углы θ_x и θ_y малы. Угловые переменные θ заменим на p, определяемые формулами $p_x = n\theta_x$ и $p_y = n\theta_y$. (Если среда анизотропна, например, для некоторых кристаллов, то соотношение между θ и p становится более сложным. Но здесь мы не будем рассматривать такие среды.) Конечно, все происходит в некоторой фиксированной плоскости, перпендикулярной оси z. Если мы будем рассматривать две плоскости z_1 и z_2, то лучу будут соответствовать два вектора для каждой плоскости соответственно:

$$\mathbf{u}_1 = \begin{pmatrix} q_{x1} \\ q_{y1} \\ p_{x1} \\ p_{y1} \end{pmatrix} \quad \text{и} \quad \mathbf{u}_2 = \begin{pmatrix} q_{x2} \\ q_{y2} \\ p_{x2} \\ p_{y2} \end{pmatrix}.$$

Наша задача — найти соотношение между $\mathbf{u}_1$ и $\mathbf{u}_2$. Поскольку мы пренебрегаем членами порядка выше первого, можно написать

$$\mathbf{u}_2 = M\mathbf{u}_1,$$

где M — матрица размера 4×4.

Наша задача — установить, какие матрицы 4×4 возникают в линейной оптике. Первая очевидная мысль, что это матрицы, удовлетворяющие условию $\operatorname{Det} M = 1$. Однако это не так. Безусловно, верно, что все оптические матрицы должны иметь единичный детерминант, но не верно, что все матрицы 4×4 с единичным детерминантом могут быть матрицами преобразований в линейной оптике. Должно быть более строгое ограничение. Чтобы объяснить, в чем состоит это более строгое ограничение, вернемся назад и еще раз сформулируем условие, при котором матрица 2×2 имеет определитель, равный единице. После этого можно будет вернуться к соответствующему условию для случая четырех переменных. Пусть на плоскости даны два вектора:

$$\mathbf{w} = \begin{pmatrix} q \\ p \end{pmatrix} \quad \text{и} \quad \mathbf{w}' = \begin{pmatrix} q' \\ p' \end{pmatrix}.$$

В разделе 4.9 мы определили антисимметричное произведение $\omega(\mathbf{w}, \mathbf{w}')$ между двумя векторами согласно формуле

$$\omega(\mathbf{w}, \mathbf{w}') = qp' - q'p.$$

Геометрический смысл $\omega(\mathbf{w}, \mathbf{w}')$ — это ориентированная площадь параллелограмма, построенного на векторах $\mathbf{w}$ и $\mathbf{w}'$ (см. рис. 9.17). Из определения и из геометрического смысла следует антисимметрия ω:

$$\omega(\mathbf{w}, \mathbf{w}') = -\omega(\mathbf{w}', \mathbf{w}).$$

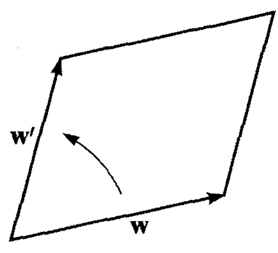


Рис. 9.17

Матрица 2×2 сохраняет площадь и ориентацию тогда и только тогда, когда ее определитель равен 1, т. е. $\operatorname{Det} M = 1$ тогда и только тогда, когда

$$\omega(M\mathbf{w}, M\mathbf{w}') = \omega(\mathbf{w}, \mathbf{w}')$$

для всех $\mathbf{w}$ и $\mathbf{w}'$. А сейчас предположим, что в четырехмерном пространстве есть два вектора

$$\mathbf{u} = \begin{pmatrix} q_x \\ q_y \\ p_x \\ p_y \end{pmatrix} \quad \text{и} \quad \mathbf{u}' = \begin{pmatrix} q'_x \\ q'_y \\ p'_x \\ p'_y \end{pmatrix}.$$

Введем по определению

$$\omega(\mathbf{u}, \mathbf{u}') = q_x p'_x - q'_x p_x + q_y p'_y - q'_y p_y.$$

Произведение ω по-прежнему остается антисимметричным:

$$\omega(\mathbf{u}', \mathbf{u}) = -\omega(\mathbf{u}, \mathbf{u}'),$$

но геометрический смысл его не очень понятен.

Оказывается, что матрица M (4×4) может быть матрицей преобразования в линейной оптике только при выполнении условия

$$\omega(M\mathbf{u}, M\mathbf{u}') = \omega(\mathbf{u}, \mathbf{u}')$$

для всех векторов $\mathbf{u}$ и $\mathbf{u}'$. Матрицы такого типа называются в физической литературе *линейными каноническими преобразованиями*, а в математической литературе они называются *линейными симплектическими преобразованиями*. Эти матрицы (и их обобщения на случай большего числа измерений) играют важнейшую роль в теоретической механике и геометрии.

Сначала сформулируем ряд основных сведений относительно линейных симплектических преобразований в пространстве четырех переменных. Тогда мы сможем увидеть, что наши рассуждения, показывающие, что гауссова оптика эквивалентна группе $Sl(2, \mathbb{R})$, могут быть использованы для демонстрации эквивалентности линейной оптики группе $Sp(4, \mathbb{R})$ — группе линейных симплектических преобразований в пространстве четырех измерений.

В общем случае пусть V обозначает векторное (вещественное конечномерное) пространство. Билинейная форма Ω в пространстве V — это любая функция $\Omega : V \times V \to \mathbb{R}$, линейная по каждой переменной при условии, что другая фиксирована, т. е. $\Omega(\mathbf{u}, \mathbf{v})$ —

линейная функция $\mathbf{v}$ при фиксированном $\mathbf{u}$ и линейная функция $\mathbf{u}$ при фиксированном $\mathbf{v}$. Мы говорим, что Ω антисимметрична, если $\Omega(\mathbf{u}, \mathbf{v}) = -\Omega(\mathbf{v}, \mathbf{u})$ для всех $\mathbf{u}$ и $\mathbf{v}$ в пространстве V. Мы говорим, что Ω невырождена, если при любом фиксированном $\mathbf{u}$ линейная функция $\Omega(\mathbf{u}, \cdot)$ не тождественный нуль (если только сам вектор $\mathbf{u}$ не равен нулю). Антисимметричная невырожденная билинейная форма в пространстве V называется *симплектической формой*. Векторное пространство, снабженное данной симплектической формой, называется *симплектическим векторным пространством*, говорят, что оно имеет *симплектическую структуру*. Пусть V — симплектическое векторное пространство с симплектической формой Ω и пусть A — линейное преобразование пространства V в себя; тогда мы говорим, что A — *симплектическое преобразование*, если $\Omega(A\mathbf{u}, A\mathbf{v}) = \Omega(\mathbf{u}, \mathbf{v})$ для всех $\mathbf{u}$ и $\mathbf{v}$ в пространстве V. Существует теорема (Guillemin, Sternberg, *Symplectic Techniques in Physics*, Chapter II), утверждающая, что каждое симплектическое векторное пространство должно быть пространством четного числа измерений, и что каждое симплектическое линейное преобразование должно иметь детерминант, равный единице, а значит, быть обратимым. Очевидно, что преобразование, обратное к симплектическому преобразованию, должно быть симплектическим, и что произведение любых двух симплектических преобразований тоже должно быть симплектическим. Множество всех симплектических линейных преобразований называется симплектической группой в пространстве V и обозначается $Sp(V)$.

Предположим, что $V = \mathbb{R}^n + \mathbb{R}^n$. Это — пространство, в котором вектор можно записать в виде

$$\mathbf{u} = \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix}, \quad \text{где} \quad \mathbf{q} = \begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix} \quad \text{и} \quad \mathbf{p} = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}.$$

Симплектическая форма Ω определяется формулой

$$\Omega(\mathbf{u}, \mathbf{u}') = \mathbf{p} \cdot \mathbf{q}' - \mathbf{p}' \cdot \mathbf{q},$$

где символ $\cdot$ обозначает обычное скалярное произведение в пространстве $\mathbb{R}^n$. Если использовать запись скалярного произведения

$\mathbf{u}\cdot\mathbf{u}' = \mathbf{q}\cdot\mathbf{q}' + \mathbf{p}\cdot\mathbf{p}'$, то форму можно переписать по-другому

$$\Omega(\mathbf{u},\mathbf{u}') = \mathbf{u}'\cdot J\mathbf{u},$$

где J — матрица $2n\times 2n$ вида $\begin{pmatrix} 0 & \mathbb{I} \\ -\mathbb{I} & 0 \end{pmatrix}$, а $\mathbb{I}$ — единичная матрица $n \times n$. Повторим: линейное преобразование T в пространстве V симплектично тогда и только тогда, когда для всех $\mathbf{u}$ и $\mathbf{u}'$ выполняется равенство

$$\Omega(T\mathbf{u}, T\mathbf{u}') = \Omega(\mathbf{u},\mathbf{u}').$$

Это можно переписать в виде

$$T^t JT\mathbf{u}\cdot\mathbf{u}' = J\mathbf{u}\cdot\mathbf{u}',$$

где T^t обозначает транспонированную матрицу (т. е. матрицу, сопряженную к T по отношению к скалярному произведению в пространстве V). Поскольку это равенство выполняется для всех $\mathbf{u}$ и $\mathbf{u}'$, мы получаем

$$T^t JT = J.$$

Отображение T можно записать в виде

$$T\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix} = \begin{pmatrix} A\mathbf{q} + B\mathbf{p} \\ C\mathbf{q} + D\mathbf{p}\end{pmatrix},$$

где A, B, C и D — матрицы $n\times n$, т. е.

$$T = \begin{pmatrix} A & B \\ C & D\end{pmatrix}.$$

Тогда

$$T^t = \begin{pmatrix} A^t & C^t \\ B^t & D^t\end{pmatrix},$$

где A^t обозначает матрицу, транспонированную к A, и т. д. Условие $T^tJT = J$ переходит с совокупность условий: $A^tC = C^tA$, $B^tD = D^tB$, и $A^tD - C^tB = \mathbb{I}$. Заметим, что матрица T^{-1}, которая также симплектична, имеет вид

$$T^{-1} = \begin{pmatrix} D^t & -B^t \\ -C^t & A^t\end{pmatrix},$$

причем выполняются условия:

$$DC^t = CD^t \quad \text{и} \quad BA^t = AB^t.$$

А теперь вернемся к проблеме доказательства того, что группа линейных симплектических преобразований (в пространстве четырех измерений) совпадает с множеством всех преобразований в линейной оптике. Как и в случае гауссовой оптики, доказательство распадается на две части. Первая, физическая часть доказательства показывает, что (в линейном приближении) матрица $\begin{pmatrix} \mathbb{I} & 0 \\ -P & \mathbb{I} \end{pmatrix}$, где $P = P^t$ — симметричная матрица, соответствует преломлению на поверхности раздела между двумя областями с постоянным коэффициентом преломления (для каждой поверхности появляется своя матрица P), и что матрица $\begin{pmatrix} \mathbb{I} & d\mathbb{I} \\ 0 & \mathbb{I} \end{pmatrix}$ соответствует движению в среде с постоянным коэффициентом преломления, где d — оптическая длина вдоль оси. Вторая часть доказательства (математическая) позволяет доказать, что любая симплектическая матрица может быть записана как произведение матриц этого вида.

Здесь мы не будем заниматься математическим доказательством, которое является довольно хитроумным обобщением доказательства, приведенного в разделе 9.2. Читатель может найти его в книге Guillemin, Sternberg, *Symplectic Techniques in Physics*, стр. 27–30. Давайте рассмотрим физическую часть этой задачи. Напомним, что мы определяем падающий луч заданием его направления $\mathbf{v} = (v_x, v_y, v_z)$, $\|\mathbf{v}\| = v_x^2 + v_y^2 + v_z^2 = 1$, и точкой, где луч пересекает плоскость, параллельную плоскости xy и находящуюся в точке z на оптической оси. Тогда

$$v_z = (1 - v_x^2 - v_y^2)^{1/2} = 1 - \frac{1}{2}(v_x^2 + v_y^2) + \cdots \doteq 1.$$

Мы предполагаем v_x и v_y малыми и поэтому пренебрегаем квадратичными членами. Пусть

$$p_x = nv_x, \qquad p_y = nv_y,$$

где n — коэффициент преломления. Смещение на расстояние t вдоль оптической оси — это с точностью до квадратичных членов по v_x и v_y то же, что и смещение на расстояние t вдоль вектора $\mathbf{v}$. Следовательно,

$$q_{2x} - q_{1x} = tv_x$$

и

$$q_{2y} - q_{1y} = tv_y.$$

Тогда мы имеем

$$\begin{pmatrix} \mathbf{q}_2 \\ \mathbf{p}_2 \end{pmatrix} = \begin{pmatrix} \mathbb{I} & d\mathbb{I} \\ 0 & \mathbb{I} \end{pmatrix} \begin{pmatrix} \mathbf{q}_1 \\ \mathbf{p}_1 \end{pmatrix},$$

где $d = t/n$ (см. рис. 9.18).

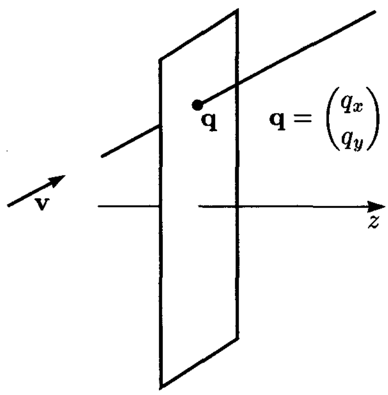


Рис. 9.18

Теперь перейдем к преломлению света. Предположим, что преломляющая поверхность квадратична и определяется формулой

$$z' - z = \frac{1}{2} k\,\mathbf{q} \cdot \mathbf{q},$$

где k — симметричная матрица 2×2. Нормаль к этой поверхности в точке $\mathbf{q}$ определяется формулой

$$\mathbf{u} = (k\mathbf{q}, -1).$$

(С точностью до членов порядка выше первого вектор $\mathbf{u}$ имеет единичную длину.) Проекция вектора $\mathbf{v}$ на плоскость, касательную к поверхности в точке $\mathbf{q}$, равна

$$\mathbf{v} - (\mathbf{v} \cdot \mathbf{u})\mathbf{u}.$$

Запишем $\mathbf{v} = (v_x, v_y, 1) = (\boldsymbol{\nu}, 1)$, тогда $\mathbf{v} \cdot \mathbf{u} = k\mathbf{q} \cdot \boldsymbol{\nu} - 1$ и

$$\mathbf{v} - (\mathbf{v} \cdot \mathbf{u})\mathbf{u} = (\boldsymbol{\nu}, 1) - (k\mathbf{q} \cdot \boldsymbol{\nu} - 1)(k\mathbf{q}, -1).$$

Откуда, пренебрегая квадратичным членом $k\mathbf{q} \cdot \boldsymbol{\nu}$, получаем для проекции выражение

$$(\boldsymbol{\nu} + k\mathbf{q}, 0).$$

Закон Снеллиуса говорит, что $n_1(\mathbf{v}_1 - (\mathbf{v}_1 \cdot \mathbf{u})\mathbf{u}) = n_2(\mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{u})\mathbf{u})$. В линейном приближении и с учетом того, что $\mathbf{p}_1 = n\boldsymbol{\nu}_1$ и $\mathbf{p}_2 = n\boldsymbol{\nu}_2$, получаем

$$\mathbf{p}_1 - n_1 k\mathbf{q} = \mathbf{p}_2 - n_2 k\mathbf{q}$$

или

$$\mathbf{p}_2 = \mathbf{p}_1 - P\mathbf{q},$$

где

$$P = -(n_1 - n_2)k.$$

Итак, мы получили матрицу преломления $\begin{pmatrix} \mathbb{I} & 0 \\ -P & \mathbb{I} \end{pmatrix}$. Таким образом, мы показали, что группа преобразований линейной оптики совпадает с группой $Sp(4, \mathbb{R})$.

Есть еще один заслуживающий внимания момент в гауссовой оптике. Предположим, что оптическая система обладает вращательной симметрией. Тогда на каждой преломляющей поверхности матрица оптической силы P имеет вид $P = m\mathbb{I}$, где m — скалярный коэффициент и $\mathbb{I}$ — единичная матрица 2×2. Очевидно, что произведение таких матриц на матрицу вида $\begin{pmatrix} \mathbb{I} & e\mathbb{I} \\ 0 & \mathbb{I} \end{pmatrix}$ может быть записано в виде $\begin{pmatrix} a\mathbb{I} & b\mathbb{I} \\ c\mathbb{I} & d\mathbb{I} \end{pmatrix}$. Заметим, что когда такая матрица действует на вектор

$$\begin{pmatrix} q_x \\ q_y \\ p_x \\ p_y \end{pmatrix},$$

то получится то же самое, когда матрица $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ действует отдельно на $\begin{pmatrix} q_x \\ p_x \end{pmatrix}$ и $\begin{pmatrix} q_y \\ p_y \end{pmatrix}$. Итак, когда мы изучали гауссову оптику, то не обязательно было ограничиваться изучением коаксиальных лучей. Можно было рассматривать произвольные лучи, разложив их на x- и y-компоненты. Это является следствием принятого линейного приближения.

Основная формула для оптической длины

$$L = L_o + \frac{1}{2}(\mathbf{p}_2 \cdot \mathbf{q}_2 - \mathbf{p}_1 \cdot \mathbf{q}_1)$$

(где теперь $\mathbf{p}$ и $\mathbf{q}$ — векторы) доказывается так же, как и в гауссовом случае, проверяя, что происходит с базисными компонентами. Нет смысла повторять это еще раз.

Две плоскости называются несопряженными, если оптическая матрица B, связывающая их, несингулярна. Тогда можно решить уравнения

$$\mathbf{q}_2 = A\mathbf{q}_1 + B\mathbf{p}_1$$

и

$$\mathbf{p}_2 = C\mathbf{q}_1 + D\mathbf{p}_1$$

относительно $\mathbf{p}_1$ и $\mathbf{p}_2$, которые оказываются равны

$$\mathbf{p}_1 = -B^{-1}A\mathbf{q}_1 + B^{-1}\mathbf{q}_2$$

и

$$\mathbf{p}_2 = (C - DB^{-1}A)\mathbf{q}_1 + DB^{-1}\mathbf{q}_2.$$

Теперь можно написать

$$L = L_o + W(\mathbf{q}_1, \mathbf{q}_2),$$

где

$$W(\mathbf{q}_1, \mathbf{q}_2) = \frac{1}{2}\left[DB^{-1}\mathbf{q}_2 \cdot \mathbf{q}_2 + B^{-1}A\mathbf{q}_1 \cdot \mathbf{q}_1 - (2B^t)^{-1}\mathbf{q}_1 \cdot \mathbf{q}_2\right].$$

(При выводе этой формулы мы использовали тождество

$$-(B^t)^{-1} = C - DB^{-1}A,$$

которое следует из несингулярности B и условия $A^tD - B^tC = \mathbb{I}$.) Прямое вычисление (с использованием этого тождества) дает (в очевидном смысле)

$$\frac{\partial L}{\partial \mathbf{q}_2} = \mathbf{p}_2 \quad \text{и} \quad \frac{\partial L}{\partial \mathbf{q}_1} = -\mathbf{p}_1.$$

Таким образом, знание L дает возможность выразить $\mathbf{p}_1$ и $\mathbf{p}_2$ через $\mathbf{q}_1$ и $\mathbf{q}_2$.

Мы можем теперь коротко описать переход к (нелинейной) геометрической оптике. Сформулируем условие симплектичности матрицы A немного по-другому. В пространстве $\mathbb{R}^4$ рассмотрим 2-форму

$$\omega = dq_x \wedge dp_x + dq_y \wedge dp_y.$$

Линейное отображение $A : \mathbb{R}^4 \to \mathbb{R}^4$ симплектично тогда и только тогда, когда

$$A^*\omega = \omega.$$

Другими словами, перенос формы ω в результате действия A дает снова ω. Поэтому дадим следующее определение: дифференцируемое отображение ϕ называется *симплектическим*, если

$$\phi^*\omega = \omega.$$

Мы не требуем здесь, чтобы ϕ было линейно. (В старой литературе симплектические отображения назывались *каноническими преобразованиями*.) Гамильтон показал, что отображения в геометрической оптике (от входа до выхода) являются в точности симплектическими отображениями. Кроме того, он показал, что в предположении *неконгруэнтности* симплектическое отображение определяется характеристической функцией L, где $L(\mathbf{q}_1, \mathbf{q}_2)$ есть *оптическая длина* пути, соединяющего точки $\mathbf{q}_1$ и $\mathbf{q}_2$. (Конечно, в этом случае формула для L будет гораздо сложнее.)

Примерно через десять лет после написания фундаментальных работ по оптике Гамильтон сделал потрясающее наблюдение: тот же самый формализм применим и в механике точечных частиц. Пусть $q_1, \ldots, q_n$ — обобщенные координаты системы частиц, а $p_1, \ldots, p_n$ — их обобщенные импульсы. Вместо оптической оси z

рассматривается время. Тогда преобразование от начальных координат и импульсов к конечным всегда будет симплектическим. В XIX веке это открытие привело к громадному прогрессу в теоретической механике. А почти столетием позже, в 1920 году, аналогия между оптикой и механикой, обнаруженная Гамильтоном, сыграла ключевую роль в развитии квантовой механики.

Резюме

А. Матричная формулировка гауссовой оптики

Вы должны уметь использовать двухкомпонентный вектор для описания луча, проходящего через плоскость отсчета.

Вы должны уметь получать и использовать матрицы 2×2 для описания эффектов распространения, преломления света и тонкой линзы.

В. Система линз

Вы должны уметь вычислять матрицу для системы преломляющих поверхностей или системы тонких линз между двумя плоскостями отсчета.

Для данной системы линз надо уметь находить главные и фокальные плоскости. Надо уметь пользоваться ими для нахождения изображения объекта.

С. Гамильтонова оптика

Для гауссовой оптической системы надо уметь находить точечные характеристики Гамильтона между плоскостями отсчета и использовать их для определения траектории луча, соединяющего две точки в двух плоскостях.

Задачи

9.1. На рис. 9.19 изображены фокальные и главные плоскости для тонкой линзы. Лучи, падающие слева параллельно оси системы, преломляются и проходят через фокус в плоскости F_2. Лучи, выходящие из фокуса в плоскости F_1, преломляются и выходят параллельными оси. Главные плоскости H_1 и H_2 связаны с фокальными плоскостями F_1 и F_2 соответственно.

(a) С помощью луча R_1, изображенного на рисунке, определяется положение в плоскости z. Нарисуйте ход луча R_1 и еще *двух* других лучей.

(b) Вычислите координаты изображения из предыдущего пункта с помощью уравнения Ньютона относительно любой из плоскостей, изображенных на рис. 9.19.

(c) Постройте матрицу оптической системы между плоскостями z_1 и z_2. С помощью этой матрицы определите положение и наклон луча R_1, выходящего из линзы в плоскости z_2.

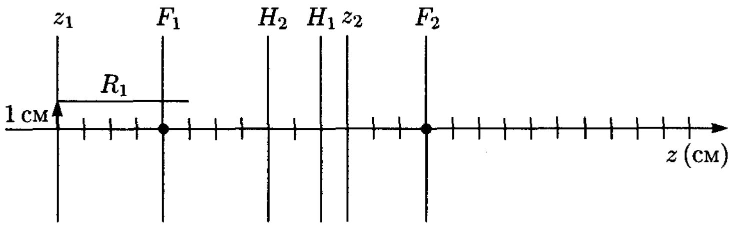


Рис. 9.19

9.2. На рис. 9.20 изображена толстая линза, сделанная из стекла с показателем преломления $n = 3/2$. Постройте матрицу распространения луча между плоскостями отсчета z_1 и z_2. Определите положение фокальных плоскостей F_1 и F_2, главных плоскостей H_1 и H_2 и покажите их на рисунке. Нарисуйте ход лучей и определите положение изображения объекта, находящегося слева от z_1 на расстоянии 1 см. Используя уравнение Ньютона, проверьте полученные результаты.

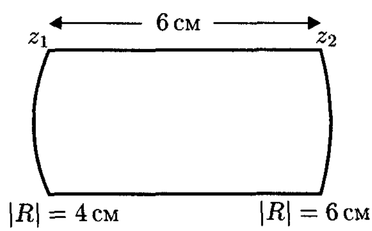


Рис. 9.20

9.3. Предположим, что ход лучей считается фундаментальной характеристикой свойств тонкой линзы, т. е. предположим, что пересечение луча, проходящего через центр линзы, с лучом, параллельным оси слева от линзы и проходящим через ее фокус справа, определяет пересечение всех лучей от выбранного объекта.

(a) Используя это допущение, получите уравнение тонкой линзы. Рассмотрите только тот случай, когда p, q и f положительны.

(b) Используя это допущение, докажите, что тонкая линза может быть описана матрицей 2×2 и получите вид этой матрицы.

9.4. Шарик радиуса 6 см (рис. 9.21) сделан из стекла с показателем преломления $n = 3/2$. Для лучей, проходящих вблизи диаметра, этот шар ведет себя как линейная толстая линза (толстая линза — это цилиндрическая сердцевина, осью которой является диаметр шара). Постройте матрицы для этой линзы между плоскостями отсчета z_1 и z_2, между фокальными и главными плоскостями. Покажите эти плоскости на рисунке.

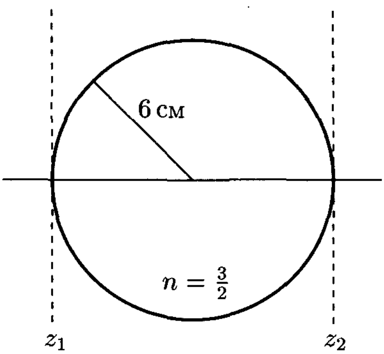


Рис. 9.21

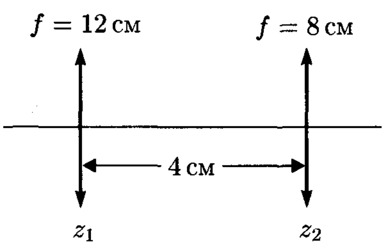


Рис. 9.22

9.5. (a) Для системы из двух собирающих линз, изображенной на рис. 9.22, постройте матрицу для движения луча между плоскостями z_1 и z_2.

(b) Найдите положение фокальных плоскостей z_1' и z_2', используя уравнение тонкой линзы и матрицу этой оптической системы. Постройте матрицу между двумя фокальными плоскостями.

(c) Найдите положение главных плоскостей H_1 и H_2 и постройте соответствующую матрицу. Это легко сделать, учтя тот факт, что фокусное расстояние $f = 6$ см и каждая главная плоскость находится на расстоянии 6 см от соответствующей фокальной плоскости. Обратите внимание, что обе главные плоскости лежат между линзами, и в этом случае H_1 находится справа от H_2.

(d) На рисунке покажите положение фокальных и главных плоскостей этой системы.

(e) Пусть плоскость z_3 находится на расстоянии 12 см слева от z_1. Четырьмя способами найдите плоскость, сопряженную этой: (1) перемножая матрицы, (2) дважды используя уравнение тонкой линзы, (3) используя уравнение Ньютона $x_1 x_2 = f^2$ и (4) рисуя ход лучей.

9.6. Оптическая система состоит из двух тонких линз с фокусными расстояниями f_1 и f_2, расположенными на расстоянии t друг от друга (рис. 9.23). Первая фокальная плоскость находится слева от первой линзы на расстоянии l_1; вторая фокальная плоскость находится справа от второй линзы на расстоянии l_2. Докажите, что фокусное расстояние f этой системы удовлетворяет уравнению $f^2 - tf - l_1 l_2 = 0$. Имея в виду, что все параметры l_1, l_2, f, f_1 и f_2 могут иметь любой знак, определите, какой из корней этого квадратного уравнения имеет физический смысл.

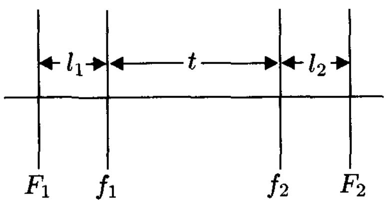


Рис. 9.23

9.7. Придумайте систему тонких линз, оптическая матрица которой имеет вид $\begin{pmatrix} \mathbb{I} & 0 \\ 0 & \mathbb{I} \end{pmatrix}$. (Указание: система состоит из нескольких линз. Можно начать с построения системы, матрица которой удовлетворяет уравнению $M^2 = \mathbb{I}$.)

9.8. На рис. 9.24 изображена оптическая система. В плоскости z_1 падающий луч имеет координаты $\begin{pmatrix} q_1 \\ p_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Найдите координаты луча, выходящего из плоскости z_2.

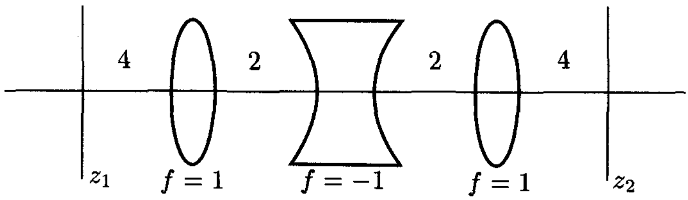


Рис. 9.24

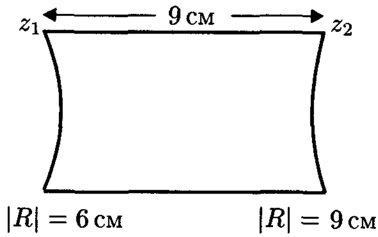


Рис. 9.25

9.9. На рис. 9.25 изображена толстая линза, сделанная из стекла с показателем преломления $n = 3/2$.

(a) Постройте матрицу, соответствующую движению луча между плоскостями z_1 и z_2.

(b) Выясните, какой падающий луч $\begin{pmatrix} q_1 \\ p_1 \end{pmatrix}$ выходит из системы в плоскости z_2 с координатами $\begin{pmatrix} q_2 \\ p_2 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/6 \end{pmatrix}$.

9.10. На рис. 9.26 изображена собирающая линза с фокусным расстоянием $f = 10$ см. Она сделана из стекла с показателем преломления $n = 1.4$. Выпуклые поверхности линзы имеют одинаковый радиус кривизны R.

(a) Вычислите радиус кривизны R, найдите зависимость толщины линзы b от расстояния q от оси. (Указание: $b(2) = 0$.)

(b) Луч проходит через точку A, потом по линии ACF. Покажите, что время движения вдоль этой траектории минимально по сравнению с любой траекторией, которая проходит через линзу с другими значениями q.

(c) Покажите, что время движения луча вдоль траектории ACB максимально по сравнению со временем движения по любой траектории, связывающей точки A и B.

(d) Напишите функцию $W(q_A, q_F)$ для плоскостей A и F. Покажите, что уравнения Гамильтона дают правильный наклон луча, заданного параметрами $q_A = 1$ и $q_F = 0$. Сделайте то же самое для плоскостей A и B. Используйте функцию $W(q_A, q_B)$ для определения траектории луча, проходящего через A и B.

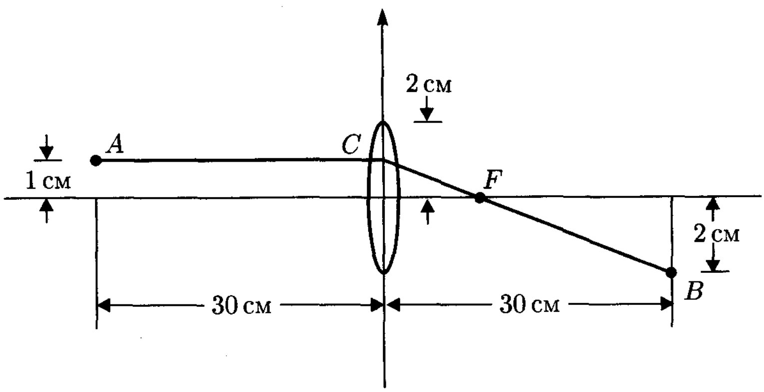


Рис. 9.26

9.11. Пусть $\mathbf{v}_1 = \begin{pmatrix} q_1 \\ p_1 \end{pmatrix}$ и $\tilde{\mathbf{v}}_1 = \begin{pmatrix} \tilde{q}_1 \\ \tilde{p}_1 \end{pmatrix}$ обозначают два луча, входящие в произвольную гауссову оптическую систему. Симплектическое скалярное произведение этих векторов определяется формулой

$$\omega(\mathbf{v}_1, \tilde{\mathbf{v}}_1) = q_1\tilde{p}_1 - \tilde{q}_1 p_1.$$

(a) Покажите, что это скалярное произведение сохраняется при прохождении оптической системы, т. е. что $\omega(\mathbf{v}_2, \tilde{\mathbf{v}}_2) = \omega(\mathbf{v}_1, \tilde{\mathbf{v}}_1)$.

(b) Покажите, что $\omega(\mathbf{v}_1, \tilde{\mathbf{v}}_1) = 0$, если $\mathbf{v}_1$ и $\tilde{\mathbf{v}}_1$ обозначают лучи, пересекающиеся на оптической оси.

(c) Допустим, что два луча проходят через одну точку q_1 в плоскости отсчета z_1 с углом ϕ_1 между ними. Если эти лучи пересекаются в сопряженной плоскости z_2 с углом ϕ_2 между ними, то чему равно расстояние от оси q_2? (Пусть $n = 1$ в плоскостях z_1 и z_2.)

Глава 10

Векторные пространства и линейные преобразования

В главе 10 мы возвращаемся назад и доказываем основные свойства конечномерных пространств и их линейных преобразований. Здесь, главным образом, обобщаются результаты, полученные в первых четырех главах для двумерного пространства. Излагается новый алгоритм — редукция по строкам. Вводятся два новых понятия: понятие сопряженного пространства и факторпространства. (Их было сложно ввести раньше.) Эти понятия будут играть решающую роль в дальнейшем[1].

Введение

Мы уже довольно подробно изучили двумерное пространство. Но при этом рассматривались только две конкретные модели. Векторное пространство V было либо множеством перемещений в аффинной плоскости, либо это было пространство $\mathbb{R}^2$ — множество упорядоченных пар вещественных чисел. Вводя координаты,

---

[1]В параграфе 10.9 авторы выходят за пределы линейной алгебры и доказывают важную теорему о локальной структуре дифференцируемого отображения. — *Прим. ред.*

мы могли отождествить любое двумерное пространство с пространством $\mathbb{R}^2$, в котором линейное преобразование представлялось матрицей 2×2.

Сейчас мы начнем изучение произвольного векторного пространства в абстрактном виде с аксиоматической точки зрения. Преимущество такого подхода состоит в том, что мы сможем рассматривать векторные пространства, которые не определяются ни через геометрические понятия, ни как множество наборов из n вещественных чисел. Окажется, что такое векторное пространство, содержащее только конечное число *линейно независимых* элементов, может быть отождествлено с пространством $\mathbb{R}^n$ для некоторого целого числа n, так что в конечном счете мы вернемся к изучению пространства $\mathbb{R}^n$ и использованию матриц для описания линейных преобразований. В последующем надо все время иметь в виду известную двумерную геометрическую модель векторного пространства, чтобы напоминать себе, что последующие аксиомы и определения вполне разумны. Однако, мы специально будем давать примеры векторных пространств без геометрического содержания. Такие векторные пространства являются частью естественного математического языка в различных областях физики, особенно в электромагнитной теории и квантовой механике.

10.1. Свойства векторных пространств

Начнем с повторения основных определений.

Векторное пространство V, иногда называемое линейным пространством, состоит из элементов — векторов, для которых справедливы ряд аксиом. Эти пространства будем обозначать заглавными буквами, напрмер, V, W^*, C_1, а их элементы — строчными буквами жирным шрифтом, например, $\mathbf{v}$, $\mathbf{w}_2$, $\mathbf{b}^1$.

В векторном пространстве V вводится правило, согласно которому любым двум элементам $\mathbf{v}_1$ и $\mathbf{v}_2$ ставится в соответствие третий элемент $\mathbf{v}$, называемый их *суммой* и обозначаемый $\mathbf{v}_1+\mathbf{v}_2$. Эта операция удовлетворяет тем же аксиомам, что и сложение вещественных чисел:

Коммутативный закон:

$$\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1. \tag{10.1}$$

Ассоциативный закон:

$$(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3). \tag{10.2}$$

Существование нуля: существует такой элемент $\mathbf{0}$, что

$$\mathbf{v} + \mathbf{0} = \mathbf{v} \quad \text{для всех} \quad \mathbf{v}. \tag{10.3}$$

Существование противоположного: для любого $\mathbf{v}$ существует такой элемент $-\mathbf{v}$, что

$$-\mathbf{v} + \mathbf{v} = \mathbf{0}. \tag{10.4}$$

В некоторых случаях сложение определяется прямо через сложение вещественных чисел, поэтому сразу бывает ясно, что эти аксиомы выполняются. Например, в пространстве $\mathbb{R}^2$ мы определяем

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix}.$$

Аналогично, можно рассматривать двумерное векторное пространство всех функций, определенных на двухточечном множестве $\{A, B\}$, где сложение определяется отдельно для каждой точки, т. е. $\mathbf{h} = \mathbf{f} + \mathbf{g}$ — функция, обладающая свойствами

$$\mathbf{h}(A) = \mathbf{f}(A) + \mathbf{g}(A),$$
$$\mathbf{h}(B) = \mathbf{f}(B) + \mathbf{g}(B).$$

В качестве еще одного примера рассмотрим пространство всех непрерывных функций на отрезке $[0, 1]$, где также сложение определяется отдельно для каждой точки. Если $\mathbf{f}$ и $\mathbf{g}$ — элементы этого пространства, то их суммой будет функция $\mathbf{h}$, задаваемая формулой $\mathbf{h}(x) = \mathbf{f}(x) + \mathbf{g}(x)$. В этом случае важно отметить, что для любых $\mathbf{f}$ и $\mathbf{g}$ сумма $\mathbf{h}$ будет также непрерывной функцией и поэтому принадлежит векторному пространству.

Ясно, что во всех этих примерах нулевой элемент и противоположный элемент существуют и единственны. Это верно для любого векторного пространства, что легко доказывается, исходя из приведенных аксиом. Читателю предоставляется возможность сделать это самостоятельно.

В векторном пространстве должна быть определена еще одна операция: умножение вектора на скаляр. В этой главе мы всегда будем считать, что скаляр — вещественное число. Поэтому векторное пространство будет *вещественным векторным пространством.* Позднее рассмотрим комплексные векторные пространства, в которых элементы могут умножаться на комплексные числа. И в этом случае должны быть справедливы аксиомы обычного умножения, а именно: если c_1 и c_2 — скаляры, а $\mathbf{v}_1$ и $\mathbf{v}_2$ — векторы, то справедливы

Ассоциативный закон:

$$c_1(c_2\mathbf{v}) = (c_1c_2)\mathbf{v}. \tag{10.5}$$

Дистрибутивные законы:

$$\begin{aligned}(c_1 + c_2)\mathbf{v} &= c_1\mathbf{v} + c_2\mathbf{v},\\ c(\mathbf{v}_1 + \mathbf{v}_2) &= c\mathbf{v}_1 + c\mathbf{v}_2.\end{aligned} \tag{10.6}$$

Умножение на 1 дает тождество:

$$1\mathbf{v} = \mathbf{v} \quad \text{для всех} \quad \mathbf{v}. \tag{10.7}$$

Поскольку аксиомы сложения и умножения на скаляр в векторном пространстве точно такие же, как и в арифметике, то в векторной алгебре сохраняются все арифметические свойства. Они легко доказываются, исходя из аксиом. Ниже приводится перечень этих свойств. Подумайте и постарайтесь доказать их, так как в определениях они не содержатся.

(a) $0 \cdot \mathbf{v} = \mathbf{0}$.
(b) $c \cdot \mathbf{0} = \mathbf{0}$.
(c) $(-c) \cdot \mathbf{v} = -(c \cdot \mathbf{v}) = c \cdot (-\mathbf{v})$.
(d) $\mathbf{v} + \mathbf{v} = 2\mathbf{v}$, $\mathbf{v} + \mathbf{v} + \mathbf{v} = 3\mathbf{v}$, и т. д.
(e) Если $a \cdot \mathbf{v} = \mathbf{0}$, то либо $a = 0$, либо $\mathbf{v} = \mathbf{0}$.
(f) $-(\mathbf{v} + \mathbf{w}) = -\mathbf{v} + (-\mathbf{w})$.

10.2. Дуальное пространство

Для выбранного векторного пространства V можно рассмотреть множество всех *линейных функций* из V в $\mathbb{R}$. Сейчас мы покажем, что они образуют векторное пространство, называемое *дуальным пространством*, *сопряженным* к пространству V, и обозначаемое V^*. Элементы этого пространства будем обозначать греческими буквами жирным шрифтом и договоримся писать у них индексы сверху, в отличие от элементов пространства V. Таким образом, $\mathbf{v}_1, \mathbf{v}_2, \ldots$ — элементы пространства V, а $\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots$ — элементы пространства V^*. Действие элемента из V^* на элемент из V будем обозначать $\boldsymbol{\alpha}[\mathbf{v}]$.

Сумму двух элементов пространства V^* определим как для функций, т. е. для любого $\mathbf{v} \in V$ положим $(\boldsymbol{\alpha}^1 + \boldsymbol{\alpha}^2)[\mathbf{v}] = \boldsymbol{\alpha}^1[\mathbf{v}] + \boldsymbol{\alpha}^2[\mathbf{v}]$. Поскольку сумма линейных функций — тоже линейная функция, $\boldsymbol{\alpha}^1 + \boldsymbol{\alpha}^2$ является элементом пространства V^*. Легко видеть, что все аксиомы сложения (10.1)–(10.4) выполняются; нулевым элементом V^* будет нулевая функция, которая, конечно, линейна. Аналогично определяется умножение на скаляр:

$$(c\,\boldsymbol{\alpha})[\mathbf{v}] = c(\boldsymbol{\alpha}[\mathbf{v}]).$$

Очевидно, что $c\,\boldsymbol{\alpha}$ — линейная функция, и что для введенной таким образом операции умножения на скаляр выполняются все аксиомы умножения.

Приведем несколько очень разных примеров определения элементов дуального пространства.

1. Пусть V — двумерное пространство $\mathbb{R}^2$. Элементы этого пространства обозначены $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$. Тогда можно считать, что элементами пространства V^* будут вектор-строки. Например, $\boldsymbol{\alpha} = (a, b)$, причем $\boldsymbol{\alpha}[\mathbf{v}] = (a, b)\begin{pmatrix} x \\ y \end{pmatrix} = ax + by$.

2. Пусть V — пространство всех функций множества, состоящего из двух элементов $\{A, B\}$. Тогда правило $\boldsymbol{\alpha}^A : V \to \mathbb{R}$ приписывает каждому элементу $\mathbf{f} \in V$ его значение на элементе A, т. е. $\boldsymbol{\alpha}^A[\mathbf{f}] = \mathbf{f}(A)$ будет элементом множества V^*. В этом случае общий вид элемента из V^* можно записать в форме

$$\boldsymbol{\alpha}[\mathbf{f}] = a\mathbf{f}(A) + b\mathbf{f}(B)$$

для произвольных a и b. В этом примере интересно то, что мы идентифицировали A с $\boldsymbol{\alpha}^A$ и B с $\boldsymbol{\alpha}^B$. Выражение $aA+bB$ бессмысленно, а выражение $a\boldsymbol{\alpha}^A+b\boldsymbol{\alpha}^B$ определяет элемент пространства V^*. Таким образом, мы получили процедуру, связывающую векторное пространство с любым конечным множеством так, чтобы все элементы множества стали векторами. Для этого надо взять дуальное пространство к пространству функций на этом множестве. Такая конструкция оказывается очень полезной в теории электрических цепей.

3. Пусть V — пространство дифференцируемых функций $\mathbf{f}(t)$ на отрезке $[0,1]$. Тогда все выписанные ниже отображения будут элементами пространства V^*:

$$\begin{aligned}
\boldsymbol{\alpha}&: \ \mathbf{f} \mapsto \mathbf{f}(0),\\
\boldsymbol{\beta}&: \ \mathbf{f} \mapsto \mathbf{f}'(0),\\
\boldsymbol{\gamma}&: \ \mathbf{f} \mapsto \int_0^1 \mathbf{f}(t)\,dt,\\
\boldsymbol{\delta}&: \ \mathbf{f} \mapsto \int_0^1 t\mathbf{f}(t)\,dt,\\
\boldsymbol{\epsilon}&: \ \mathbf{f} \mapsto \mathbf{f}\left(\frac{1}{2}\right)+\mathbf{f}'\left(\frac{3}{4}\right)+\int_{1/3}^{2/3} t\mathbf{f}(t)\,dt.
\end{aligned}$$

10.3. Подпространства

Часто векторное пространство W является частью большего векторного пространства V, причем процедуры сложения и умножения на скаляр определены в них одинаково. Поскольку известно, что V удовлетворяет всем аксиомам векторного пространства, то нет необходимости проверять их для W. Чтобы показать, что подпространство W является векторным пространством, достаточно доказать, что W *замкнуто* по отношению к операциям сложения и умножения, т. е. что сумма двух любых элементов $\mathbf{w}_1, \mathbf{w}_2 \in W$, равная $\mathbf{w}_1+\mathbf{w}_2$, является элементом W, и для произвольного вещественного числа c и любого $\mathbf{w} \in W$ элемент $c\mathbf{w}$ принадлежит W. В частности, нулевой вектор должен быть элементом W.

На практике подпространство обычно определяется одним из двух методов: конкретизируется некоторое множество элементов из V, либо множество элементов из V^*.

Метод 1. Пусть $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ — векторы в пространстве V. Тогда множество W всех линейных комбинаций вида

$$\mathbf{w} = \sum_{i=1}^{k} c_i \mathbf{w}_i$$

образует подпространство пространства V. (Конечно, это может быть и все пространство V.)

Метод 2. Пусть $\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^k$ — элементы V^*. Тогда множество W элементов $\mathbf{v} \in V$, удовлетворяющих условиям

$$\begin{gathered} \boldsymbol{\alpha}^1[\mathbf{v}] = 0, \\ \boldsymbol{\alpha}^2[\mathbf{v}] = 0, \\ \vdots \\ \boldsymbol{\alpha}^k[\mathbf{v}] = 0, \end{gathered}$$

образует подпространство пространства V. Это доказывается очень просто. Пусть $\mathbf{w}_1$ и $\mathbf{w}_2$ — два вектора из множества W. Тогда вследствие линейности функций $\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots$ мы получаем

$$\boldsymbol{\alpha}^i[\mathbf{w}_1 + \mathbf{w}_2] = \boldsymbol{\alpha}^i[\mathbf{w}_1] + \boldsymbol{\alpha}^i[\mathbf{w}_2] = 0, \quad i = 1, 2, \ldots, k,$$

так что $\mathbf{w}_1 + \mathbf{w}_2 \in W$. Точно так же, если $\mathbf{w} \in W$, то

$$\boldsymbol{\alpha}^i[c\mathbf{w}] = c\,\boldsymbol{\alpha}^i[\mathbf{w}] = 0,$$

следовательно, $c\mathbf{w} \in W$. Таким образом, мы показали, что множество W замкнуто по отношению к операциям сложения и умножения на скаляр, т. е. является подпространством. (Конечно, это может быть $\{\mathbf{0}\}$ — нулевое подпространство, состоящее из единственного элемента — нулевого вектора.)

Хорошо известный пример применения этих двух методов дает построение плоскости, проходящей через начало координат в пространстве $\mathbb{R}^3$. Согласно методу 1, плоскость задается двумя векторами, например,

$$\begin{pmatrix}1\\1\\0\end{pmatrix} \quad \text{и} \quad \begin{pmatrix}0\\1\\2\end{pmatrix} \qquad \text{или} \qquad \begin{pmatrix}1\\2\\2\end{pmatrix} \quad \text{и} \quad \begin{pmatrix}-1\\0\\2\end{pmatrix}.$$

Согласно методу 2, плоскость описывается линейным уравнением, например,

$$2x - 2y + z = 0,$$

которое соответствует $\boldsymbol{\alpha}[\mathbf{w}] = 0$, где

$$\boldsymbol{\alpha} = (2, -2, 1) \quad \text{и} \quad \mathbf{w} = \begin{pmatrix}x\\y\\z\end{pmatrix}.$$

Рассмотрим еще один пример: пространство V образовано множеством полиномов, степень которых ≤ 2. Элемент этого множества имеет вид

$$f(t) = a + bt + ct^2.$$

Согласно методу 1, в качестве одномерного подпространства W можно рассматривать пространство всех полиномов, имеющих вид умноженной на любую константу функции $1 - t^2$. Согласно методу 2, это же подпространство можно определить двумя условиями

$$f(1) = 0 \quad \text{и} \quad f(-1) = 0,$$

или, более замаскированно, условиями

$$f'(0) = 0 \quad \text{и} \quad \int_0^{\sqrt{3}} f(t)\,dt = 0.$$

10.4. Размерность и базис

Чтобы двигаться дальше в изучении векторных пространств, нам необходимы понятия линейной зависимости и линейной независимости набора векторов. Говорят, что множество векторов

$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ *линейно зависимо*, если существуют такие вещественные числа $\lambda_1, \lambda_2, \ldots, \lambda_k$, не все равные нулю, что

$$\lambda \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_k \mathbf{v}_k = \mathbf{0}.$$

Если это равенство выполняется только для $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$, то такое множество векторов *линейно независимо.*

Приведем несколько примеров, иллюстрирующих эти важные понятия.

1. Пусть V — пространство $\mathbb{R}^3$. Рассмотрим

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix}.$$

Векторы, образующие набор $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, линейно зависимы, потому что

$$2\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix} = \mathbf{0}.$$

С другой стороны, набор векторов $\{\mathbf{v}_1, \mathbf{v}_2\}$ линейно независим, потому что

$$\lambda_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_1 + 2\lambda_2 \\ \lambda_2 \end{pmatrix}.$$

Очевидно, что результирующий вектор равен нулю, только если $\lambda_1 = 0$ и $\lambda_2 = 0$.

2. Пусть V — пространство функций на отрезке $[0, 2\pi]$. Рассмотрим

$$\mathbf{v}_1 = \cos^2 t, \quad \mathbf{v}_2 = \sin^2 t, \quad \mathbf{v}_3 = \cos 2t.$$

Этот набор линейно зависимый, потому что $\mathbf{v}_1 - \mathbf{v}_2 - \mathbf{v}_3 = \mathbf{0}$.

3. Пусть V — пространство функций на множестве $\{A, B\}$. Рассмотрим

$$\begin{aligned}
\mathbf{f}_1 &: \quad f_1(A) = 1, \quad & f_1(B) = 2, \\
\mathbf{f}_2 &: \quad f_2(A) = 2, \quad & f_2(B) = -3, \\
\mathbf{f}_3 &: \quad f_3(A) = -3, \quad & f_3(B) = 1.
\end{aligned}$$

Функции этого набора линейно зависимы, потому что $\mathbf{f}_1+\mathbf{f}_2+\mathbf{f}_3 = \mathbf{0}$ (т. е. мы получили нулевую функцию).

4. Пусть V — пространство полиномов степени ≤ 2. Рассмотрим следующие элементы пространства V^*:

$$\boldsymbol{\alpha}: \ \mathbf{f} \to \int_{-2}^{2} t\mathbf{f}(t)\,dt,$$
$$\boldsymbol{\beta}: \ \mathbf{f} \to \mathbf{f}'(0).$$

Написав $\mathbf{f}(t) = A + Bt + Ct^2$, мы получим

$$\boldsymbol{\alpha}[\mathbf{f}] = \int_{-2}^{2} (At + Bt^2 + Ct^3)\,dt = \frac{16}{3}B$$

и

$$\boldsymbol{\beta}[\mathbf{f}] = B.$$

Следовательно, $\left(\boldsymbol{\alpha} - \dfrac{16}{3}\boldsymbol{\beta}\right)[\mathbf{f}] = \mathbf{0}$, и набор векторов $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$ линейно зависим.

Из приведенных примеров, вероятно, стало понятным, что существуют ситуации, когда не очевидно, является ли некоторый набор векторов линейно зависимым или нет. Поэтому мы должны разработать процедуру исследования этого вопроса.

Мы говорим, что пространство V *натянуто* на множество векторов $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, если V состоит из векторов, которые можно записать как линейную комбинацию $\sum_{i=1}^{k} \mu_i \mathbf{v}_i$. (Набор векторов $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ может быть линейно зависимым, но тогда коэффициенты $\mu_1, \mu_2, \ldots, \mu_k$ определены неоднозначно.) Рассмотрим следующие примеры.

1. Пусть V — пространство $\mathbb{R}^3$. Очевидно, что набор векторов

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

порождает пространство $\mathbb{R}^3$, потому что любой элемент $\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}$

этого пространства может быть записан в виде

$$\begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\end{pmatrix} = \mu_1\begin{pmatrix}1\\0\\0\end{pmatrix} + \mu_2\begin{pmatrix}0\\1\\0\end{pmatrix} + \mu_3\begin{pmatrix}0\\0\\1\end{pmatrix}.$$

Менее очевидно, что набор векторов

$$\left\{\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix}\right\}$$

тоже порождает пространство $\mathbb{R}^3$, а вот набор

$$\left\{\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}\right\}$$

пространство $\mathbb{R}^3$ не порождает.

2. Пусть V — пространство функций $\mathbf{f}(t)$ на интервале $[0, \infty)$, удовлетворяющих дифференциальному уравнению

$$\mathbf{f}'' + 3\mathbf{f}' + 2\mathbf{f} = \mathbf{0}.$$

Векторы e^{-t} и e^{-2t} порождают пространство V, потому что общее решение уравнения имеет вид

$$\mathbf{f}(t) = Ae^{-t} + Be^{-2t}.$$

3. Пусть V — пространство функций вида $\mathbf{f}(t) = A + Bt^3$. Тогда векторы

$$\boldsymbol{\alpha}:\ \mathbf{f} \to \mathbf{f}(0),$$
$$\boldsymbol{\beta}:\ \mathbf{f} \to \int_{-1}^{1} t\mathbf{f}(t)\,dt$$

порождают дуальное пространство V^*. В самом деле,

$$\boldsymbol{\alpha}[\mathbf{f}] = A,$$
$$\boldsymbol{\beta}[\mathbf{f}] = \int_{-1}^{1} (At + Bt^4)\,dt = \frac{2}{5}B.$$

Однако любой элемент $\gamma \in V^*$ должен иметь вид $\gamma[\mathbf{f}] = aA + bB$, где a и b — некоторые постоянные. Таким образом, $\gamma = a\boldsymbol{\alpha} + \frac{5}{2}b\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$ определяют пространство V^*.

Пусть $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ — конечный линейно независимый набор векторов, порождающих векторное пространство V. Число векторов n называется *размерностью* пространства V. Чтобы установить, что размерность определяется однозначно, т.е. что все такие наборы для данного пространства содержат одинаковое количество элементов, мы должны доказать следующую теорему.

Теорема. *Пусть набор векторов* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ *порождает векторное пространство* V. *Тогда любой набор* $k+1$ *векторов в этом пространстве* V *линейно зависим.*

Теорему будем доказывать по индукции: сначала проверим ее для $k = 1$, потом покажем, что если теорема верна для пространства, натянутого на $k-1$ вектор, то она справедлива и для пространства, натянутого на k векторов.

Итак, пусть $k = 1$. Теорема в этом случае утверждает, что если V натянуто на один вектор $\mathbf{v}$, то любые два вектора в этом пространстве V линейно зависимы. Действительно, рассмотрим два таких вектора $\mathbf{w}_1$ и $\mathbf{w}_2$. Так как $\mathbf{v}$ порождает пространство V, то существуют такие вещественные числа μ_1 и μ_2, что $\mathbf{w}_1 = \mu_1\mathbf{v}$ и $\mathbf{w}_2 = \mu_2\mathbf{v}$. Тогда очевидно, что

$$\mu_2\mathbf{w}_1 - \mu_1\mathbf{w}_2 = \mu_2\mu_1\mathbf{v} - \mu_1\mu_2\mathbf{v} = \mathbf{0},$$

откуда следут линейная зависимость векторов $\mathbf{w}_1$ и $\mathbf{w}_2$.[2]

Теперь предположим, что теорема верна для любого набора k векторов в пространстве, натянутом на $k-1$ вектор. Рассмотрим набор $k+1$ векторов $\{\mathbf{w}_1, \ldots, \mathbf{w}_{k+1}\}$ в пространстве, определенном векторами $\mathbf{v}_1, \ldots, \mathbf{v}_k$. Мы можем написать $\mathbf{w}_1 = a_{11}\mathbf{v}_1 + a_{12}\mathbf{v}_2 + \cdots + a_{1k}\mathbf{v}_k$, потому что векторы $\{\mathbf{v}_i\}$ порождают пространство V. Если $\mathbf{w}_1 = \mathbf{0}$, то $r\mathbf{w}_1 = 0$, где $r \neq 0$, дает нетривиальную связь между векторами $\mathbf{w}_i$, и тогда все доказано. Поэтому предположим, что $\mathbf{w}_1 \neq 0$. В этом случае можно

[2] Определение линейной зависимости векторов $\{\mathbf{w}_1, \mathbf{w}_2\}$ требует, чтобы по крайней мере одно из чисел μ_1, μ_2 не равнялось нулю. На случае $\mu_1 = \mu_2 = 0$ авторы не останавливаются. Можно порекомендовать читателю рассмотреть его в качестве самостоятельного (очень простого) упражнения. — *Прим. ред.*

упорядочить векторы $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ так, что $a_{11} \neq 0$. Тогда

$$\frac{1}{a_{11}}\mathbf{w}_1 = \mathbf{v}_1 + \frac{a_{12}}{a_{11}}\mathbf{v}_2 + \cdots + \frac{a_{1k}}{a_{11}}\mathbf{v}_k.$$

Далее, пусть

$$\mathbf{w}_2 = a_{21}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + \cdots + a_{2k}\mathbf{v}_k.$$

Можно написать равенство

$$\mathbf{w}_2 - \frac{a_{21}}{a_{11}}\mathbf{w}_1 = \left(a_{22} - \frac{a_{21}a_{12}}{a_{11}}\right)\mathbf{v}_2 + \cdots + \left(a_{2k} - \frac{a_{21}a_{1k}}{a_{11}}\right)\mathbf{v}_k.$$

Аналогично, можно выразить $\mathbf{w}_3 - \dfrac{a_{31}}{a_{11}}\mathbf{w}_1$ и т. д. через $k-1$ вектор из набора $\mathbf{v}_2, \ldots, \mathbf{v}_k$. Но мы предположили, что теорема верна для $k-1$ векторов. Следовательно, набор k векторов

$$\left\{\mathbf{w}_2 - \frac{a_{21}}{a_{11}}\mathbf{w}_1,\ \mathbf{w}_3 - \frac{a_{31}}{a_{11}}\mathbf{w}_1, \ldots,\ \mathbf{w}_{k+1} - \frac{a_{(k+1)1}}{a_{11}}\mathbf{w}_1\right\}$$

линейно зависим. Это значит, что существуют такие константы $\lambda_2, \ldots, \lambda_{k+1}$, не все равные нулю, что

$$\lambda_2\left(\mathbf{w}_2 - \frac{a_{21}}{a_{11}}\mathbf{w}_1\right) + \lambda_3\left(\mathbf{w}_3 - \frac{a_{31}}{a_{11}}\mathbf{w}_1\right) + \cdots$$
$$+ \lambda_{k+1}\left(\mathbf{w}_{k+1} - \frac{a_{(k+1)1}}{a_{11}}\mathbf{w}_1\right) = \mathbf{0}.$$

Отсюда вытекает, что набор векторов $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{k+1}\}$ линейно зависим, что и требовалось доказать.

Теперь мы можем легко показать, что размерность векторного пространства определена однозначно. Предположим, что у нас есть набор линейно независимых векторов $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$, на который натянуто пространство V. Кроме того, есть еще один набор $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$, на который тоже натянуто это пространство. Из только что доказанной теоремы следует, что $n \leq k$. В противном случае векторы $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ будут линейно зависимы. С другой стороны, по той же теореме, $k \leq n$. Иначе векторы $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$

будут линейно зависимы. Следовательно, $k = n$, и любой конечный набор линейно независимых векторов, определяющих пространство V, содержит одинаковое число элементов. Это число и называется размерностью пространства V.

На самом деле в векторном пространстве V, имеющем размерность n, любой набор n линейно независимых векторов порождает это пространство. Действительно, пусть $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ — набор линейно независимых векторов в пространстве V, и пусть $\mathbf{w}$ — произвольный вектор, не равный нулю. Множество векторов $\{\mathbf{w}, \mathbf{v}_1, \ldots, \mathbf{v}_n\}$, состоящее их $n+1$ элементов, должно быть линейно зависимым. Следовательно, существуют такие постоянные $\lambda_0, \lambda_1, \ldots, \lambda_n$, что

$$\lambda_0 \mathbf{w} + \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n = \mathbf{0}.$$

В этом выражении константа λ_0 не может быть равна нулю, так как иначе векторы $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ будут линейно зависимы, что противоречит начальному предположению. Поэтому можно написать

$$\mathbf{w} = -\frac{1}{\lambda_0}(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n).$$

Таким образом, произвольный ненулевой вектор $\mathbf{w}$ выражен как линейная комбинация векторов $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, порождающих пространство.

Линейно независимый набор векторов $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, порождающий векторное пространство V, *записанный в определенном порядке*, называется *базисом* пространства V. Таким образом, наборы $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n\}$ и $\{\mathbf{v}_2, \mathbf{v}_1, \mathbf{v}_3, \ldots, \mathbf{v}_n\}$ являются разными базисами.

Рассмотрим векторы $\mathbf{w}_1, \ldots, \mathbf{w}_k$ $(k < n)$ в n-мерном векторном пространстве V. Мы всегда можем найти вектор $\mathbf{w}_{k+1}$, который не принадлежит подпространству $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$. Продолжая этот процесс $n-k$ раз, мы получим базис в пространстве V, который включает векторы $\mathbf{w}_1, \ldots, \mathbf{w}_k$. В частности, для данного векторного пространства V с размерностью n, содержащего подпространство W с размерностью k, мы всегда можем *построить такой базис для V, что его первые k векторов будут базисом подпространства W*. Эта процедура называется *расширением* базиса подпространства W до базиса пространства V.

Пусть в пространстве V выбран базис $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. Тогда любой элемент этого пространства может быть *единственным* образом записан как линейная комбинация базисных векторов:

$$\mathbf{v} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n.$$

Числа $x_1, \ldots, x_n$ называются *компонентами* вектора $\mathbf{v}$ в данном базисе. Чтобы показать единственность такого разложения, допустим, что $\mathbf{v}$ может быть записан по-другому:

$$\mathbf{v} = y_1\mathbf{e}_1 + y_2\mathbf{e}_2 + \cdots + y_n\mathbf{e}_n.$$

Вычтем из первого равенства второе и получим

$$\mathbf{0} = (x_1 - y_1)\mathbf{e}_1 + (x_2 - y_2)\mathbf{e}_2 + \cdots + (x_n - y_n)\mathbf{e}_n.$$

Поскольку базисные векторы линейно независимы, то

$$x_1 - y_1 = x_2 - y_2 = \cdots = x_n - y_n = 0,$$

что и доказывает единственность компонент вектора $\mathbf{v}$.

Из сказанного видно, что с помощью базиса в пространстве V можно определить изоморфизм L пространства V с $\mathbb{R}^n$, положив

$$L\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad L\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \text{и т. д.}$$

Обратно, если L — такой изоморфизм, то набор векторов

$$\mathbf{v}_1 = L^{-1}\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = L^{-1}\begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \text{и т. д.}$$

является базисом пространства V. Таким образом, мы можем отождествить базис $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ с соответствующим изоморфизмом L точно так же, как мы делали в главе 1 в двумерном случае.

Пусть $L : V \to \mathbb{R}^n$ и $L' : V \to \mathbb{R}^n$ — два базиса в одном n-мерном пространстве V.

$$\begin{tikzcd} & V \arrow[dl, "L"'] \arrow[dr, "L'=BL"] & \\ \mathbb{R}^n \arrow[rr, "B"'] & & \mathbb{R}^n \end{tikzcd}$$

Тогда $B = L' \circ L^{-1}$ — линейный изоморфизм $\mathbb{R}^n \to \mathbb{R}^n$, т. е. обратимая матрица $n \times n$. Эта матрица называется *матрицей замены базиса.*

Пусть V — векторное пространство размерности k, а W — векторное пространство размерности l, и пусть $T : V \to W$ — линейное преобразование. Предположим, что мы выбрали базисы в пространствах V и W. Тогда у нас есть изоморфизмы $L : V \to \mathbb{R}^k$ и $M : W \to \mathbb{R}^l$, и мы можем определить отображение

$$MTL^{-1} : \mathbb{R}^k \to \mathbb{R}^l.$$

Выражение MTL^{-1} можно рассматривать как матрицу, у которой l строк и k столбцов. Мы называем MTL^{-1} матрицей отображения T относительно базисов L и M и обозначаем ее $\mathrm{Mat}_{L,M}(T)$, т. е.

$$\mathrm{Mat}_{L,M}(T) = MTL^{-1}.$$

Эта ситуация может быть изображена на диаграмме

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "L"'] & W \arrow[d, "M"] \\ \mathbb{R}^k \arrow[r, "\mathrm{Mat}_{L,M}(T)"'] & \mathbb{R}^l \end{tikzcd}$$

Если мы выберем другие базисы: $L' = PL$ в пространстве V и $M' = QM$ в пространстве W, то

$$L'^{-1} = L^{-1}P^{-1}, \quad \text{так что} \quad M'TL'^{-1} = QMTL^{-1}P^{-1},$$

или

$$\mathrm{Mat}_{L',M'}(T) = Q(\mathrm{Mat}_{L,M}(T))P^{-1},$$

т. е. получена формула для замены базисов, из которой видно, как матричное представление линейного преобразования изменяется при такой замене.

10.5. Дуальный базис

Построив базис векторного пространства V, мы можем приступить к построению *дуального базиса* для дуального пространства V^*. Пусть $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ — базис в пространстве V. Тогда любой вектор $\mathbf{v} \in V$ может быть записан единственным образом

$$\mathbf{v} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n.$$

Пусть $\boldsymbol{\alpha}$ — элемент дуального пространства V^*. Поскольку $\boldsymbol{\alpha}$ — *линейная* функция в пространстве V, мы имеем

$$\boldsymbol{\alpha}[\mathbf{v}] = x_1\boldsymbol{\alpha}[\mathbf{e}_1] + x_2\boldsymbol{\alpha}[\mathbf{e}_2] + \cdots + x_n\boldsymbol{\alpha}[\mathbf{e}_n].$$

Это значит, что $\boldsymbol{\alpha}$ полностью определяется своими значениями на базисных векторах $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$. Введем в пространстве V^* систему векторов $\boldsymbol{\epsilon}^1, \ldots, \boldsymbol{\epsilon}^n$, задаваемых равенствами

$$\boldsymbol{\epsilon}^i[\mathbf{e}_j] = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Чтобы доказать линейную независимость этой системы, рассмотрим $\sum \lambda_i \boldsymbol{\epsilon}^i$. Применив эту линейную функцию к $\mathbf{e}_j$, получаем

$$\sum_{i=1}^{n} \lambda_i \boldsymbol{\epsilon}^i[\mathbf{e}_j] = \lambda_j.$$

Таким образом, если $\sum \lambda_i \boldsymbol{\epsilon}^i$ является нулевым элементом в V^*, то $\lambda_j = 0$ для всех j. Это доказывает, что набор векторов $\{\boldsymbol{\epsilon}^1, \ldots, \boldsymbol{\epsilon}^n\}$ линейно независим.

Для любого элемента $\boldsymbol{\alpha} \in V^*$ можно написать

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}[\mathbf{e}_1]\boldsymbol{\epsilon}^1 + \boldsymbol{\alpha}[\mathbf{e}_2]\boldsymbol{\epsilon}^2 + \cdots + \boldsymbol{\alpha}[\mathbf{e}_n]\boldsymbol{\epsilon}^n.$$

Очевидно, что обе части этого равенства имеют одинаковое значение для любого базисного вектора $\mathbf{e}_j$, т. е. являются одним и тем же элементом дуального пространства V^*. Это доказывает, что пространство V^* натянуто на векторы $\boldsymbol{\epsilon}^1, \ldots, \boldsymbol{\epsilon}^n$, которые, кроме того, линейно независимы. Из этого следует, что пространство V^*

имеет ту же размерность n, и набор векторов $\{\boldsymbol{\epsilon}^1, \ldots, \boldsymbol{\epsilon}^n\}$ является базисом в пространстве V^*.

Этот базис можно использовать, чтобы идентифицировать V^* с пространством $\mathbb{R}^{n*}$. Если мы разложим элемент $\boldsymbol{\alpha} \in V^*$ по дуальному базису

$$\boldsymbol{\alpha} = \lambda_1 \boldsymbol{\epsilon}^1 + \lambda_2 \boldsymbol{\epsilon}^2 + \cdots + \lambda_n \boldsymbol{\epsilon}^n,$$

то обнаружим, что удобно рассматривать элементы $\mathbb{R}^{n*}$ как векторы-строки. Тогда $\boldsymbol{\alpha}$ отождествляется с *вектором-строкой* $(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Преимущество такого обозначения в том, что действие $\boldsymbol{\alpha}$ на $\mathbf{v}$ описывается обычными правилами умножения матриц

$$\boldsymbol{\alpha}[\mathbf{v}] = (\lambda_1, \lambda_2, \ldots, \lambda_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n.$$

Важно не забывать, что эта техника применима только в случае, когда базис, определяющий соответствие между V^* и $\mathbb{R}^{n*}$, является дуальным к базису, определяющему соответствие между V и $\mathbb{R}^n$.

Предположим, что у нас есть n-мерное пространство V и k-мерное подпространство W. В пространстве V мы можем выбрать такой базис, что его первые k векторов образуют базис подпространства W:

$$\{\mathbf{v}_1, \ldots, \mathbf{v}_k; \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}.$$

После этого построим дуальный базис

$$\{\boldsymbol{\alpha}^1, \ldots, \boldsymbol{\alpha}^k; \boldsymbol{\alpha}^{k+1}, \ldots, \boldsymbol{\alpha}^n\},$$

так что

$$\boldsymbol{\alpha}^i[\mathbf{v}_j] = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Подпространство в V^* размерности $n - k$, натянутое на векторы $\{\boldsymbol{\alpha}^{k+1}, \ldots, \boldsymbol{\alpha}^n\}$, называется *аннулирующим пространством* для W и обозначается $W^\perp$. Если $\boldsymbol{\alpha} \in W^\perp$ и $\mathbf{w} \in W$, то $\boldsymbol{\alpha}[\mathbf{w}] = 0$, т. е. $W^\perp$ «аннулирует» подпространство W. Отсюда и происхождение этого термина. То, что мы раньше называли методом 2 для описания подпространства, по существу, было описанием аннулирущего пространства. Например, вектор (a, b, c) определяет одномерное подпространство $W^\perp$ дуального пространства $\mathbb{R}^{3*}$. Подпространство W пространства $\mathbb{R}^3$, аннулированное $W^\perp$, будет двумерным: это плоскость $ax + by + cz = 0$. Если мы выберем два независимых элемента дуального пространства $\mathbb{R}^{3*}$, например, (a_1, b_1, c_1) и (a_2, b_2, c_2), тогда подпространство $\mathbb{R}^3$, аннулированное ими, будет одномерным: это будет прямая, задаваемая системой уравнений

$$\begin{aligned} a_1x + b_1y + c_1z &= 0, \\ a_2x + b_2y + c_2z &= 0. \end{aligned}$$

Заметим, что аннулирующее пространство $W^\perp$ подпространства $W \subseteq V$ не зависит от конкретного выбора базиса пространства V. Мы вводили базис только для того, чтобы было удобно вычислить размерность $W^\perp$.

Пока нам недостает систематической процедуры вычисления размерности подпространства, натянутого на определенные элементы векторного пространства, или размерности подпространства, аннулированного конкретными элементами дуального векторного пространства. Эту процедуру, которая называется *редукцией по строкам*, мы опишем позднее.

10.6. Факторпространство

Продолжим рассмотрение n-мерного векторного пространства V, содержащего подпространство W с размерностью k.

Кажется естественным, что должно существовать подпространство с размерностью $n - k$, которое в некотором смысле будет «разницей» между V и W. Такое подпространство называется *факторпространством* и обозначается V/W. Однако, его элементы не будут элементами пространства V. Они являются множествами элементов пространства V и называются *классами эквивалентности*. Прежде чем дать определение этих классов, нам следует сначала увидеть, почему что-то более простое не будет достаточным.

В качестве конкретного примера пространства V с подпространством W рассмотрим плоскость $\mathbb{R}^2$ и прямую W на этой плоскости, как это изображено на рис. 10.1. «Разностью» между V и W могло бы быть множество элементов V, не входящих в W. Но ведь они порождают всю плоскость. На рис. 10.1 видно, что векторы $\mathbf{v}_1$ и $\mathbf{v}_2$, не лежащие на прямой W, порождают всю плоскость.

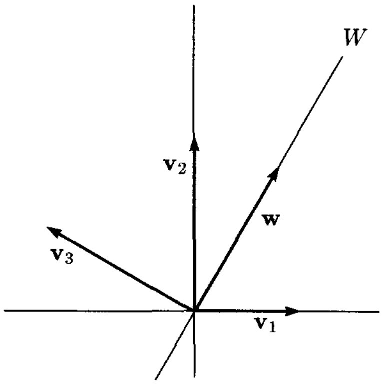


Рис. 10.1

Можно попробовать поступить по-другому — взять k базисных векторов подпространства W и расширить это множество до базиса пространства V. Тогда у нас добавится $n - k$ базисных векторов, не входящих в W. Возникает подпространство нужной размерности, но оно зависит от выбора базисных векторов и поэтому «плохо» определено. Например, на рис. 10.1 мы взяли $\mathbf{w}$ в качестве первого базисного вектора. После этого в качестве второго базисного вектора можно брать $\mathbf{v}_1$, $\mathbf{v}_2$ или $\mathbf{v}_3$. При этом мы каждый раз будем получать новое пространство. Если бы в пространстве V было определено скалярное произведение, то можно было бы выбрать подпространство, ортогональное к W. Но у нас нет скалярного произведения, поэтому и нет возможности определить процедуру предпочтительного получения второго базисного элемента.

Работающая конструкция дает определение *классов эквивалентности* (по модулю W) как множеств, каждое из которых состоит из векторов в пространстве V, разность между которыми лежит в W. Класс эквивалентности вектора $\mathbf{v}$ будем обозначать черточкой над вектором. Таким образом, $\overline{\mathbf{v}}$ обозначает множество векторов вида $\mathbf{v} + \mathbf{w}$, где $\mathbf{v}$ — заданный элемент множества V, а $\mathbf{w}$ — произвольный элемент W.

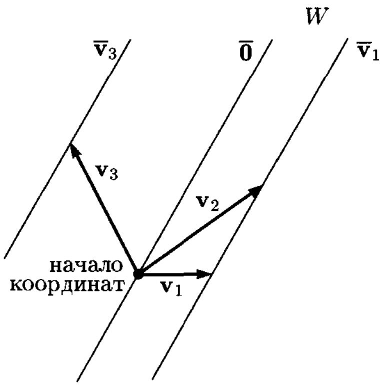


Рис. 10.2

Например, на рис. 10.2 видно, что $\overline{\mathbf{0}}$ (класс эквивалентности нулевого вектора) есть подпространство W, т. е. прямая, проходящая через начало координат. Векторы $\mathbf{v}_1$ и $\mathbf{v}_2$, отличающиеся на элемент подпространства W, принадлежат одному и тому же классу эквивалентности, который можно обозначать $\overline{\mathbf{v}}_1$ или $\overline{\mathbf{v}}_2$. Этот класс эквивалентности является прямой, не проходящей через начало координат. Класс эквивалентности $\overline{\mathbf{v}}_3$ будет другой прямой, которая тоже не проходит через начало координат. В этом случае классы эквивалентности составляют семейство прямых, параллельных W. В более общем случае мы можем считать подпространство W k-мерной *гиперплоскостью*[3], проходящей через начало координат n-мерного пространства V. Тогда классами эквивалентности по модулю W будет семейство гиперплоскостей, параллельных ей.

[3] В русскоязычной математической литературе термин «гиперплоскость» обычно относят только к случаю $k = n - 1$. — *Прим. ред.*

Чтобы ввести операцию сложения классов эквивалентности, сначала рассмотрим арифметику целых чисел по *модулю* 4, с которой читатель, возможно, уже знаком. В этом случае есть четыре класса эквивалентности:

$$\begin{aligned}
\overline{0} &= \{0, 4, -4, \; 8, -8, \dots\} = \{4n\}, \\
\overline{1} &= \{1, 5, -3, \; 9, -7, \dots\} = \{4n+1\}, \\
\overline{2} &= \{2, 6, -2, 10, -6, \dots\} = \{4n+2\}, \\
\overline{3} &= \{3, 7, -1, 11, -5, \dots\} = \{4n+3\}.
\end{aligned}$$

Чтобы сложить два класса эквивалентности, выбираем любое целое число из каждого класса, складываем их и получаем класс эквивалентности суммы. Например, чтобы сложить $\overline{2}$ и $\overline{3}$, выбираем число 6 из $\overline{2}$ и число 3 из $\overline{3}$. Складываем эти числа $6+3 = 9$ и видим, что сумма принадлежит классу эквивалентности $\overline{1}$. Таким образом, $\overline{2}+\overline{3} = \overline{1}$. Любой другой выбор, скажем, $-2+(-1) = -3$, дает тот же результат. Значит, процедура сложения классов эквивалентности определена корректно.

Сложение классов эквивалентности векторов по модулю подпространства W определяется аналогично. По определению, $\overline{\mathbf{v}}_1 + \overline{\mathbf{v}}_2 = \overline{(\mathbf{v}_1 + \mathbf{v}_2)}$, т. е. складываем любые два вектора из классов $\overline{\mathbf{v}}_1$ и $\overline{\mathbf{v}}_2$ и находим класс, к которому принадлежит сумма. Допустим, что мы взяли $\mathbf{v}_1 + \mathbf{w}_1$ из класса $\overline{\mathbf{v}}_1$ и $\mathbf{v}_2 + \mathbf{w}_2$ из класса $\overline{\mathbf{v}}_2$, где $\mathbf{w}_1$ и $\mathbf{w}_2$ — произвольные элементы подпространства W. Тогда сумма $\overline{\mathbf{v}}_1 + \overline{\mathbf{v}}_2$ есть класс эквивалентности, содержащий $(\mathbf{v}_1 + \mathbf{v}_2) + (\mathbf{w}_1 + \mathbf{w}_2)$, т. е. класс $\overline{(\mathbf{v}_1 + \mathbf{v}_2)}$ не зависит от выбора $\mathbf{w}_1$ и $\mathbf{w}_2$.

Процедура сложения изображена геометрически на рис. 10.3. Сумма $\overline{\mathbf{v}_1} + \overline{\mathbf{v}_2} = \overline{\mathbf{v}_3}$ независимо от того, какую пару векторов — $\mathbf{v}_1$ и $\mathbf{v}_2$ или $\mathbf{u}_1$ и $\mathbf{u}_2$ — мы брали в качестве представителей классов $\overline{\mathbf{v}_1}$ и $\overline{\mathbf{v}_2}$.

Аналогично определим операцию умножения класса эквивалентности на скаляр: $c\overline{\mathbf{v}}_1 = \overline{(c\mathbf{v}_1)}$, т. е. умножая произвольный элемент из класса $\overline{\mathbf{v}}_1$ на c, мы получаем класс эквивалентности произведения. Результат единственный, потому что W является подпространством.

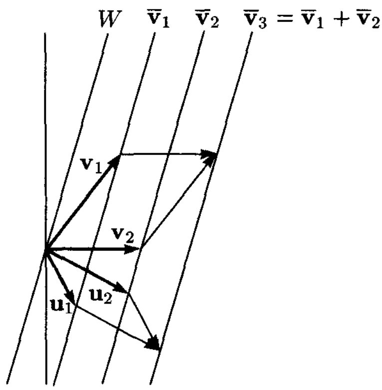


Рис. 10.3

Нетрудно проверить, что классы эквивалентности пространства V по модулю W образуют векторное пространство. Это пространство и называется *факторпространством* V/W.

Чтобы построить для него базис, возьмем базис подпространства W и расширим его до базиса во всем пространстве V. Итак, пусть базисными элементами V являются векторы $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ (элементы W) и $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-k}$ (не являются элементами W). Утверждается, что классы эквивалентности $\overline{\mathbf{v}}_1, \overline{\mathbf{v}}_2, \ldots, \overline{\mathbf{v}}_{n-k}$ образуют базис для V/W. Чтобы это доказать, мы должны показать, что они линейно независимы и на них натянуто пространство V/W.

Сначала рассмотрим проблему линейной независимости. Допустим, что $\overline{\mathbf{v}}_1, \overline{\mathbf{v}}_2, \ldots, \overline{\mathbf{v}}_{n-k}$ линейно зависимы. Тогда существуют такие константы $\lambda_1, \lambda_2, \ldots, \lambda_{n-k}$, что выполняется равенство

$$\lambda_1\overline{\mathbf{v}}_1 + \lambda_2\overline{\mathbf{v}}_2 + \cdots + \lambda_{n-k}\overline{\mathbf{v}}_{n-k} = \overline{\mathbf{0}}.$$

Значит,

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_{n-k}\mathbf{v}_{n-k} \in W,$$

что противоречит нашему предположению о линейной независимости векторов $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k; \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-k}\}$.

Любой вектор $\mathbf{v} \in V$ можно записать как сумму векторов

$$\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_{n-k}\mathbf{v}_{n-k} + \text{элемент из } W,$$

откуда следует, что

$$\overline{\mathbf{v}} = x_1\overline{\mathbf{v}}_1 + x_2\overline{\mathbf{v}}_2 + \dots + x_{n-k}\overline{\mathbf{v}}_{n-k}.$$

Это доказывает, что классы эквивалентности $\overline{\mathbf{v}}_1, \overline{\mathbf{v}}_2, \dots, \overline{\mathbf{v}}_{n-k}$ порождают пространство V/W. Таким образом, $\overline{\mathbf{v}}_1, \overline{\mathbf{v}}_2, \dots, \overline{\mathbf{v}}_{n-k}$ образуют базис V/W и

$$\dim\,(V/W) = \dim V - \dim W.$$

(Здесь и в дальнейшем dim обозначает размерность пространства.)

А теперь приведем несколько примеров факторпространств.

Пример 1. Пусть пространство V — это $\mathbb{R}^3$, а одномерное подпространство W натянуто на вектор $\begin{pmatrix}1\\1\\0\end{pmatrix}$. Векторы $\begin{pmatrix}1\\1\\0\end{pmatrix}$, $\begin{pmatrix}0\\1\\0\end{pmatrix}$ и $\begin{pmatrix}0\\0\\1\end{pmatrix}$ порождают пространство $\mathbb{R}^3$. Поэтому векторы $\overline{\begin{pmatrix}0\\1\\0\end{pmatrix}}$ и $\overline{\begin{pmatrix}0\\0\\1\end{pmatrix}}$ могут быть базисными векторами факторпространства V/W. Например,

$$\begin{pmatrix}3\\1\\1\end{pmatrix} = 3\begin{pmatrix}1\\1\\0\end{pmatrix} - 2\begin{pmatrix}0\\1\\0\end{pmatrix} + \begin{pmatrix}0\\0\\1\end{pmatrix}.$$

Следовательно,

$$\overline{\begin{pmatrix}3\\1\\1\end{pmatrix}} = -2\overline{\begin{pmatrix}0\\1\\0\end{pmatrix}} + \overline{\begin{pmatrix}0\\0\\1\end{pmatrix}}.$$

Класс эквивалентности любого вектора в пространстве $\mathbb{R}^3$ можно выразить как линейную комбинацию векторов $\overline{\begin{pmatrix}0\\1\\0\end{pmatrix}}$ и $\overline{\begin{pmatrix}0\\0\\1\end{pmatrix}}$.

Заметим, между прочим, что векторы $\overline{\begin{pmatrix}1\\0\\0\end{pmatrix}}$ и $\overline{\begin{pmatrix}0\\1\\0\end{pmatrix}}$ не могут быть базисом V/W. Сумма этих векторов является элементом W, поэтому они не будут линейно независимыми элементами факторпространства V/W:

$$\begin{pmatrix}1\\0\\0\end{pmatrix}+\begin{pmatrix}0\\1\\0\end{pmatrix}=\begin{pmatrix}1\\1\\0\end{pmatrix}, \quad \text{поэтому} \quad \overline{\begin{pmatrix}1\\0\\0\end{pmatrix}}+\overline{\begin{pmatrix}0\\1\\0\end{pmatrix}}=\overline{\mathbf{0}}.$$

Пример 2. Пусть V — пространство полиномов $\mathbf{f}(t)$ степени ≤ 2, и пусть двумерное подпространство W образовано полиномами, удовлетворяющими условию $\mathbf{f}(1) = 0$. Базисными полиномами W будут $\mathbf{f}_1 = 1 - t$ и $\mathbf{f}_2 = 1 - t^2$. Базисом факторпространства V/W будет класс эквивалентности $\overline{\mathbf{1}}$. В этом случае общий вид элемента пространства V — это

$$\mathbf{f}(t) = A + Bt + Ct^2,$$

или, что то же самое,

$$\mathbf{f}(t) = A + B + C - B(1 - t) - C(1 - t^2).$$

Это означает, что

$$\overline{\mathbf{f}(t)} = (A + B + C)\overline{\mathbf{1}}.$$

Если элементами факторпространства V/W считать плоскости, то подпространство W — это плоскость $\mathbf{f}(1) = 0$, и класс эквивалентности любой функции $\mathbf{f}(t)$ определяется значением $\mathbf{f}(1)$.

А теперь воспользуемся понятиями дуального пространства и факторпространства и получим очень важный результат. Раньше мы обнаружили, что если W является подпространством V, то аннулирующее пространство $W^{\perp}$ является подпространством дуального пространства V^*. Предположим, что $\boldsymbol{\alpha}$ — элемент $W^{\perp}$, и рассмотрим его действие на класс эквивалентности $\overline{\mathbf{v}}$. Поскольку $\boldsymbol{\alpha}[\mathbf{w}] = 0$ для всех $\mathbf{w} \in W$ и $\boldsymbol{\alpha}$ линейно, то

$$\boldsymbol{\alpha}[\mathbf{v} + \mathbf{w}] = \boldsymbol{\alpha}[\mathbf{v}].$$

Это значит, что $\boldsymbol{\alpha}$ имеет одинаковое значение для *любого* элемента класса эквивалентности и поэтому может рассматриваться как линейная функция в пространстве классов эквивалентности. И обратное утверждение: любая линейная функция в V/W может считаться линейной функцией в пространстве V. Определим $\boldsymbol{\beta}[\mathbf{v}]$ формулой

$$\boldsymbol{\beta}[\mathbf{v}] = \boldsymbol{\beta}[\overline{\mathbf{v}}].$$

Тогда $\boldsymbol{\beta}[\mathbf{w}] = \boldsymbol{\beta}[\overline{\mathbf{0}}] = \mathbf{0}$. Следовательно, $\boldsymbol{\beta}$ является элементом $W^{\perp}$. *Поэтому $W^{\perp}$ может быть отождествлено с дуальным пространством факторпространства V/W.* Вспомним, что $W^{\perp}$ и V/W имеют одинаковую размерность: $\dim V - \dim W$.

Аналогично можно рассмотреть факторпространство $V^*/W^{\perp}$, элементами которого являются классы эквивалентности $\overline{\boldsymbol{\beta}}$, элементы которых, в свою очередь, отличаются на элементы $W^{\perp}$, т. е.

$$\overline{\boldsymbol{\beta}} = \{\boldsymbol{\beta} + \boldsymbol{\alpha} : \boldsymbol{\alpha} \in W^{\perp}\}.$$

Определим действие $\overline{\boldsymbol{\beta}}$ на вектор $\mathbf{w} \in W$ согласно формуле $\overline{\boldsymbol{\beta}}[\mathbf{w}] = \boldsymbol{\beta}[\mathbf{w}]$. Это вполне законно, потому что для любого $\boldsymbol{\alpha} \in W^{\perp}$ мы имеем $(\boldsymbol{\beta} + \boldsymbol{\alpha})[\mathbf{w}] = \boldsymbol{\beta}[\mathbf{w}]$. Если $\overline{\boldsymbol{\beta}}[\mathbf{w}] = \mathbf{0}$ для всех $\mathbf{w} \in W$, то $\boldsymbol{\beta}[\mathbf{w}] = \mathbf{0}$ для всех $\mathbf{w} \in W$. А это значит, что $\boldsymbol{\beta} \in W^{\perp}$ или $\overline{\boldsymbol{\beta}} = \mathbf{0}$. Следовательно, $\overline{\boldsymbol{\beta}}$ полностью определено линейной функцией, заданной в подпространстве W.

Обратное утверждение: каждый элемент $\gamma \in W^*$ имеет вид $\gamma = \overline{\boldsymbol{\beta}}$ для некоторого $\boldsymbol{\beta} \in V^*$. Действительно, выберем базис $\mathbf{w}_1, \ldots, \mathbf{w}_k$ в подпространстве W и расширим его до базиса $\mathbf{w}_1, \ldots, \mathbf{w}_k, \mathbf{v}_1, \ldots, \mathbf{v}_{n-k}$ пространства V.

Пусть $\boldsymbol{\beta}$ — произвольная линейная функция, удовлетворяющая условию $\boldsymbol{\beta}[\mathbf{w}_i] = \gamma[\mathbf{w}_i]$ для всех i, и пусть $\boldsymbol{\beta}$ принимает любые значения для всех $\mathbf{v}_j$. Тогда $\overline{\boldsymbol{\beta}} = \gamma$. *Поэтому мы можем пространство этих классов эквивалентности, т. е. факторпространство $V^*/W^{\perp}$, отождествить с дуальным пространством W^*.*

Этот результат можно изобразить на следующей схеме:

$$\begin{aligned} V^*/W^{\perp} \leftarrow V^* \leftarrow W^{\perp}, \\ W \rightarrow V \ \ \rightarrow V/W. \end{aligned}$$

Здесь пространства, дуальные друг другу, расположены одно над другим: V и V^* (размерность n), W и $V^*/W^\perp$ (размерность k), V/W и $W^\perp$ (размерность $n-k$).

На этой теореме основывается много результатов линейной алгебры и ее применений в теории электрических цепей, что, безусловно, заслуживает более тщательного рассмотрения.

Приведем один пример. Пусть пространство V образовано полиномами $\mathbf{f}(t)$ степени ≤ 2, а двумерное подпространство W — *четными* полиномами. Тогда факторпространство V/W будет одномерным, базисный элемент, обозначаемый $\mathbf{h}_0$, — это класс эквивалентности функции $\mathbf{f}(t) = t$. Таким образом, если $\mathbf{f}(t) = A + Bt + Ct^2$, то $\overline{\mathbf{f}} = B\mathbf{h}_0$.

В этом случае аннулирующее пространство $W^\perp$ тоже одномерно. Базисный элемент может быть выбран в виде линейного оператора

$$\boldsymbol{\alpha}: \ \mathbf{f} \mapsto \int_{-1}^{1} t\mathbf{f}(t)\,dt.$$

Поскольку

$$\int_{-1}^{1} t(A + Bt + Ct^2)\,dt = \frac{2}{3}B,$$

то $\boldsymbol{\alpha}$ на самом деле аннулирует любой четный полином и приписывает значение $2/3$ полиному $\mathbf{f}(t) = t$, который определяет базисный элемент $\mathbf{h}_0$ факторпространства V/W. Таким образом, $3\boldsymbol{\alpha}/2$ — это базисный элемент, дуальный к $\mathbf{h}_0$.

Мы можем расширить $\boldsymbol{\alpha}$ до полного базиса V^*, добавив базисные элементы

$$\boldsymbol{\beta}_1: \ \mathbf{f} \mapsto \mathbf{f}(0),$$
$$\boldsymbol{\beta}_2: \ \mathbf{f} \mapsto \frac{1}{2}\mathbf{f}''(0),$$

действие которых состоит в выборе коэффициентов A и C соответственно, т. е.

$$\boldsymbol{\beta}_1[A + Bt + Ct^2] = A,$$
$$\boldsymbol{\beta}_2[A + Bt + Ct^2] = C.$$

Классы эквивалентности $\overline{\boldsymbol{\beta}_1}$ и $\overline{\boldsymbol{\beta}_2}$, образующие базис $V^*/W^\perp$, также образуют базис W^*, поскольку они являются дуальными базисными элементами для базисных элементов 1 и t^2 подпространства четных полиномов пространства W.

10.7. Линейные отображения

Рассмотрим линейное преобразование

$$A : V \to W,$$

где V — векторное пространство размерности n, а W — пространство размерности k. Как мы уже знаем, линейность отображения A означает, что $A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2$.

С линейным отображением $A : V \to W$ связаны два подпространства: *ядро* A и *образ* A.

Ядро A, в дальнейшем обозначаемое $\operatorname{Ker} A$, — это по определению множество всех таких векторов $\mathbf{v} \in V$, что $A\mathbf{v} = \mathbf{0}$. Проверим, что $\operatorname{Ker} A$ является подпространством пространства V. Заметим, что если $\mathbf{v}_1$ и $\mathbf{v}_2$ принадлежат ядру A, то $A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 = \mathbf{0}$, откуда следует, что $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \in \operatorname{Ker} A$. Это и значит, что ядро A является подпространством.

Образ A, в дальнейшем обозначаемый $\operatorname{Im} A$, — это множество всех векторов $\mathbf{w} \in W$, которые имеют вид $A\mathbf{v}$ для каких-либо $\mathbf{v} \in V$. Если $\mathbf{w}_1$ и $\mathbf{w}_2$ — два вектора из множества $\operatorname{Im} A$, то $\mathbf{w}_1 = A\mathbf{v}_1$ и $\mathbf{w}_2 = A\mathbf{v}_2$ для некоторых $\mathbf{v}_1, \mathbf{v}_2 \in V$. Поскольку отображение A линейно, мы имеем

$$A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 = c_1\mathbf{w}_1 + c_2\mathbf{w}_2,$$

откуда видно, что элемент $c_1\mathbf{w}_1 + c_2\mathbf{w}_2$ является элементом множества $\operatorname{Im} A$. Это доказывает, что A является подпространством пространства W.

Размерности пространств $\operatorname{Ker} A$ и $\operatorname{Im} A$, как мы сейчас докажем, связаны соотношением

$$\boxed{\dim(\operatorname{Im} A) + \dim(\operatorname{Ker} A) = \dim V.} \tag{10.8}$$

Размерность образа A называется *рангом* отображения A, а размерность ядра A — *дефектом* отображения A. Соотношение (10.8) называется *теоремой о ранге и дефекте*. Читатель уже знаком с этой теоремой в частном случае преобразований плоскости в себя. Вспомним, что в этом случае было три возможности для $A : \mathbb{R}^2 \to \mathbb{R}^2$:

1. Образом преобразования A является вся плоскость и отсутствует ненулевой вектор, который преобразуется в нулевой (ранг равен 2, дефект равен 0);

2. A преобразует плоскость в прямую, а прямую — в начало координат (ранг равен 1, дефект равен 1);

3. A переводит всю плоскость в начало координат (ранг равен 0, дефект равен 2).

Чтобы доказать эту теорему в общем случае в пространстве V, выберем удобный базис. Допустим, что $\dim V = n$ и $\dim(\operatorname{Ker} A) = k$. Выберем базис $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ для $\operatorname{Ker} A$ и расширим его до базиса всего пространства V, *выбрав* $r = n - k$ векторов $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$. Для удобства упорядочим этот базис в виде:

$$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r; \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\},$$

где *первые* r векторов *не* принадлежат подпространству $\operatorname{Ker} A$. Теперь надо показать, что r векторов вида $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}$ являются базисом $\operatorname{Im} A$.

Сначала покажем, что векторы $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}$ линейно независимы. Предположим противное:

$$\sum_{i=1}^{r} \lambda_i A\mathbf{v}_i = \mathbf{0}.$$

Преобразование A линейно, что означает

$$A\left(\sum_{i=1}^{r} \lambda_i \mathbf{v}_i\right) = \mathbf{0}.$$

Отсюда сразу же следует, что $\sum_{i=1}^{r} \lambda_i \mathbf{v}_i \in \operatorname{Ker} A$. Однако, векторы $\{\mathbf{v}_i\}$ вместе с векторами $\{\mathbf{u}_i\}$ образуют базис V. Поэтому из равенства

$$\sum_{i=1}^{r} \lambda_i \mathbf{v}_i = \mathbf{u}, \quad \text{где} \quad \mathbf{u} \in \operatorname{Ker} A,$$

следует, что все λ_i равны нулю, и вектора $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}$ линейно независимы.

Теперь надо показать, что вектора $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}$ порождают $\operatorname{Im} A$. Для этого рассмотрим произвольный вектор $\mathbf{w} \in \operatorname{Im} A$. Существует такой вектор $\mathbf{v} \in V$, для которого $\mathbf{w} = A\mathbf{v}$. Тогда мы можем написать

$$\mathbf{v} = \sum_{i=1}^{r} a_i \mathbf{v}_i + \sum_{j=1}^{k} b_j \mathbf{u}_j.$$

Однако, все базисные вектора $\mathbf{u}_j$ принадлежат множеству $\operatorname{Ker} A$, значит,

$$\mathbf{w} = A\left(\sum_{i=1}^{r} a_i \mathbf{v}_i + \sum_{j=1}^{k} b_j \mathbf{u}_j\right) = \sum_{i=1}^{r} a_i (A\mathbf{v}_i),$$

т. е. любой вектор $\mathbf{w}$ можем быть записан в виде линейной комбинации векторов $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}$.

Отсюда мы делаем вывод, что r векторов $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}$ образуют базис $\operatorname{Im} A$. Кроме того, $\dim(\operatorname{Im} A) = r = n - k$, где $\dim V = n$ и $\dim(\operatorname{Ker} A) = k$. Таким образом, $\dim(\operatorname{Im} A) = \dim V - \dim(\operatorname{Ker} A)$, что равносильно формуле (10.8).

Теорема о ранге и дефекте доказывает еще один факт об аннуляторе подпространства, о котором читатель, возможно, уже догадался. Предположим, что $\{\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^n\}$ — элементы пространства V^*. Определим линейное преобразование

$$A : V \to \mathbb{R}^n$$

формулой

$$A\mathbf{v} = \begin{pmatrix} \boldsymbol{\alpha}^1[\mathbf{v}] \\ \boldsymbol{\alpha}^2[\mathbf{v}] \\ \vdots \\ \boldsymbol{\alpha}^n[\mathbf{v}] \end{pmatrix}.$$

На векторы $\{\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^n\}$ натянуто подпространство $U^\perp \subset V^*$, которое аннулирует подпространство $\operatorname{Ker} A$. Доказанная нами теорема говорит, что $\dim(\operatorname{Im} A) = \dim V - \dim(\operatorname{Ker} A)$. Но в параграфе 10.4 мы показали, что

$$\dim(U^\perp) = \dim V - \dim(\operatorname{Ker} A),$$

откуда следует, что

$$\dim(\operatorname{Im} A) = \dim U^\perp.$$

На языке матриц каждый элемент $\boldsymbol{\alpha}^i$ является *строкой* матрицы и $\dim U^\perp$ есть размерность подпространства V^*, натянутого на *строки* матрицы. В тоже время $\dim(\operatorname{Im} A)$ есть размерность подпространства W, натянутого на *столбцы* матрицы. Обе эти размерности равны (обозначим их r), и число r называется *рангом* матрицы.

Представление строк матрицы как элементов дуального пространства V^* особенно полезно для решения системы линейных уравнений. Например, система уравнений

$$\begin{aligned} x + y + z &= 0, \\ x + 2y + 3z &= 0, \\ 2x + 3y + 4z &= 0 \end{aligned}$$

может быть записана в форме $A\mathbf{v} = \mathbf{0}$, где

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 2 & 3 & 4 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Здесь в *строках* матрицы A записаны коэффициенты каждого уравнения. Поскольку третье уравнение является суммой первых двух, то три строки определяют только двумерное подпространство пространства V^*, ранг матрицы A равен 2, и ее дефект (размерность ядра) равен $3 - 2 = 1$. Поэтому $\operatorname{Ker} A$ — одномерное подпространство, и существует одномерное подпространство решений уравнения $A\mathbf{v} = \mathbf{0}$.

10.8. Редукция по строкам

Рассмотрим линейное отображение

$$T : V \to W,$$

где V — векторное пространство размерности m, а W — пространство размерности n. Отображение T имеет ранг r. Мы уже видели, что при выборе базисов для V и W отображение T представляется в виде матрицы. В качестве базиса пространства V выберем векторы

$$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r; \mathbf{v}_{r+1}, \ldots, \mathbf{v}_m\},$$

где последние $m - r$ векторов являются базисом $\operatorname{Ker} T$. Тогда

$$T\mathbf{v}_{r+1} = T\mathbf{v}_{r+2} = \cdots = T\mathbf{v}_m = 0.$$

Следовательно, векторы $T\mathbf{v}_1$, $T\mathbf{v}_2, \ldots, T\mathbf{v}_r$ являются базисом $\operatorname{Im} T$. Теперь выберем векторы

$$\mathbf{w}_1 = T\mathbf{v}_1, \quad \mathbf{w}_2 = T\mathbf{v}_2, \quad \ldots, \quad \mathbf{w}_r = T\mathbf{v}_r$$

в качестве базиса $\operatorname{Im} T$ и расширим этот этот набор до базиса всего пространства W. В этих базисах матричное представление преобразования T имеет очень простой вид:

$$\underbrace{\left.\begin{pmatrix} \mathbb{I}_r & 0 \\ 0 & 0 \end{pmatrix}\right\}}_{m \text{ столбцов}} n \text{ строк} \tag{10.9}$$

В левом верхнем углу по диагонали этой матрицы стоят r единиц, а все остальные матричные элементы равны нулю.

Увы, обычно отображение T описывается матрицей A, которая представляет его в других, менее удобных базисах. Важная задача — найти такую замену базисов для пространств V и W, которое преобразует матрицу A, заданную в произвольном базисе, к $\mathbb{I}_r$. Наиболее эффективно эта задача решается с помощью *редукции по строкам*, что, по существу, является систематической процедурой решения линейных уравнений методом исключения неизвестных[4]. Сначала мы покажем процедуру на конкретном примере, а потом объясним, почему таким образом решается задача в общем случае.

Предположим, что нам дана матрица

$$M = \begin{pmatrix} 0 & 4 & -4 & 8 \\ 2 & 4 & 0 & 2 \\ 3 & 0 & 6 & -9 \end{pmatrix}.$$

Назовем *индексом* любой строки матрицы M, где не все элементы равны нулю, положение первого ненулевого элемента, называемого *ведущим* элементом этой строки. Для первой строки нашей матрицы M индекс равен 2, а ведущий элемент равен 4. Для третьей строки индекс равен 1, а ведущий элемент равен 3.

В качестве первого шага определим строку с минимальным индексом, переместив ее вверх матрицы, переставляя строки, и разделив ее на ведущий элемент. Для данной матрицы M мы переставляем первую и вторую строки:

$$\begin{pmatrix} 2 & 4 & 0 & 2 \\ 0 & 4 & -4 & 8 \\ 3 & 0 & 6 & -9 \end{pmatrix}$$

и делим верхнюю строку на ее ведущий элемент:

$$\begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 4 & -4 & 8 \\ 3 & 0 & 6 & -9 \end{pmatrix}.$$

В качестве второго шага мы должны «очистить» столбец под ведущим элементом верхней строки. Для этого первая строка умножается на соответствующий коэффициент и вычитается последовательно из нижних строк. В нашем случае мы умножаем первую строку на 3 и вычитаем из третьей строки:

$$\begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 4 & -4 & 8 \\ 0 & -6 & 6 & -12 \end{pmatrix}.$$

[4]Другое название этой процедуры, принятое в русскоязычной литературе, — метод Гаусса. — *Прим. ред.*

Теперь у матрицы есть ведущий элемент 1 в верхней строке, а все другие строки имеет более высокий индекс. Далее, передвигаем строку со следующим по величине индексом на вторую позицию и делим ее на ведущий элемент. В нашем примере вторая строка уже имеет следующий по величине индекс, поэтому нам остается разделить ее на ведуший элемент 4. В результате получаем

$$\begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & -1 & 2 \\ 0 & -6 & 6 & -12 \end{pmatrix}.$$

Теперь надо очистить столбец, соответствующий ведущему элементу второй строки, умножая вторую строку на соответствующий коэффициент и вычитая ее из других строк. В нашем случае мы умножаем вторую строку на 2 и вычитаем ее из первой строки, потом умножаем вторую строку -6 и вычитаем ее из третьей строки. Получаем матрицу

$$\begin{pmatrix} 1 & 0 & 2 & -3 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{10.10}$$

Теперь первый и второй столбцы содержат по единственному ненулевому элементу, равному 1, которые являются ведущими элементами соответствующих строк.

В общем случае мы снова переставляли бы строку со следующим по величине индексом на третью позицию и делили бы ее на ведущий элемент. Далее умножали бы третью строку на соответствующие коэффициенты и вычитали бы из остальных строк, чтобы очистить третий столбец. И так далее. В нашем случае больше нет ненулевых строк, и мы получили матрицу в редуцированной форме. Перечислим свойства редуцированной матрицы, которые можно увидеть из (10.10).

(a) Все нулевые строки, если они есть, находятся внизу.

(b) Ненулевые строки расположены в порядке возрастания индексов.

(c) Все элементы столбца, содержащего ведущий элемент (он равен 1) ненулевой строки, равны нулю, (кроме, естественно, этого ведущего элемента).

Каждая операция в процедуре редукции по строкам может быть получена умножением *слева* на обратимую матрицу $n \times n$. Например, умножение нашей матрицы слева на матрицу

$$S_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

приводит к перестановке первой и второй строк:

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 4 & -4 & 8 \\ 2 & 4 & 0 & 2 \\ 3 & 0 & 6 & -9 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 0 & 2 \\ 0 & 4 & -4 & 8 \\ 3 & 0 & 6 & -9 \end{pmatrix}.$$

Умножение слева на матрицу

$$S_2 = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

соответствует делению первой строки на 2:

$$\begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 4 & 0 & 2 \\ 0 & 4 & -4 & 8 \\ 3 & 0 & 6 & -9 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 4 & -4 & 8 \\ 3 & 0 & 6 & -9 \end{pmatrix}.$$

Умножение слева на матрицу

$$S_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}$$

соответствует трехкратному вычитанию первой строки из третьей:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 4 & -4 & 8 \\ 3 & 0 & 6 & -9 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 4 & -4 & 8 \\ 0 & -6 & 6 & -12 \end{pmatrix}.$$

Таким образом, редуцированная по строкам матрица B может быть записана в виде

$$B = S_k S_{k-1} \cdots S_3 S_2 S_1 M$$

или

$$B = SM,$$

где S — обратимая матрица $n \times n$. Заметим, что поскольку S обратима, $\dim \operatorname{Im} B = \dim \operatorname{Im} M$.

Образ и ядро редуцированной матрицы B легко определить. Очевидно, что образ является r-мерным подпространством, соответствующим r ненулевым строкам матрицы B и натянутым на столбцы

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots,$$

которые содержат ведущие элементы всех ненулевых строк. Согласно теореме о ранге и дефекте, ядро матрицы B имеет размерность $m - r$, равную числу столбцов, которые не содержат ведущих элементов. Базис для $\operatorname{Ker} B$ будем искать следующим образом. Возьмем векторы, имеющие 1 в одной из $m - r$ позиций, соответствующих столбцам с неведущими элементами, и нулю в остальных $m - r - 1$ позициях. Для нахождения первых r компонент этих векторов умножим поочередно строки матрицы B на эти векторы-столбцы, приравняем результаты к нулю и из получившихся уравнений найдем компоненты, соответствующие ведущим элементам столбцов.

Например, очевидно, что для матрицы

$$B = \begin{pmatrix} 1 & 0 & 2 & -3 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

базисом $\operatorname{Im} B$ являются вектора $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ и $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. Третий и четвертый столбцы не содержат ведущих элементов. Поэтому базисные вектора $\operatorname{Ker} B$ ищем в форме

$$\mathbf{u}_1 = \begin{pmatrix} x_1 \\ x_2 \\ 1 \\ 0 \end{pmatrix} \quad \text{и} \quad \mathbf{u}_2 = \begin{pmatrix} y_1 \\ y_2 \\ 0 \\ 1 \end{pmatrix}.$$

Положив $B\mathbf{u}_1 = \mathbf{0}$, получаем

$$x_1 + 2 = 0, \quad x_2 - 1 = 0.$$

Следовательно,

$$\mathbf{u}_1 = \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

Положив $B\mathbf{u}_2 = \mathbf{0}$, получаем

$$y_1 - 3 = 0, \quad y_2 + 2 = 0$$

и

$$\mathbf{u}_2 = \begin{pmatrix} 3 \\ -2 \\ 0 \\ 1 \end{pmatrix}.$$

Конечно, нам интересно знать ядро и образ первоначальной матрицы M, а не редуцированной по сторокам матрицы B. Однако, $B = SM$, где S — *обратимая* матрица. Тогда

$$M = S^{-1}B.$$

Из двух последних равенств видно, что любой вектор из ядра отображения B принадлежит ядру M. Поэтому *получив ядро B, тем самым, мы получаем ядро M*. Чтобы найти образ M, надо обратить S и подействовать матрицей S^{-1} на образ B. Итак, мы имеем $B = SM$, $\operatorname{Ker} B = \operatorname{Ker} M$ и $\dim \operatorname{Im} B = \dim \operatorname{Im} M$.

Предположим, что нам надо решить уравнение

$$M\mathbf{v} = \mathbf{w}.$$

Проделаем процедуру редукции по строкам с обеими частями этого уравнения:

$$SM\mathbf{v} = S\mathbf{w} \quad \text{или} \quad B\mathbf{v} = \mathbf{u},$$

где B — редуцированная по строкам матрица, а $\mathbf{u} = S\mathbf{w}$. Это уравнение имеет вид

$$\begin{pmatrix} 1 & 0 & 2 & -3 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}.$$

Оно может быть исследовано и решено следующим образом.

1. Если хотя бы одна компонента вектора $\mathbf{u}$, соответствующая нулевой строке B, отлична от нуля, то уравнение не имеет решения.

2. Если все компоненты вектора $\mathbf{u}$, соответствующие нулевым строкам B, равны нулю, то

$$\mathbf{v}_0 = \begin{pmatrix} u_1 \\ u_2 \\ 0 \\ 0 \end{pmatrix}$$

является частным решением уравнения.

3. Общее решение уравнения имеет вид $\mathbf{v}_0 + \mathbf{v}$, где $\mathbf{v} \in \operatorname{Ker} M$.

Решая конкретную задачу, прежде чем проводить редукцию по строкам, удобно объединить матрицу M и вектор $\mathbf{w}$ в одну таблицу так, чтобы проводить редукцию одновременно. Приведем пример. Нам надо решить уравнение

$$M\mathbf{v} = \mathbf{w},$$

где

$$M = \begin{pmatrix} 2 & 4 & 2 & 2 \\ 1 & 3 & 2 & 0 \\ 3 & 1 & -2 & 8 \end{pmatrix} \quad \text{и} \quad \mathbf{w} = \begin{pmatrix} 0 \\ -1 \\ 5 \end{pmatrix}.$$

Объединяем M и $\mathbf{w}$ в таблицу

$$\left(\begin{array}{cccc|c} 2 & 4 & 2 & 2 & 0 \\ 1 & 3 & 2 & 0 & -1 \\ 3 & 1 & -2 & 8 & 5 \end{array}\right)$$

и проводим редукцию, получая последовательно

$$\left(\begin{array}{cccc|c} 1 & 2 & 1 & 1 & 0 \\ 1 & 3 & 2 & 0 & -1 \\ 3 & 1 & -2 & 8 & 5 \end{array}\right),$$

$$\left(\begin{array}{cccc|c} 1 & 2 & 1 & 1 & 0 \\ 0 & 1 & 1 & -1 & -1 \\ 0 & -5 & -5 & 5 & 5 \end{array}\right),$$

$$\left(\begin{array}{cccc|c} 1 & 0 & -1 & 3 & 2 \\ 0 & 1 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Так мы получаем частное решение нашего уравнения

$$\begin{pmatrix} 2 \\ -1 \\ 0 \\ 0 \end{pmatrix}.$$

Чтобы найти общее решение, мы должны построить базис ядра M. Один базисный вектор имеет 1 на третьей позиции и 0 на четвертой, он равен $\begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}$. Другой имеет 0 на третьей позиции и 1 на четвертой, он равен $\begin{pmatrix} -3 \\ 1 \\ 0 \\ 1 \end{pmatrix}$. Поэтому решением уравнения $M\mathbf{v} = \mathbf{w}$ будет

$$\mathbf{v} = \begin{pmatrix} 2 \\ -1 \\ 0 \\ 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -3 \\ 1 \\ 0 \\ 1 \end{pmatrix},$$

где λ_1 и λ_2 — произвольные вещественные числа.

Отметим характерную форму решения относительно столбцов редуцированной матрицы, которые не содержат ведущих элементов строк (в нашем случае это третий и четвертый столбцы). Частное решение уравнения имеет нули на третьей и четвертой позициях. А вот базисные вектора ядра M имеют ненулевой элемент (единицу) ровно на одной из этих позиций. Общее решение уравнения можно записать многими способами, но этот вид — самый простой.

Для несингулярной квадратной матрицы A эта техника дает эффективный метод получения обратной матрицы. Преобразование S, которое редуцирует по строкам матрицу A к единичной, и есть A^{-1}. Поэтому обратная матрица может быть получена, если каждый отдельный шаг процедуры редукции применять к матрице, которая получается добавлением к матрице A справа единичной матрицы. Допустим, что $A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & 3 \\ -1 & -1 & 0 \end{pmatrix}$. Эту матрицу и единичную матрицу объединим в одну таблицу

$$\left(\begin{array}{rrr|rrr} 1 & 2 & 1 & 1 & 0 & 0 \\ 2 & 3 & 3 & 0 & 1 & 0 \\ -1 & -1 & 0 & 0 & 0 & 1 \end{array}\right).$$

Первую строку умножим на -2 и сложим со второй, потом прибавляем первую строку к третьей:

$$\left(\begin{array}{rrr|rrr} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & -2 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \end{array}\right).$$

Умножим вторую строку на -1:

$$\left(\begin{array}{rrr|rrr} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 2 & -1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \end{array}\right).$$

Вторую строку умножим на -2 и прибавим к первой, потом вычтем вторую строку из третьей:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 3 & -3 & 2 & 0 \\ 0 & 1 & -1 & 2 & -1 & 0 \\ 0 & 0 & 2 & -1 & 1 & 1 \end{array}\right).$$

Разделим третью строку на 2:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 3 & -3 & 2 & 0 \\ 0 & 1 & -1 & 2 & -1 & 0 \\ 0 & 0 & 1 & -1/2 & 1/2 & 1/2 \end{array}\right).$$

Третью строку умножим на -3 и сложим с первой, потом прибавим третью строку ко второй:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -3/2 & 1/2 & -3/2 \\ 0 & 1 & 0 & 3/2 & -1/2 & 1/2 \\ 0 & 0 & 1 & -1/2 & 1/2 & 1/2 \end{array}\right).$$

Отсюда следует, что матрица A^{-1} равна:

$$\frac{1}{2}\begin{pmatrix} -3 & 1 & -3 \\ 3 & -1 & 1 \\ -1 & 1 & 1 \end{pmatrix}.$$

Вернемся теперь к случаю прямоугольной матрицы. Вместо того, чтобы выполнять редукцию по строкам, мы можем выполнить редукцию по столбцам. Процедура точно такая же, только слово «строка» надо заменить на «столбец». В результате получим матрицу

$$C = MT,$$

где T — обратимая квадратная матрица, а C — матрица, редуцированная по столбцам. Отметим свойства матрицы, редуцированной по столбцам.

(a$'$) Все нулевые столбцы матрицы C, если они есть, расположены справа.

(b$'$) Ненулевые столбцы матрицы C расположены в порядке возрастания индекса. (Индекс ненулевого столбца равен номеру первого ненулевого элемента.)

(c′) Каждая строка, содержащая ведущий элемент ненулевого столбца, имеет ведущий элемент, равный 1, а все остальные элементы строки равны нулю.

Отметим, что
$$\operatorname{Im} C = \operatorname{Im} M,$$
и r ненулевых столбцов матрицы C будут линейно независимы. Следовательно, они образуют базис $\operatorname{Im} M$. Таким образом, подведем итоги.

> Чтобы найти базис $\operatorname{Im} M$, применяется редукция по столбцам. Базисом являются ненулевые столбцы получившейся матрицы C.
>
> Чтобы найти базис $\operatorname{Ker} M$, применяется редукция по строкам. Получившиеся строки редуцированной матрицы $B = SM$ дают систему r уравнений для $\operatorname{Ker} B$, с помощью которых затем и находится ядро M и его базис.

Мы можем также проделать редукцию и по строкам, и по столбцам. Например, допустим, что надо сделать редукцию по столбцам с матрицей, редуцированной по строкам:
$$B = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Первый столбец умножим на 2 и вычтем из третьего, потом первый столбец умножим на 3 и вычтем из четвертого. В результате получим
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Теперь второй столбец сложим с третьим, потом второй столбец умножим на 2 и прибавим к четвертому. Получаем матрицу
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

В общем случае, проводя редукцию по столбцам с матрицей B, редуцированной по строкам, мы делаем так, чтобы ведущими столбцами стали первые r столбцов. (В нашем примере этого не надо было делать.) Затем, умножая первые r столбцов на соответствующие коэффициенты и вычитая их из остающихся $m-r$ столбцов, мы получаем матрицу, у которой на главной диагонали стоят r единиц, а остальные элементы равны нулю, т. е. матрицу вида (10.9). Итак, мы описали эффективный метод получения таких матриц S и T, что матрица SMT имеет вид (10.9).

10.9. Локальная структура дифференцируемого отображения

Если собрать вместе результаты предыдущего параграфа с результатами параграфа 6.3, то можно получить очень важную информацию о структуре дифференцируемых отображений. Пусть V и W — векторные пространства размерности m и n соответственно, и пусть O — некоторая (открытая) область в пространстве V. Далее, предположим, что $f : O \to W$ — дифференцируемое отображение. В каждой точке $\mathbf{p} \in O$ мы можем вычислить дифференциал $df_{\mathbf{p}}$. Этот дифференциал является линейным отображением V на W, и поэтому мы можем вычислить его ранг. Конечно, этот ранг может зависеть от точки $\mathbf{p}$. Сейчас мы докажем следующую теорему.

Теорема о постоянном ранге. *Предположим, что существует такое постоянное целое число r, что ранг $df_{\mathbf{p}}$ равен r для всех $\mathbf{p} \in O$. Тогда для любого $\mathbf{x} \in O$ существует такое взаимно-однозначное дифференцируемое отображение ϕ, преобразующее окрестность $\mathbf{x}$ в некоторую окрестность в $\mathbb{R}^m$, такое, что ϕ имеет обратное дифференцируемое. Кроме того, можно найти такое взаимно-однозначное дифференцируемое отображение ψ, преобразующее окрестность $f(\mathbf{x})$ в пространстве W в некоторую окрестность в $\mathbb{R}^n$, причем оно также имеет обратное дифференцируемое. При этом композиционное отображение*

$$\psi \circ f \circ \phi^{-1} : \mathbb{R}^m \to \mathbb{R}^n$$

является линейным отображением с матрицей вида (10.9).

Короче говоря, эта теорема утверждает, что для дифференцируемых отображений постоянного ранга выполняется основная теорема о редукции по строкам: мы можем «сделать такие замены переменных», т. е. найти такие отображения ϕ и ψ, что выполняется равенство

$$(\psi \circ f \circ \psi^{-1})\left(\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}\right) = \begin{pmatrix} x_1 \\ \vdots \\ x_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Доказательство. Сделав трансляцию в пространстве V, мы можем считать, что $\mathbf{x} = \mathbf{0}$. Сделав еще одну трансляцию в пространстве W, можно предположить, что $f(\mathbf{x}) = \mathbf{0}$. Иными словами, чтобы упростить обозначения, можно считать, что $f(\mathbf{0}) = \mathbf{0}$, и мы интересуемся отображениями f в окрестности $\mathbf{0}$. Проделав редукцию по строкам, мы можем привести *линейное* отображение $df_{\mathbf{0}}$ к виду (10.9), т. е. мы можем найти обратимые линейные отображения $R : W \to \mathbb{R}^m$ и $S : V \to \mathbb{R}^n$, для которых

$$d(RfS^{-1})_{\mathbf{0}} = R\,df_{\mathbf{0}}S^{-1}$$

имеет вид (10.9). Тогда можно рассмотреть отображение

$$RfS^{-1} = \begin{pmatrix} f_1 \\ \vdots \\ f_r \\ f_{r+1} \\ \vdots \\ f_n \end{pmatrix},$$

где матрица

$$\left(\frac{\partial f_i}{\partial x_j}\right)_{\substack{i=1,\dots,r \\ j=1,\dots,r}}$$

есть единичная матрица в $\mathbf{0}$. Следовательно, она обратима в некоторой окрестности точки $\mathbf{0}$. Рассмотрим отображение g, определенное вблизи $\mathbf{0}$ в пространстве $\mathbb{R}^m$ формулой

$$g\left(\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}\right) = \begin{pmatrix} f_1 \\ \vdots \\ f_r \\ x_{r+1} \\ \vdots \\ x_m \end{pmatrix}.$$

Здесь первые r компонент отображения g — это введенные выше функции $f_1, \ldots, f_r$, а остальные $m-r$ компонент равны соответствующим координатам. Заметим, что матрица

$$\left(\frac{\partial g_i}{\partial x_j}\right) = \begin{array}{cc} \left(\begin{array}{c|ccc} \multicolumn{4}{c}{\dfrac{\partial f_i}{\partial x_j} \qquad \cdots} \\ \hline & 1 & & 0 \\ 0 & & \ddots & \\ & 0 & & 1 \end{array}\right) & \begin{array}{c} r \\ \text{строк} \\ \\ m-r \\ \text{строк} \end{array} \\ \begin{array}{cc} r & m-r \\ \text{столбцов} & \text{столбцов} \end{array} & \end{array}$$

обратима в $\mathbf{0}$ и, следовательно, обратима в окрестности $\mathbf{0}$.

А теперь применим теорему об обратной функции из параграфа 6.3 к отображению g. Еще раз рассмотрев то доказательство, мы видим, что оно работает для произвольных конечномерных пространств. (Заметим также, что сохраняется и доказательство теоремы о неявной функции, если x и y рассматривать как векторные переменные, а условие неравенства нулю производной $\partial g/\partial y$ (вместо нее теперь матрица) становится требованием обратимости матрицы.)

Вернемся к нашей задаче. Мы можем найти обратное преобразование для g. Пусть

$$h = (RfS^{-1}) \circ g^{-1}.$$

Тогда

$$(RfS^{-1})\left(\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}\right) = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$$

и

$$g\left(\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}\right) = \begin{pmatrix} f_1 \\ \vdots \\ f_r \\ x_{r+1} \\ \vdots \\ x_m \end{pmatrix}.$$

Следовательно,

$$h\left(\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}\right) = \begin{pmatrix} y_1 \\ \vdots \\ y_r \\ g_{r+1}(y) \\ \vdots \\ g_n(y) \end{pmatrix}$$

при некоторых функциях $g_{r+1}, \ldots, g_n$, определенных в пространстве $\mathbb{R}^m$ в окрестности $\mathbf{0}$. Воспользуемся цепным правилом и получим

$$dh_{\mathbf{q}} = d(Rf_{\mathbf{p}}S^{-1}) \circ dg_{\mathbf{q}}^{-1}, \quad \text{где} \quad \mathbf{q} = g(S\mathbf{p}).$$

Отсюда получаем, что в некоторой окрестности $\mathbf{0}$ ранг $dh_{\mathbf{q}}$ тождественно равен r. Рассмотрев матрицу

$$\left(\frac{\partial h_i}{\partial y_j}\right) = \begin{array}{cc} \left(\begin{array}{c|c} \mathbb{I}_r & 0 \\ \hline \vdots & \dfrac{\partial g_i}{\partial y_j} \end{array}\right) & \begin{array}{c} r \text{ строк} \\[2em] n-r \text{ строк} \end{array} \\ \begin{array}{cc} r \text{ столбцов} & n-r \text{ столбцов} \end{array} & \end{array},$$

мы видим, что это может случиться, только если все частные производные, появляющиеся в правом нижнем углу, тождественно обращаются в нуль. Это значит, что последние $n-r$ компонент отображения h зависят только от *первых* r координат y, т.е.

$$g_{r+1} = g_{r+1}(y_1, \ldots, y_r, 0, \ldots, 0), \quad \text{и т.д.}$$

Введем преобразование H в пространстве $\mathbb{R}^n$ согласно формуле

$$H\left(\begin{pmatrix} z_1 \\ \vdots \\ z_r \\ z_{r+1} \\ \vdots \\ z_n \end{pmatrix}\right) = \begin{pmatrix} z_1 \\ \vdots \\ z_r \\ z_{r+1} - g_{r+1}(z_1, \ldots, z_r, 0, \ldots, 0) \\ \vdots \\ z_n - g_n(z_1, \ldots, z_r, 0, \ldots, 0) \end{pmatrix}.$$

Это гладкое отображение, определенное в окрестности $z = 0$ в $\mathbb{R}^n$. Очевидно, что оно обратимо — обратное преобразование определяется формулой

$$H^{-1}\left(\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}\right) = \begin{pmatrix} w_1 \\ \vdots \\ w_r \\ w_{r+1} + g_{r+1}(w_1, \ldots, w_r, 0, \ldots, 0) \\ \vdots \\ w_n + g_n(w_1, \ldots, w_r, 0, \ldots, 0) \end{pmatrix}.$$

Тогда

$$(H \circ h)\left(\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}\right) = \begin{pmatrix} y_1 \\ \vdots \\ y_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Подставив определение $h = (RfS^{-1}) \circ g^{-1}$ в $H \circ h$, получаем

$$H \circ h = (HR) \circ f \circ (S^{-1}g^{-1}).$$

Определяя

$$\psi = H \circ R$$

и

$$\phi = g \circ S,$$

видим, что $\psi \circ f \circ \phi^{-1}$ имеет требуемую теоремой форму

$$\psi \circ f \circ \phi^{-1}\left(\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}\right) = \begin{pmatrix} y_1 \\ \vdots \\ y_r \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

что и требовалось доказать.

Наиболее важное применение этой теоремы относится к случаю $r = n$ (напомним, что n — это размерность пространства W). Если дифференциал $df_{\mathbf{p}}$ непрерывен в некоторой точке $\mathbf{p}$ и имеет ранг n, то утверждается, что он имеет ранг n во всех точках, достаточно близких к $\mathbf{p}$. Действительно, с помощью редукции по строкам мы можем найти такие R и S, что $Rdf_{\mathbf{p}}S^{-1}$ имеет вид (10.9) с $r = n$. Если $\mathbf{q}$ достаточно близко к $\mathbf{p}$, то верхний левый блок матрицы $Rdf_{\mathbf{p}}S^{-1}$ будет близок к единичной матрице. Следовательно, размерность образа $df_{\mathbf{q}}$ не меньше n. Поскольку размерность этого образа не может быть больше $\dim W = n$, мы делаем вывод, что ранг дифференциала $df_{\mathbf{q}}$ тождественно равен n для всех $\mathbf{q}$ в некоторой окрестности $\mathbf{p}$. Из теоремы о постоянном ранге мы делаем следующий вывод.

Теорема о множестве решений. *Предположим, что* $f : O \to W$ *— непрерывно дифференцируемое отображение и что* $df_{\mathbf{p}}$ *сюръективно в точке* $\mathbf{p}$, *т. е. что ранг* $df_{\mathbf{p}}$ *равен* $\dim W$. *Тогда мы можем найти дифференцируемые отображения:* ϕ, *отображающее окрестность* $\mathbf{p}$ *в* $\mathbb{R}^m$, *и* ψ, *отображающее окрестность* $f(\mathbf{p})$ *в* $\mathbb{R}^n$, *так что оба отображения имеют непрерывно-дифференцируемые обратные, причем* $\phi(\mathbf{p}) = \mathbf{0}$ *в* $\mathbb{R}^m$, $\psi(f(\mathbf{q})) =$

0 *в* $\mathbb{R}^n$ *и локально выполняется равенство*

$$\psi f \phi^{-1}\left(\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}\right) = \begin{pmatrix} x_1 \\ \vdots \\ x_m \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

В частности, точка **x** *является решением уравнения*

$$f(\mathbf{x}) = f(\mathbf{p}) \quad (\mathbf{x} \text{ вблизи } \mathbf{p})$$

тогда и только тогда, когда

$$\mathbf{x} = \psi^{-1}\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \\ y_{m+1} \\ \vdots \\ y_n \end{pmatrix}\right),$$

причем все $y_{m+1}, y_{m+2}, \ldots, y_n$ *лежат в окрестности* **0**.

Введем несколько терминов, чтобы сделать теорему о множестве решений более краткой. Пусть H_{m-n} обозначает подпространство $\mathbb{R}^m$, определенное условиями

$$x_1 = 0, \quad \ldots, \quad x_n = 0.$$

(Здесь предполагается, что n должно быть $\leq m$.) Подмножество M m-мерного векторного пространства V называется *подмногообразием коразмерности* n, если оно обладает следующим свойством: около каждой точки $\mathbf{x} \in M$ мы можем найти окрестность O в пространстве V и дифференцируемое отображение $\phi : O \to \mathbb{R}^m$ с $\phi(\mathbf{x}) = \mathbf{0}$, так что ϕ взаимно-однозначным образом отображает O в окрестность U вблизи **0**, при этом ϕ^{-1} дифференцируемо и

$$\phi(M \cap O) = U \cap H_{m-n}.$$

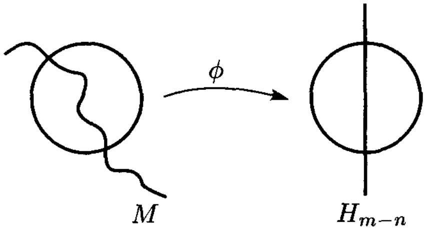


Рис. 10.4

Другими словами, это условие на M означает, что вблизи каждой точки $\mathbf{x} \in M$ мы можем найти такую гладкую деформацию всей окрестности, задаваемую отображением ϕ, что M «уплощается» и становится куском гиперплоскости. Например, окружность $x^2 + y^2 = 1$ является подмногообразием коразмерности 1 в пространстве $\mathbb{R}^2$, потому что мы можем ввести отображение

$$\phi(x, y) = \begin{pmatrix} r^2 - 1 \\ \arctan(y/x) \end{pmatrix}$$

во всех ее точках с $x \neq 0$ (с последующим сдвигом в вертикальном направлении к началу координат образа). В точках, где $x = 0$, мы можем пользоваться функцией $\arctan(x/y)$. А вот периметр квадрата не является подмногообразием, потому что не существует гладкого способа выпрямить углы.

Пусть $f : O \to W$ — непрерывно-дифференцируемое отображение. Точка $\mathbf{p} \in O$ называется *регулярной точкой* отображения f, если $df_{\mathbf{p}}$ сюръективно, т. е. если ранг $df_{\mathbf{p}}$ равен $\dim W$. Точка, которая не является регулярной, называется критической точкой. Если $W = \mathbb{R}$, то $\mathbf{p}$ будет критической точкой при условии $df_{\mathbf{p}} = 0$. Это согласуется с нашими предыдущими обозначениями.

Точка $\mathbf{q} \in W$ называется *регулярным значением*, если все точки $\mathbf{p}$ в $f^{-1}(\mathbf{q})$ являются регулярными точками. Теперь дадим еще одну формулировку нашей теоремы.

Если $\mathbf{q}$ — регулярное значение f, то $f^{-1}(\mathbf{q})$ — подмногообразие.

Приведем несколько примеров.

(a) Пусть $n = 1$ и $f : \mathbb{R}^m \to \mathbb{R}^1 = \mathbb{R}$ определено формулой

$$f\left(\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}\right) = x_1^2 + \cdots + x_m^2.$$

Тогда $df_{\mathbf{p}} \neq 0$, если $\mathbf{p} \neq \mathbf{0}$. Таким образом, любое ненулевое значение f будет регулярным значением. Для $c > 0$, $f^{-1}(c)$ есть сфера радиуса $\sqrt{c}$. Значит, сферы являются подмногообразиями.

(b) Пусть $M(k)$ обозначает векторное пространство всех матриц $k \times k$, тогда $m = \dim V = k^2$. Пусть $W = S(R)$ обозначает векторное пространство всех симметричных матниц $k \times k$, т. е. матриц D, удовлетворяющих условию $D^t = D$. Тогда

$$n = \dim W = \frac{1}{2}k(k+1).$$

Рассмотрим отображение $f : V \to W$, где $f(A) = AA^t$. Легко проверить, что $df_A(B) = BA^T + AB^t$. Мы утверждаем, что единичная матрица $\mathbb{I}$ является регулярным значением отображения f. Для доказательства этого утверждения следует проверить, что df_A сюръективно, если $AA^t = \mathbb{I}$. Другими словами, при каждом A уравнение

$$BA^t + AB^t = C$$

относительно B должно иметь решение для любой симметричной матрицы C. Действительно, пусть $B = CA/2$. Тогда

$$\begin{aligned} BA^t + AB^t &= \frac{1}{2}CAA^t + \frac{1}{2}AA^tC^t \\ &= \frac{1}{2}C + \frac{1}{2}C^t, && \text{т. к.} \quad AA^t = A^tA = \mathbb{I} \\ &= C, && \text{т. к.} \quad C = C^t. \end{aligned}$$

Итак, множество всех ортогональных матриц, т. е. матриц, удовлетворяющих условию $AA^t = \mathbb{I}$, является подмногообразием пространства всех матриц.

10.10. Сопряженное отображение

Изучая отображения одной аффинной плоскости в другую, мы пользовались понятием переноса функции (pullback). Вспомним, что при отображении ϕ аффинной плоскости A в другую аффинную плоскость B естественным образом возникало *линейное* отображение функций, определенных на плоскости B, в функции, определенные на плоскости A. Эта ситуация изображена на рис. 10.5. Перенос определенной на плоскости B функции f, обозначаемый $\phi^* f$, является функцией на плоскости A, определяемой соотношением

$$(\phi^* f)(P) = f(\phi(P)).$$

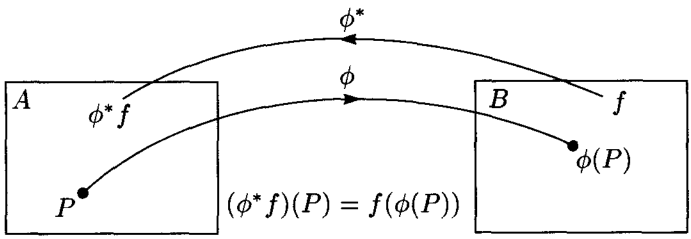


Рис. 10.5

Понятие переноса может быть расширено на случай отображений одного векторного пространства V в другое векторное пространство W. Нам особенно интересен перенос линейных функций в пространстве W (т. е. элементов дуального пространства W^*), возникающий как следствие *линейного* отображения A пространства V в W. В этом случае перенесенное отображение из W^* в V^* называется отображением, *сопряженным* к A. Мы будем обозначать его A^*. Его действие на элемент $\boldsymbol{\beta} \in W^*$ определяется формулой

$$(A^*\boldsymbol{\beta})[\mathbf{v}] = \boldsymbol{\beta}[A\mathbf{v}].$$

Доказательство того, что таким образом определенное отображение A^* линейно по $\boldsymbol{\beta}$, точно такое же, как доказательство того, что вообще перенос — линейное отображение. Если элементы $\boldsymbol{\beta}^1$ и $\boldsymbol{\beta}^2$ являются элементами W^*, то

$$\begin{aligned}(A^*(c_1\boldsymbol{\beta}^1 + c_2\boldsymbol{\beta}^2))[\mathbf{v}] &= (c_1\boldsymbol{\beta}^1 + c_2\boldsymbol{\beta}^2)[A\mathbf{v}]\\ &= c_1\boldsymbol{\beta}^1[A\mathbf{v}] + c_2\boldsymbol{\beta}^2[A\mathbf{v}]\\ &= c_1 A^*\boldsymbol{\beta}^1[\mathbf{v}] + c_2 A^*\boldsymbol{\beta}^2[\mathbf{v}].\end{aligned}$$

Отсюда получаем равенство

$$A^*(c_1\boldsymbol{\beta}^1 + c_2\boldsymbol{\beta}^2) = c_1 A^*\boldsymbol{\beta}^1 + c_2 A^*\boldsymbol{\beta}^2,$$

которое и означает линейность A^*. Отметим, что из линейности A и $\boldsymbol{\beta}$ следует линейность функции $A^*\boldsymbol{\beta}$ по $\mathbf{v}$, т. е. что действительно $A^*\boldsymbol{\beta}$ находится в пространстве V^*.

Важно отметить, что сопряженное отображение A^* действует «в направлении, противоположном» A. Это значит, что если A преобразует вектор $\mathbf{v} \in V$ в вектор $A\mathbf{v} \in W$, то сопряженное отображение A^* преобразует вектор $\boldsymbol{\beta} \in W^*$ в вектор $A^*\boldsymbol{\beta} \in V^*$. На схеме это выглядит так:

$$\begin{gathered} V^* \xleftarrow{A^*} W^*,\\ V \xrightarrow{A} W. \end{gathered}$$

Рассмотрим пример сопряженного отображения: пусть V — трехмерное пространство четных полиномов степени ≤ 4 и пусть W — двумерное пространство нечетных полиномов степени ≤ 3. Тогда операция дифференцирования определяет линейное отображение D из V в W:

$$D: \quad \mathbf{f} \mapsto \mathbf{f}'.$$

Пусть элемент пространства W^* таков:

$$\boldsymbol{\beta}: \quad \mathbf{g} \mapsto \mathbf{g}(1).$$

Например, $\boldsymbol{\beta}(t + 2t^3) = 1 + 2 = 3$. Чтобы вычислить $D^*\boldsymbol{\beta}$, воспользуемся определением

$$D^*\boldsymbol{\beta}[\mathbf{f}] = \boldsymbol{\beta}[D\mathbf{f}].$$

В нашем случае $D\mathbf{f} = \mathbf{f}'$ и $\boldsymbol{\beta}[D\mathbf{f}] = \mathbf{f}'(1)$. Отсюда следует, что $D^*\boldsymbol{\beta}$ — линейная функция в пространстве V (элемент пространства V^*):

$$D^*\boldsymbol{\beta}: \quad \mathbf{f} \to \mathbf{f}'(1).$$

Возьмем другой элемент пространства W^*:

$$\boldsymbol{\alpha}: \quad \mathbf{g} \mapsto \int_0^1 \mathbf{g}(t)\,dt.$$

Тогда

$$D^*\boldsymbol{\alpha}[\mathbf{f}] = \boldsymbol{\alpha}[D\mathbf{f}] = \int_0^1 \mathbf{f}'(t)\,dt = \mathbf{f}(1) - \mathbf{f}(0),$$

т. е. $D^*\boldsymbol{\alpha}$ — линейная функция

$$D^*\boldsymbol{\alpha}: \quad \mathbf{f} \mapsto \mathbf{f}(1) - \mathbf{f}(0).$$

Если ввести базисы в пространствах V, W, V^* и W^*, то описание сопряженного отображения становится особенно простым. Предположим, что V — m-мерное пространство с базисом $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$. Введем дуальный базис $\{\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^m\}$ в дуальном пространстве V^*. Аналогично, пусть $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ и $\{\boldsymbol{\beta}^1, \boldsymbol{\beta}^2 \ldots, \boldsymbol{\beta}^n\}$ будут базисами пространств W и W^* соответственно. Если A — линейное отображение из V в W, то

$$A\mathbf{v}_i = \sum_{j=1}^{n} a_{ji}\mathbf{w}_j,$$

где величины $\{a_{1i}, a_{2i}, \ldots, a_{ni}\}$ образуют i-й *столбец* матрицы, задающей отображение A. Вычислим, как A^* действует на базисный элемент пространства W^*:

$$(A^*\boldsymbol{\beta}^k)[\mathbf{v}_i] = \boldsymbol{\beta}^k[A\mathbf{v}_i] = \boldsymbol{\beta}^k\left[\sum_{j=1}^{n} a_{ji}\mathbf{w}_j\right] = \sum_{j=1}^{n} a_{ji}\boldsymbol{\beta}^k[\mathbf{w}_j].$$

Поскольку базис $\{\boldsymbol{\beta}^1, \boldsymbol{\beta}^2 \ldots, \boldsymbol{\beta}^n\}$ дуален базису $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$, мы имеем

$$\boldsymbol{\beta}^k[\mathbf{w}_j] = \begin{cases} 1, & j = k, \\ 0, & j \neq k, \end{cases}$$

т. е.

$$(A^*\boldsymbol{\beta}^k)[\mathbf{v}_i] = a_{ki}.$$

Итак, мы можем выразить $A^*\boldsymbol{\beta}^k$ через базисные вектора $\{\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^m\}$ в пространстве V^*:

$$A^*\boldsymbol{\beta}^k = \sum_{l=1}^{n} a_{kl}\boldsymbol{\alpha}^l.$$

Величины $\{a_{kl}\} = \{a_{k1}, a_{k2}, \ldots, a_{km}\}$ образуют k-й *столбец* матрицы, задающей отображение A^*, но одновременно они являются k-й строкой матрицы, задающей A. Это значит, что матрица отображения A^* является *транспонированной* матрицей отображения A. Например, если V — трехмерное пространство, W — двумерное и $A : V \to W$ представляется матрицей

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix},$$

то сопряженное отображение $A^* : W^* \to V^*$ задается матрицей

$$A^* = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix}.$$

Вспомним, что элементы дуального пространства могут быть записаны в виде векторов-строк. В нашем примере элемент пространства W^* можно записать в виде двухкомпонентного вектора-строки:

$$\boldsymbol{\beta} = (\lambda_1, \lambda_2),$$

а элемент пространства V — как трехкомпонентный вектор-столбец:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}.$$

Тогда $(A^*\boldsymbol{\beta})[\mathbf{v}] = \boldsymbol{\beta}[A\mathbf{v}]$ можно записать как произведение

$$(\lambda_1, \lambda_2)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}.$$

Естественное истолкование этого произведения: матрица A, действуя сначала на столбец, расположенный от нее справа, дает $A\mathbf{v}$, на который уже действует $\boldsymbol{\beta}$. С другой стороны, можно думать, что матрица сначала действует на вектор-строку, стоящую *слева*:

$$(\lambda_1, \lambda_2)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} = (\mu_1, \mu_2, \mu_3),$$

где (μ_1, μ_2, μ_3) представляет вектор $(A^*\boldsymbol{\beta}) \in V^*$, который потом действует на $\mathbf{v} \in V$. Таким образом, если мы изменим обычные правила умножения матриц и позволим матрице сначала действовать налево на вектор-строку, то одна и та же матрица будет представлять A и A^*. Если же мы хотим записывать A^* традиционно, т. е. в виде матрицы, действующей на столбец справа, то мы должны писать

$$\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}.$$

И мы видим, что матрица, представляющая A^*, будет транспонированной матрицей A.

Исследуем теперь ядро и образ сопряженного отображения A^*. Если $\boldsymbol{\beta} \in W^*$ и принадлежит ядру A^*, то $A^*\boldsymbol{\beta}[\mathbf{v}] = \mathbf{0}$ для всех $\mathbf{v} \in V$, т. е. $\boldsymbol{\beta}[A\mathbf{v}] = \mathbf{0}$. Это значит, что $\boldsymbol{\beta}$ аннулирует все вектора вида $A\mathbf{v}$. Отсюда можно сделать вывод, что *ядро A^* аннулирует образ A*.

Предположим, что $\boldsymbol{\alpha} \in V^*$ и принадлежит образу A^*, т. е.

$$\boldsymbol{\alpha} = A^*\boldsymbol{\beta} \quad \text{для некоторых} \quad \boldsymbol{\beta} \in W^*.$$

Предположим, что $\mathbf{v}$ есть элемент $\operatorname{Ker} A$. Тогда

$$\boldsymbol{\alpha}[\mathbf{v}] = A^*\boldsymbol{\beta}[\mathbf{v}] = \boldsymbol{\beta}[A\mathbf{v}] = \mathbf{0}.$$

Отсюда следует, что *образ A^* аннулирует ядро A*.

Объединив эти результаты с общими результатами для дуальных пространств и факторпространств, доказанными в конце параграфа 10.5, мы можем построить две схемы, изображающие общую картину векторных пространств и линейных преобразований. Итак, для подпространств V и V^* мы имеем

$$\begin{aligned} V^*/\operatorname{Im} A^* \leftarrow V^* &\leftarrow \operatorname{Im} A^*, \\ \operatorname{Ker} A \to V \ &\to V/\operatorname{Ker} A. \end{aligned}$$

Эта схема отражает тот факт, что факторпространство $V^*/\operatorname{Im} A$ может быть идентифицировано с сопряженным к $\operatorname{Ker} A$, а образ A^* сопряжен с $V/\operatorname{Ker} A$.

Для подпространств W и W^* мы получаем схемы:

$$\begin{aligned} W^*/\operatorname{Ker} A^* \leftarrow W^* &\leftarrow \operatorname{Ker} A^*, \\ \operatorname{Im} A \to W \ &\to W/\operatorname{Im} A. \end{aligned}$$

Здесь $W^*/\operatorname{Ker} A^*$ может быть идентифицировано с дуальным $\operatorname{Im} A$, а $\operatorname{Ker} A^*$ дуально к $W/\operatorname{Im} A$. Многочисленные примеры этих соотношений будут нами рассмотрены при изучении теории электрических цепей.

Резюме

A. Векторные пространства

Вы должны знать аксиомы векторных пространств и уметь применять их.

Для данного базиса векторного пространства надо уметь находить дуальный базис для дуального пространства.

Для данного подпространства U векторного пространства V надо уметь строить базисы для аннулирующего пространства $U^\perp$ и факторпространства V/U и использовать их.

B. Линейные преобразования

Вы должны уметь писать матрицу, представляющую линейное преобразование $A : V \to W$ в заданных базисах.

Вы должны знать формулировку теоремы о ранге и дефекте, уметь доказывать и применять ее.

Для данной матрицы линейного преобразования A надо уметь пользоваться редукцией по строкам для определения базисов для $\operatorname{Ker} A$ и $\operatorname{Im} A$ и получать общее решение уравнения $A\mathbf{v} = \mathbf{w}$.

Вы должны знать определение сопряженного отображения A^* для линейного преобразования A. Надо знать формулы, доказывать и применять соотношения между ядром и образом A, между ядром и образом A^*.

Задачи

10.1. Рассмотрите пятимерное векторное пространство V полиномов $f(t)$ степени ≤ 4. Выясните, будет ли подпространством множество в каждом из перечисленных ниже примеров. Если нет, то объясните, почему. Если да, то найдите базис.

(a) Полиномы удовлетворяют условию $f(t) = f(-t)$.

(b) Полиномы удовлетворяют условию $f(0) = 1$.

(c) Полиномы удовлетворяют условию $f(1) = f(-1)$.

(d) Полиномы удовлетворяют условию $\int_{-1}^{1} tf(t)\,dt = 0$.

10.2. (a) Найдите базис для подпространства $\mathbb{R}^3$, определенного условием $\boldsymbol{\alpha}[\mathbf{v}] = 0$, где

$$\boldsymbol{\alpha} = (2, -3, 1).$$

(b) Найдите базис для подпространства $\mathbb{R}^3$, определенного условиями $\boldsymbol{\alpha}[\mathbf{v}] = 0$ и $\boldsymbol{\beta}[\mathbf{v}] = 0$, где

$$\boldsymbol{\alpha} = (2, -3, 1) \quad \text{и} \quad \boldsymbol{\beta} = (2, 1, -1).$$

(c) Найдите базис аннулирующего пространства для подпространства $W \subset \mathbb{R}^3$, натянутого на вектора

$$\begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}.$$

10.3. (a) Покажите, что множество функций $f(t)$, удовлетворяющих уравнению $f'' + 5f' + 6f = 0$, является векторным пространством V.

(b) Возьмем три элемента пространства V^*, дуального к этому пространству:

$$\begin{aligned} \boldsymbol{\alpha}^1 &: \; f \mapsto f(0), \\ \boldsymbol{\alpha}^2 &: \; f \mapsto f'(0), \\ \boldsymbol{\alpha}^3 &: \; f \mapsto \int_0^{+\infty} e^t f(t)\,dt. \end{aligned}$$

Найдите соотношение между $\boldsymbol{\alpha}^1$, $\boldsymbol{\alpha}^2$ и $\boldsymbol{\alpha}^3$, показывающее, что они *линейно зависимы*.

(c) В качестве базиса пространства V выберем вектора

$$\mathbf{v}_1 = e^{-2t} + e^{-3t},$$
$$\mathbf{v}_2 = 2e^{-2t} - e^{-3t}.$$

Выразите дуальные базисные элементы $\boldsymbol{\beta}^1$ и $\boldsymbol{\beta}^2$ через $\boldsymbol{\alpha}^1$ и $\boldsymbol{\alpha}^2$, определенные в предыдущем пункте.

10.4. (a) Покажите, что если S и T — подпространства векторного пространства V, то их пересечение $S \cap T$ (множество элементов, общих для S и T) является подпространством V.

(b) Покажите, что $S+T$ (множество векторов, являющихся линейными комбинациями векторов из S и T) будет подпространством V.

(c) Покажите, что $\dim(S+T) = \dim(S) + \dim(T) - \dim(S \cap T)$. (Указание: начните с базиса $S \cap T$ и расширьте его до базиса $S+T$.)

(d) Предположим, что V есть $\mathbb{R}^4$, S натянуто на векторы $\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$ и $\begin{pmatrix} 2 \\ 1 \\ 2 \\ 1 \end{pmatrix}$, а T соответственно на векторы $\begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}$ и $\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$. Постройте базисы для $S \cap T$, $S+T$ и для пространства, аннулирующего $S+T$.

10.5. Пусть W — подпространство $\mathbb{R}^3$, натянутое на вектор $\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}$.

(a) Покажите, что $\mathbf{e}_1 = \overline{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}}$ и $\mathbf{e}_2 = \overline{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}$ образуют базис факторпространства $\mathbb{R}^3/W$. Выразите вектора $\overline{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}}$, $\overline{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}}$

и $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ в виде линейных комбинаций этих базисных векторов.

(b) Покажите, что $\boldsymbol{\alpha} = (2, -1, 0)$ и $\boldsymbol{\beta} = (4, 0, -1)$, будучи элементами дуального пространства к $\mathbb{R}^3$, являются базисом аннулирующего пространства $W^{\perp}$. Для $W^{\perp}$ постройте базис $\{\boldsymbol{\epsilon}^1, \boldsymbol{\epsilon}^2\}$ через элементы $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$, который будет дуален базису $\{\mathbf{e}_1, \mathbf{e}_2\}$ факторпространства $\mathbb{R}^3/W$.

10.6. (a) Пусть W — подпространство векторного пространства V. Пусть U — подпространство W. Докажите, что W/U является подпространством V/U и что $\dfrac{V/U}{W/U}$ естественно совпадает с V/W.

(b) Пусть векторное пространство V есть $\mathbb{R}^3$, W — плоскость $x + y + z = 0$, а U — прямая линия, определенная вектором $\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$. Получите явный вид базисов для этих пространств.

(c) Разберитесь, что происходит в дуальном пространстве, т. е. постройте подпространства V^*, которые будут дуальными или аннулирующими пространствами для всех пространств, перечисленных в пункте (a). Сначала сделайте это в общем виде, а потом для частного случая, описанного в пункте (b).

10.7. Рассмотрите пространство V, образованное функциями вида $f(t) = At + B/t$, заданными на отрезке $[1, 2]$. Пусть базисные вектора имеют вид $\mathbf{v}_1 = t$ и $\mathbf{v}_2 = 1/t$. Два элемента дуального пространства V^* равны

$$\boldsymbol{\beta}^1 : f(t) \to f(1),$$
$$\boldsymbol{\beta}^2 : f(t) \to f(2).$$

Выразите базисные элементы $\{\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2\}$, *дуальные* к $\{\mathbf{v}_1, \mathbf{v}_2\}$, в виде линейной комбинации $\boldsymbol{\beta}^1$ и $\boldsymbol{\beta}^2$.

Прежде чем выполнять упражнения 10.8–10.12, перечитайте параграф 4.2 о процессе Грама–Шмидта.

10.8. (a) Найдите размерность пространства тригонометрических полиномов, натянутого на

$$\{1,\ \sin^2 x,\ \cos^2 x,\ \sin^4 x,\ \sin^2 x \cos^2 x,\ \cos^4 x\}.$$

(b) Пусть скалярное произведение в этом пространстве равно

$$(f,g) = \frac{1}{\pi}\int_0^{\pi} f(x)g(x)\,dx.$$

С учетом этого определения постройте ортонормированный базис.

10.9. (a) Рассмотрите векторное пространство нечетных полиномов степени ≤ 3 с базисом $\{t, t^3\}$ и скалярным произведением $\int_0^1 f(t)g(t)\,dt$. Постройте ортогональный базис (его не надо нормировать).

(b) Для линейного оператора $f : V \to V$, определенного формулой $f(p(t)) = tp'(t)$, постройте матрицу A, которая представляет f в базисе $\{t, t^3\}$, и матрицу $\hat{A}$, которая представляет f в ортогональном базисе, построенном в предыдущем пункте.

(c) Три элемента дуального к V пространства V^* можно задать в виде: $\boldsymbol{\alpha}^1[p] = p(1)$, $\boldsymbol{\alpha}^2[p] = p'(0)$ и $\boldsymbol{\alpha}^3[p] = \int_0^1 tp(t)\,dt$. Получите соотношение между $\boldsymbol{\alpha}^1$, $\boldsymbol{\alpha}^2$ и $\boldsymbol{\alpha}^3$, показывающее в явном виде, что они линейно зависимы.

10.10. Пусть V — двумерное векторное пространство функций, являющихся линейными комбинациями $f_1 = 1$ и $f_2 = \cos^2 t$. Определим скалярное произведение в этом пространстве $(f,g) = \int_0^{\pi/2} f(t)g(t)\,dt$.

(a) Постройте ортонормированный базис для пространства V. (Замечание: $\dfrac{2}{\pi}\displaystyle\int_0^{\pi/2}\cos^2 t\,dt = \frac{1}{2}$, $\dfrac{2}{\pi}\displaystyle\int_0^{\pi/2}\cos^4 t\,dt = \frac{3}{8}$.)

(b) Три элемента дуального пространства V^* равны:

$$\boldsymbol{\alpha}^1 : f \to f(\pi/2), \quad \boldsymbol{\alpha}^2 : f \to \int_0^{\pi/2} f(t)\cos t\,dt,$$

$$\boldsymbol{\alpha}^3 : f \to \int_0^{\pi/2} f(t)\sin t\,dt.$$

Покажите в явном виде, что они линейно зависимы.

10.11. Рассмотрите трехмерное пространство V, элементами которого будут полиномы степени ≤ 2, умноженные на e^{-2t}. Базисом этого пространства будут три вектора:

$$\mathbf{v}_1 = e^{-2t}, \quad \mathbf{v}_2 = te^{-2t}, \quad \mathbf{v}_3 = t^2e^{-2t}.$$

(a) В этом базисе напишите матрицу D, представляющую операцию дифференцирования, т. е.

$$Df(t) = f'(t).$$

(b) Постройте матрицу $D^2 + 4D + 4\mathbb{I}$. (В этой матрице будет много нулей!)

(c) *Общее* решение уравнения $\ddot{x} + 4\dot{x} + 4x = e^{-2t}$ лежит в векторном пространстве V. Получите его с помощью матрицы, построенной в пункте (b).

10.12. Векторное пространство решений уравнения

$$\ddot{x} + 3\dot{x} + 2x = 0$$

обозначим через V. Определим скалярное произведение в этом пространстве равенством

$$(\mathbf{f}, \mathbf{g}) = \int_0^{+\infty} \mathbf{f}(t)\mathbf{g}(t)\, dt.$$

(Учтите, что $\int_0^{+\infty} e^{-\lambda t} dt = 1/\lambda$.)

(a) Пусть $\mathbf{f}_1 = e^{-2t}$ — первый базисный вектор пространства V. Постройте второй вектор $\mathbf{f}_2$, ортогональный к $\mathbf{f}_1$.

(b) В базисе $\{\mathbf{f}_1, \mathbf{f}_2\}$ постройте матрицу D, которая представляет операцию дифференцирования по t. Проверьте, что $D^2 + 3D + 2\mathbb{I} = 0$. Пусть три элемента дуального пространства V^* имеют вид

$$\begin{aligned} \boldsymbol{\alpha}_1[\mathbf{f}] &= \mathbf{f}(0), \\ \boldsymbol{\alpha}_2[\mathbf{f}] &= \mathbf{f}'(0), \\ \boldsymbol{\alpha}_3[\mathbf{f}] &= \int_0^{+\infty} \mathbf{f}(t)\, dt. \end{aligned}$$

(c) Объясните из общих соображений, почему существуют числа λ_1, λ_2, λ_3 (не все равные нулю), для которых выполняется соотношение

$$\lambda_1\boldsymbol{\alpha}_1 + \lambda_2\boldsymbol{\alpha}_2 + \lambda_3\boldsymbol{\alpha}_3 = \mathbf{0}.$$

Получите λ_1, λ_2 и λ_3.

(d) С помощью $\boldsymbol{\alpha}_1$ и $\boldsymbol{\alpha}_2$ постройте базис $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2\}$ для V^*, который будет *дуальным* к базису $\{\mathbf{f}_1, \mathbf{f}_2\}$, полученному в пункте (a).

3. Пусть V — четырехмерное векторное пространство полиномов степени ≤ 3 с базисными элементами $1, t, t^2, t^3$. Пусть D — дифференциальный оператор, действующий в этом пространстве: $Df(t) = f'(t)$ и пусть T_a — оператор трансляции: $T_a f(t) = f(t+a)$.

   (a) В этом базисе постройте матрицы D и T_a и покажите, что
   $$T_a = \mathbb{I} + Da + \frac{1}{2}D^2a^2 + \frac{1}{6}D^3a^3.$$

   (b) Докажите, что если V — пространство полиномов степени $\leq n$, то $T_a = \exp(Da)$. (Указание: вспомните формулу Тейлора, и тогда не придется строить матрицы.)

4. Пусть V — пространство 1-форм на плоскости, для которых коэффициенты при dx и dy — квадратичные функции. Базисом такого пространства V будут $x^2\,dx$, $xy\,dx$, $y^2\,dx$, $x^2\,dy$, $xy\,dy$, $y^2\,dy$. Любая кривая Γ, лежащая на плоскости, определяет элемент $\boldsymbol{\alpha}_\Gamma$ дуального пространства V^* согласно формуле
   $$\boldsymbol{\alpha}_\Gamma[\omega] = \int_\Gamma \omega.$$

   (a) Найдите нетривиальную кривую Γ, для которой $\boldsymbol{\alpha}_\Gamma$ будет нулевым элементом V^*.

   (b) Получите базис для подпространства V, которое аннулируется элементом $\boldsymbol{\alpha}_\Gamma$, где Γ — *любая* замкнутая кривая.

   (c) Пусть Γ_1, Γ_2 и Γ_3 — три *произвольные* замкнутые кривые на плоскости. Докажите, что $\boldsymbol{\alpha}_{\Gamma_1}$, $\boldsymbol{\alpha}_{\Gamma_2}$ и $\boldsymbol{\alpha}_{\Gamma_3}$ линейно зависимы. (Указание: воспользуйтесь теоремой Грина, чтобы преобразовать криволинейные интегралы к двойным.)

   (d) Найдите такие кривые $\Gamma_1, \Gamma_2, \ldots, \Gamma_6$ (это могут быть отрезки прямых линий), чтобы $\boldsymbol{\alpha}_{\Gamma_1}, \boldsymbol{\alpha}_{\Gamma_2}, \ldots, \boldsymbol{\alpha}_{\Gamma_6}$ были базисом для V^*, дуальным к базису, данному в условии задачи. Например,
   $$\int_{\Gamma_1} x^2\,dx = 1, \quad \text{остальные} \quad \int_{\Gamma_j} x^2\,dx = 0 \quad (j = 2, 3, 4, 5, 6);$$
   $$\int_{\Gamma_2} xy\,dx = 1, \quad \text{остальные} \quad \int_{\Gamma_j} xy\,dx = 0 \quad (j = 1, 3, 4, 5, 6).$$

10.15. Рассмотрите линейное преобразование f из пространства $\mathbb{R}^4$ в пространство $\mathbb{R}^3$, матрица которого имеет вид:

$$\begin{pmatrix} 1 & 2 & 0 & 1 \\ 1 & 0 & 2 & -3 \\ 0 & 1 & -1 & 2 \end{pmatrix}.$$

(a) Получите базис для ядра f и найдите общее решение уравнения

$$f(\mathbf{v}) = \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix}.$$

(b) Пусть N обозначает ядро f; пусть G — факторпространство $\mathbb{R}^4/N$. Постройте базис для G и объясните, почему полученные базисные вектора будут линейно независимыми элементами G.

10.16. Рассмотрите линейное преобразование f из $\mathbb{R}^4$ в $\mathbb{R}^3$, с матрицей

$$\begin{pmatrix} 2 & 4 & 2 & 2 \\ 1 & 3 & 2 & 0 \\ 3 & 1 & -2 & 8 \end{pmatrix}.$$

(a) Получите базис ядра f и найдите общее решение уравнения

$$f(\mathbf{v}) = \begin{pmatrix} 0 \\ -1 \\ 5 \end{pmatrix}.$$

(b) Напишите базис образа f.

(c) Два элемента факторпространства H имеют вид:

$$\mathbf{h}_1 = \overline{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}} \quad \text{и} \quad \mathbf{h}_2 = \overline{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}}.$$

Покажите в явной форме, что $\mathbf{h}_1$ и $\mathbf{h}_2$ *линейно зависимы*. (Указание: вернитесь к первому пункту этого упражнения.)

10.17. Линейное преобразование $f : \mathbb{R}^4 \to \mathbb{R}^4$ представляется матрицей

$$A = \begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 4 & -8 \\ 2 & 1 & -4 & 6 \\ 1 & 1 & -3 & 5 \end{pmatrix}.$$

(a) С помощью редукции по строкам получите ранг матрицы A и найдите базис ядра этой матрицы. Обозначьте этот базис: $\mathbf{v}_1, \ldots, \mathbf{v}_k$, где $k = \dim \operatorname{Ker} A$.

(b) Найдите решения $\mathbf{w}_1, \ldots, \mathbf{w}_k$ уравнений $A\mathbf{w}_i = \mathbf{v}_i$.

(c) Будут ли вектора $\mathbf{w}_1, \ldots, \mathbf{w}_k, \mathbf{v}_1, \ldots, \mathbf{v}_k$ базисом пространства $\mathbb{R}^4$? Что означает ответ на этот вопрос для преобразования f?

(d) Найдите базис $\operatorname{Im} A$ и выразите этот базис через вектора $\mathbf{w}_i$ и $\mathbf{v}_i$. Как этот результат связан с результатами предыдущего пункта?

10.18. Нам дана матрица 4×4:

$$A = \begin{pmatrix} 2 & 4 & 2 & 2 \\ 1 & 2 & 0 & 2 \\ 3 & 6 & 5 & 1 \\ 0 & 0 & 3 & -3 \end{pmatrix}.$$

(a) С помощью редукции по строкам получите базис ядра A и базис образа A. Найдите общее решение уравнения

$$A\mathbf{v} = \begin{pmatrix} 4 \\ 3 \\ 4 \\ -3 \end{pmatrix}.$$

(b) Получите базисные вектора $\mathbf{u}_1$ и $\mathbf{u}_2$ для факторпространства $U = \mathbb{R}^4/\operatorname{Ker} A$. Вектор $\begin{pmatrix} 4 \\ 3 \\ 4 \\ -3 \end{pmatrix}$ выразите через полученные $\mathbf{u}_1$ и $\mathbf{u}_2$.

10.19. Пусть W — векторное подпространство в $\mathbb{R}^4$, натянутое на вектора

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \end{pmatrix}, \quad \mathbf{w}_2 = \begin{pmatrix} 3 \\ 3 \\ 5 \\ 3 \end{pmatrix}, \quad \mathbf{w}_3 = \begin{pmatrix} 1 \\ 1 \\ 3 \\ 5 \end{pmatrix}.$$

Скалярное произведение в $\mathbb{R}^4$ — обычное евклидово.

(a) Постройте ортонормированный базис для W.

(b) Запишите вектор $\mathbf{v} = \begin{pmatrix} 4 \\ 0 \\ -1 \\ 7 \end{pmatrix}$ в виде суммы вектора из пространства W и вектора, ортогонального к W.

(c) Пусть $f : \mathbb{R}^4 \to \mathbb{R}^3$ определено формулой

$$f(\mathbf{v}) = \begin{pmatrix} (\mathbf{w}_1, \mathbf{v}) \\ (\mathbf{w}_2, \mathbf{v}) \\ (\mathbf{w}_3, \mathbf{v}) \end{pmatrix}.$$

Запишите матрицу, представляющую преобразование f. С помощью редукции по строкам получите общее решение уравнения

$$f(\mathbf{v}) = \begin{pmatrix} 4 \\ 6 \\ -2 \end{pmatrix}.$$

10.20. С помощью редукции по строкам получите базис подпространства $U \in \mathbb{R}^5$, аннулирующее пространство которого $U^{\perp}$ натянуто на вектор-строки

$$\begin{aligned} \boldsymbol{\alpha}^1 &= (1, 2, 0, -1, 2), \\ \boldsymbol{\alpha}^2 &= (2, 4, 3, 4, 4), \\ \boldsymbol{\alpha}^3 &= (0, 0, 1, 2, 1). \\ \boldsymbol{\alpha}^4 &= (3, 6, 5, 7, 8). \end{aligned}$$

Получите общее решение для системы линейных уравнений

$$\boldsymbol{\alpha}^1[\mathbf{v}] = 5, \quad \boldsymbol{\alpha}^2[\mathbf{v}] = 16, \quad \boldsymbol{\alpha}^3[\mathbf{v}] = 3, \quad \boldsymbol{\alpha}^4[\mathbf{v}] = 27.$$

10.21. С помощью редукции по строкам получите матрицу, обратную к

$$A = \begin{pmatrix} 1 & -2 & 1 \\ -2 & 5 & -4 \\ 1 & -4 & 6 \end{pmatrix}.$$

10.22. Пусть V — пространство функций f на $\mathbb{R}^2$, обладающих свойством $f(\lambda x, \lambda y) = \lambda^3 f(x, y)$. Базисные вектора пространства V имеют вид

$$\{\mathbf{v}_1 = x^3, \quad \mathbf{v}_2 = x^2 y, \quad \mathbf{v}_3 = x y^2, \quad \mathbf{v}_4 = y^3\}.$$

Пусть W — пространство 1-форм на плоскости, являющихся квадратичными функциями x и y. Базис этого пространства имеет вид

$$\mathbf{w}_1 = x^2\,dx,\ \ \mathbf{w}_2 = xy\,dx,\ \ \mathbf{w}_3 = y^2\,dx,$$
$$\mathbf{w}_4 = x^2\,dy,\ \ \mathbf{w}_5 = xy\,dy,\ \ \mathbf{w}_6 = y^2\,dy.$$

Оператор d является линейным преобразованием из V в W.

(a) Напишите матрицу, представляющую оператор d в заданных базисах.

(b) Получите базис для образа d и для факторпространства $G = W/(\operatorname{Im} d)$.

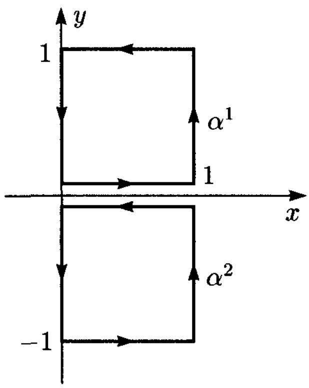


Рис. 10.6

(c) Два элемента дуального пространства W^* обозначаются $\boldsymbol{\alpha}^1$ и $\boldsymbol{\alpha}^2$. Причем $\boldsymbol{\alpha}^1$ приписывает любому элементу $\omega \in W$ значение интеграла $\int_{\alpha^1} \omega$, где α^1 является квадратом единичной площади $0 \le x \le 1$, $0 \le y \le 1$, проходимым против часовой стрелки, а для элемента $\boldsymbol{\alpha}^2$ берется другой квадрат $0 \le x \le 1$, $-1 \le y \le 0$ (рис. 10.6). Получите векторы-строки, представляющие $\boldsymbol{\alpha}^1$ и $\boldsymbol{\alpha}^2$. Найдите линейные комбинации $\boldsymbol{\alpha}^1$ и $\boldsymbol{\alpha}^2$, образующие дуальный базис для базиса пространства G.

(d) Пусть U обозначает пространство 2-форм на плоскости, которые зависят *линейно* от x и y. Базисом этого пространства являются вектора

$$\{\mathbf{u}_1 = x\,dx \wedge dy, \quad \mathbf{u}_2 = y\,dx \wedge dy\}.$$

Постройте матрицу, представляющую оператор d из W в U. Что является ядром этого оператора?

10.23. Пусть A — линейное преобразование из V в W; и пусть A^*— сопряженное преобразование из W^* в V^*. Предположим, что мы имеем произвольные (не обязательно дуальные) базисы для V и V^*: базис $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ для пространства V и базис $\{\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^m\}$ для пространства V^*, где $\boldsymbol{\alpha}^i[\mathbf{v}_j] = S_{ij}$. Числа S_{ij} образуют $m \times m$ матрицу S. Аналогично у нас есть базисы $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ и $\{\boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \ldots, \boldsymbol{\beta}^n\}$ для W и W^* соответственно, где $\boldsymbol{\beta}^k[\mathbf{w}_l] = T_{kl}$. Числа T_{kl} образуют $n \times n$ матрицу T. (Если бы мы выбрали дуальные базисы, то T и S были бы единичными матрицами.)

Если A — матрица отображения в базисах $\{\mathbf{v}_i\}$ и $\{\mathbf{w}_i\}$, то какова будет A^* — матрица сопряженного отображения в базисах $\{\boldsymbol{\beta}^i\}$ и $\{\boldsymbol{\alpha}^i\}$? Ответ выразить через транспонированную матрицу A^t и матрицы S и T.

10.24. Для векторного пространства V, где определено скалярное произведение $(\mathbf{v}_1, \mathbf{v}_2)$, сопряженное преобразование A^* для линейного преобразования $A : V \to V$ тоже будет линейным преобразованием $A^* : V \to V$, определяемым формулой

$$(A^*\mathbf{v}_1, \mathbf{v}_2) = (\mathbf{v}_1, A\mathbf{v}_2).$$

(a) Покажите, что это определение следует из определения A^*, как преобразования из V^* в V^*, с учетом обычной идентификации V^* с V, которая следует из скалярного произведения[5].

(b) Покажите, что в *ортонормированном* базисе $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ матрица, представляющая A^*, будет транспонированной к матрице A.

(c) Пусть π обозначает линейную операцию ортогонального проектирования из V на подпространство W, т. е. для любого $\mathbf{v} \in V$ вектор $\pi\mathbf{v}$ принадлежит подпространству W и вектор $\mathbf{v} - \pi\mathbf{v}$ ортогонален вектору $\pi\mathbf{v}$. (Операция π есть преобразование из V в V, причем $\operatorname{Im}\pi = W$.) Покажите, что $\pi^* = \pi$.

[5]Напомним, что упомянутая идентификация означает, что каждому вектору $\mathbf{u} \in V$ ставится в соответствие линейная функция $\boldsymbol{\alpha} \in V^*$, действующая по формуле $\boldsymbol{\alpha}(\mathbf{v}) = (\mathbf{u}, \mathbf{v})$. — *Прим. ред.*

(d) Пусть V — пространство полиномов степени ≤ 2, где определено скалярное произведение $(\mathbf{f}, \mathbf{g}) = \int_0^1 \mathbf{f}(t)\mathbf{g}(t)\,dt$. Выберем базис $\mathbf{v}_1 = 1$, $\mathbf{v}_2 = t$, $\mathbf{v}_3 = t^2$. (Этот базис не ортонормирован.) Пусть A — линейное преобразование, определенное формулой

$$A\mathbf{f}(t) = \mathbf{f}(t+1).$$

В этом базисе постройте матрицу, представляющую A^*.

10.25. Рассмотрите линейное преобразование из четырехмерного векторного пространства V в трехмерное векторное пространство W, которое представляется матрицей

$$A = \begin{pmatrix} 1 & 1 & 0 & 3 \\ 2 & 1 & 2 & 2 \\ 2 & 3 & -2 & 10 \end{pmatrix}.$$

(a) Пусть M обозначает образ A. Покажите, что вектора $\mathbf{m}_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$ и $\mathbf{m}_2 = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$ образуют базис подпространства M.

(b) Пусть $H = W/M$. Покажите, что базис этого факторпространства состоит из одного вектора $\mathbf{h}_1 = \overline{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}}$, который является классом эквивалентности, содержащим все вектора $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \mathbf{m}$, где $\mathbf{m}$ — элемент M. Покажите, что $\overline{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}} = -\frac{1}{4}\mathbf{h}_1$, т. е. найдите такой элемент $\mathbf{m}$ из пространства M, чтобы выполнялось равенство

$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = -\frac{1}{4}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \mathbf{m}.$$

Выразите $\overline{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}}$ через $\mathbf{h}_1$.

(c) Пусть N обозначает ядро A. Один элемент из N равен $\mathbf{n}_1 = \begin{pmatrix} 1 \\ -4 \\ 0 \\ 1 \end{pmatrix}$. Найдите второй вектор $\mathbf{n}_2$, так чтобы $\mathbf{n}_1$ и $\mathbf{n}_2$ были базисом пространства N.

Указание: Сначала проведите редукцию по строкам для матрицы A, а потом ищите вектор вида $\begin{pmatrix} a \\ b \\ 1 \\ 0 \end{pmatrix}$.

(d) Пусть G обозначает факторпространство V/N. Базисные вектора этого пространства имеют вид: $\mathbf{g}_1 = \overline{\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}}$ и $\mathbf{g}_2 = \overline{\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}}$. Покажите, что $\overline{\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}} = 2\mathbf{g}_1 - 2\mathbf{g}_2$, и выразите $\overline{\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}}$ через $\mathbf{g}_1$ и $\mathbf{g}_2$.

(e) Получите несингулярную 2×2 матрицу C, которая представляет A как преобразование от G (с базисом $\mathbf{g}_1$, $\mathbf{g}_2$) к M (с базисом $\mathbf{m}_1$, $\mathbf{m}_2$).

(f) Вектор $\begin{pmatrix} 2 \\ 1 \\ 7 \end{pmatrix}$ выразите через $\mathbf{m}_1$ и $\mathbf{m}_2$. (Указание: сделайте те же операции, которые выполняются при редукции по строкам.) После этого примените C^{-1}, т. е. решите уравнение $A\mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ 7 \end{pmatrix}$, записав ответ в виде $\mathbf{v} = a\mathbf{g}_1 + b\mathbf{g}_2 = \begin{pmatrix} a \\ b \\ 0 \\ 0 \end{pmatrix} + \mathbf{n}$, где $\mathbf{n}$ — произвольный элемент из N.

10.26. Пусть f^* обозначает сопряженное отображение для f в предыдущей задаче. Отображение f^* является отображением дуального пространства W^* в дуальное пространство V^* и задается формулой $f^*\boldsymbol{\beta}(\mathbf{v}) = \boldsymbol{\beta}(f\mathbf{v})$, где $\boldsymbol{\beta}$ и $\mathbf{v}$ — произвольные элементы из W^* и V. Если взять базисы, дуальные к исходным базисам в пространствах V и W, то f^* представляется матрицей, транспонированной к A:

$$A^t = \begin{pmatrix} 1 & 2 & 2 \\ 1 & 1 & 3 \\ 0 & 2 & -2 \\ 3 & 2 & 10 \end{pmatrix}.$$

(a) Покажите, что образ f^* сопряжен с пространством G. Пусть $\mathbf{g}_1^*$ и $\mathbf{g}_2^*$ — элементы V^*, дуальные к g_1 и g_2. Каждый столбец в матрице A^t выразите через $\mathbf{g}_1^*$ и $\mathbf{g}_2^*$.

(b) Покажите, что ядро f^* сопряжено с пространством H. Пусть элемент $\boldsymbol{\gamma}'$ дуален $\mathbf{h}_1$, т. е. $\boldsymbol{\gamma}'(\mathbf{h}_1) = 1$. Пусть $\boldsymbol{\beta}^1$ — элемент W^*, выбирающий первую компоненту вектора, т. е. $\boldsymbol{\beta}^1 \begin{pmatrix} a \\ b \\ c \end{pmatrix} = a$. Выразите $\boldsymbol{\beta}^1$ через $\boldsymbol{\gamma}'$. Сделайте то же самое для $\boldsymbol{\beta}^2$ и $\boldsymbol{\beta}^3$, которые выбирают вторую и третью компоненты вектора соответственно.

(c) Найдите такие вектора $\boldsymbol{\alpha}^1$ и $\boldsymbol{\alpha}^2$ в пространстве W^*, что $f^*\boldsymbol{\alpha}^1 = \boldsymbol{\beta}^1$, $f^*\boldsymbol{\alpha}^2 = \boldsymbol{\beta}^2$. Покажите, что $\boldsymbol{\alpha}^1$, $\boldsymbol{\alpha}^2$ и $\boldsymbol{\gamma}'$ образуют базис для W^*.

Глава 11

Определители

Глава 11 посвящена доказательствам основных свойств определителей матриц $n \times n$. Сначала приводятся аксиомы для определителей, потом даются методы их вычисления.

Введение

В этой главе мы обсудим свойства определителей матриц размера $n \times n$. Пусть A — матрица $n \times n$. Столбцы этой матрицы будем обозначать $A_1, \ldots, A_n$. Для единичной матрицы $\mathbb{I}$ размера $n \times n$ столбцы имеют вид:

$$I_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad I_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \text{и т. д.}$$

Тогда для произвольной матрицы A

$$A_1 = AI_1, \quad \ldots, \quad A_n = AI_n,$$

другими словами, A_i является образом I_i, т. е. i-го вектора стандартного базиса в $\mathbb{R}^n$ при отображении A.

Мы хотели бы определить $\operatorname{Det} A$ как *ориентированный объем параллелепипеда*, построенного на векторах $A_1, \ldots, A_n$. Наши знания, полученные в главах 1 и 4, позволяют говорить, что этот ориентированный объем будет полилинейным, т. е. линейным по A_1 при фиксированных $A_2, \ldots, A_n$, линейным по A_2 при фиксированных $A_1, A_3, \ldots, A_n$, и так далее. Вследствие наличия ориентации мы ожидаем, что $\operatorname{Det} A$ антисимметричен по столбцам, т. е. перестановка любых двух столбцов матрицы изменяет знак определителя. Мы должны дать определение определителя и доказать, что он обладает требуемыми свойствами. Материал излагается классически, начиная с основных аксиом для определителей. Мы выпишем основные свойства, которыми должны обладать определители, и покажем, что эти свойства задают его однозначно. Другими словами, покажем, что существует не более, чем одна функция от элементов матрицы с требуемыми свойствами. Затем получим правила вычисления определителей и покажем, что функция, удовлетворяющая аксиомам, существует. Потом дадим другие определения определителей и убедимся, что все определения дают одну и ту же функцию.

В последующем мы будем рассматривать функцию от матриц. Если эта функция вычисляется для конкретной матрицы A, то пишем $D(A)$ или $D(A_1, \ldots, A_n)$, чтобы подчеркнуть, что это функция от n векторов $A_1, \ldots, A_n$. Если все столбцы, кроме k-го, фиксированы, то мы получим функцию одного столбца, которую будем обозначать D_k. Например, будем писать

$$D_2\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) \quad \text{для} \quad D\left(\begin{pmatrix} 1 & x & 7 \\ -2 & y & 9 \\ 3 & z & -1 \end{pmatrix}\right).$$

(Строго говоря, в обозначении D_k мы должны отразить постоянные векторы $A_1 = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}$ и $A_3 = \begin{pmatrix} 7 \\ 9 \\ -1 \end{pmatrix}$ и писать

$$D_{2;\left(\begin{smallmatrix} 1 \\ -2 \\ 3 \end{smallmatrix}\right),\left(\begin{smallmatrix} 7 \\ 9 \\ -1 \end{smallmatrix}\right)}\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right),$$

но это обозначение слишком громоздко.)

11.1. Аксиомы для определителей

Функция матриц D называется *определителем*, если она удовлетворяет следующим условиям.

- Каждая функция D_k линейна:

$$\begin{aligned} D_k(A_k + A'_k) &= D_k(A_k) + D_k(A'_k), \\ D_k(cA_k) &= cD_k(A_k). \end{aligned} \tag{11.1}$$

  Другими словами, D линейна по каждому столбцу, когда другие столбцы фиксированы.

- Если два соседних столбца матрицы A равны, то

$$D(A) = 0. \tag{11.2}$$

  Наконец,

$$D(\mathbb{I}) = 1. \tag{11.3}$$

Предположим, что функция D, удовлетворяющая свойствам (11.1) и (11.2), существует. Что из этого следует?

Если один столбец умножить на какое-то число и прибавить его к соседнему, то значение функции D не изменится.

Доказательство. Прибавим к $(k+1)$-му столбцу k-й, умноженный на c. Получим

$$\begin{aligned} &D(A_1, \ldots, A_k, cA_k + A_{k+1}, A_{k+2}, \ldots, A_n) \\ &\quad = D(A_1, \ldots, A_k, A_{k+1}, A_{k+2} \ldots, A_n) \\ &\qquad + cD(A_1, \ldots, A_k, A_k, A_{k+2}, \ldots, A_n) \\ &\quad = D(A_1, \ldots, A_n). \end{aligned} \tag{11.4}$$

Здесь использовались условия (11.1) и (11.2). Например,

$$D\left(\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}\right) = D\left(\begin{pmatrix} 1 & 4-4\cdot 1 & 7 \\ 2 & 5-4\cdot 2 & 8 \\ 3 & 6-4\cdot 3 & 9 \end{pmatrix}\right) = D\left(\begin{pmatrix} 1 & 0 & 7 \\ 2 & -3 & 8 \\ 3 & -6 & 9 \end{pmatrix}\right).$$

Чтобы получить еще одно свойство определителя, прибавим k-й столбец к $(k+1)$-му, получившийся $(k+1)$-й столбец вычтем из

k-го столбца, потом еще раз прибавим полученный k-й столбец к $(k+1)$-му. В результате получим:

$$\begin{aligned}D(A) &= D(A_1, \ldots, A_{k-1}, A_k, A_k + A_{k+1}, A_{k+2}, \ldots, A_n)\\ &= D(A_1, \ldots, A_{k-1}, A_k - (A_k + A_{k+1}), A_k + A_{k+1}, \ldots, A_n)\\ &= D(A_1, \ldots, A_{k-1}, -A_{k+1}, A_k + A_{k+1}, A_{k+2}, \ldots, A_n)\\ &= D(A_1, \ldots, A_{k-1}, -A_{k+1}, A_k + A_{k+1} - A_{k+1}, \ldots, A_n)\\ &= D(A_1, \ldots, A_{k-1}, -A_{k+1}, A_k, A_{k+2}, \ldots, A_n)\\ &= -D(A_1, \ldots, A_{k-1}, A_{k+1}, A_k, A_{k+2}, \ldots, A_n). \qquad (11.5)\end{aligned}$$

Таким образом, *перестановка двух соседних столбцов изменяет знак определителя* $D(A)$. Отсюда немедленно следует, что

если два столбца матрицы A *равны, то* $D(A) = 0$. (11.6)

Действительно, если у матрицы есть два одинаковых столбца, то мы можем переставлять соседние столбцы до тех пор, пока наши столбцы не станут рядом. После этого остается воспользоваться свойствами (11.2) и (11.5). Ссылки на свойства (11.1) и (11.6) приводят нас к еще одному следствию.

Если умножить один столбец на число
и прибавить его к другому столбцу, (11.7)
то величина $D(A)$ *не изменяется.*

Продолжим наш пример вычисления определителя. Мы получили:

$$D\left(\begin{pmatrix}1 & 4 & 7\\ 2 & 5 & 8\\ 3 & 6 & 9\end{pmatrix}\right) = D\left(\begin{pmatrix}1 & 0 & 7\\ 2 & -3 & 8\\ 3 & -6 & 9\end{pmatrix}\right)$$

Теперь 1-й столбец умножим на 7 и вычтем из 3-го, а затем 2-й столбец умножим на 2 и вычтем из 3-го:

$$D\left(\begin{pmatrix}1 & 0 & 0\\ 2 & -3 & -6\\ 3 & -6 & -12\end{pmatrix}\right) = D\left(\begin{pmatrix}1 & 0 & 0\\ 2 & -3 & 0\\ 3 & -6 & 0\end{pmatrix}\right) = 0$$

по свойству (11.1).

Сделаем еще один пример.

$$D\left(\begin{pmatrix}2 & 1 & 4\\ 3 & 2 & 1\\ 4 & 2 & -1\end{pmatrix}\right) \quad \text{2-й столбец умножим на 2 и вычтем из 1-го}$$

$$= D\left(\begin{pmatrix}0 & 1 & 4\\ -1 & 2 & 1\\ 0 & 2 & -1\end{pmatrix}\right) \quad \text{1-й столбец умножим на 2 и сложим со 2-м}$$

$$= D\left(\begin{pmatrix}0 & 1 & 4\\ -1 & 0 & 1\\ 0 & 2 & -1\end{pmatrix}\right) \quad \text{1-й столбец прибавим к 3-му}$$

$$= D\left(\begin{pmatrix}0 & 1 & 4\\ -1 & 0 & 0\\ 0 & 2 & -1\end{pmatrix}\right) \quad \text{2-й столбец умножим на 4 и вычтем из 3-го}$$

$$= D\left(\begin{pmatrix}0 & 1 & 0\\ -1 & 0 & 0\\ 0 & 2 & -9\end{pmatrix}\right) \quad \text{воспользуемся аксиомой (11.1)}$$

$$= -9D\left(\begin{pmatrix}0 & 1 & 0\\ -1 & 0 & 0\\ 0 & 2 & 1\end{pmatrix}\right) \quad \text{3-й столбец умножим на 2 и вычтем из 2-го}$$

$$= -9D\left(\begin{pmatrix}0 & 1 & 0\\ -1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}\right) \quad \text{меняем местами 1-й и 2-й столбцы}$$

$$= 9D\left(\begin{pmatrix}1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 1\end{pmatrix}\right) \quad \text{воспользуемся аксиомой (11.1)}$$

$$= -9D\left(\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\right).$$

Если теперь использовать аксиому (11.3), то получаем

$$D\left(\begin{pmatrix}2 & 1 & 4\\ 3 & 2 & 1\\ 4 & 2 & -1\end{pmatrix}\right) = -9.$$

Заметим также, что из (11.7) следует (доказывается как в (11.5)), что

$$\textit{перестановка двух любых столбцов изменяет знак } D(A). \tag{11.8}$$

Можно также получить, что

$$\textit{если столбцы } A \textit{ линейно зависимы, то } D(A) = 0. \tag{11.9}$$

Действительно, если $\lambda_1 A_1 + \cdots + \lambda_n A_n = 0$ и некоторое значение $\lambda_i \neq 0$, то можно разделить это соотношение на λ_i и получить

$$A_i - c_1 A_1 - \cdots - c_n A_n = 0, \quad c_i = 0.$$

Поэтому вычитание $c_1 A_1$ и т. д. из i-го столбца не изменяет $D(A)$ (следствие (11.7)) и приводит к матрице с нулевым i-м столбцом. Тогда вследствие (11.1) $D(A) = 0$. В частности, мы знаем, что любые n векторов вида

$$\left(\begin{pmatrix} 0 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}\right)$$

(с нулем в первой строке) линейно зависимы. Поэтому

$$\begin{gathered}\textit{если все элементы верхней строки матрицы } A \\ \textit{равны нулю, то } D(A) = 0.\end{gathered} \tag{11.10}$$

Предположим, что хотя бы один элемент в верхней строке матрицы A не равен нулю. Переставляя столбцы, если это надо, всегда можно сделать так, что *первый* столбец имеет ненулевой элемент в верхней строке. Следовательно,

$$D(A) = \pm D(B), \quad \text{где } b_{11} \neq 0.$$

Но тогда

$$D(B) = b_{11} D(B'),$$

где для матрицы B' первый столбец $B'_1 = (1/b_{11}) B_1$.

Умножая первый столбец на соответствующий коэффициент и вычитая из остальных столбцов, можно сделать так, чтобы все остальные элементы в первой строке были равны нулю, т. е.

$$D(B') = D(B''),$$

где

$$B'' = \left(\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ \hline B''_{21} & & & \\ \vdots & & C & \\ B''_{n1} & & & \end{array}\right)$$

Рассмотрим $D(B'')$ как функцию матрицы C. Очевидно, что аксиомы (11.1) и (11.2) удовлетворяются. Кроме этого, если бы C была единичной матрицей $(n-1)\times(n-1)$, мы могли бы, не изменяя значения $D(B'')$, сделать все элементы первого столбца b''_{21}, b''_{31} и т. д. равными нулю (умножая остальные столбцы на соответствующие множители и вычитая из первого). Например,

$$D\left(\begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ 4 & 0 & 0 & 1 \end{pmatrix}\right) \quad \begin{array}{l}\text{2-й столбец умножаем на 2}\\ \text{и вычитаем из 1-го}\end{array}$$

$$= D\left(\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ 4 & 0 & 0 & 1 \end{pmatrix}\right) \quad \begin{array}{l}\text{3-й столбец умножаем на 3}\\ \text{и вычитаем из 1-го}\end{array}$$

$$= D\left(\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 4 & 0 & 0 & 1 \end{pmatrix}\right) \quad \begin{array}{l}\text{4-й столбец умножаем на 4}\\ \text{и вычитаем из 1-го}\end{array}$$

$$= D\left(\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right).$$

Другими словами, функция

$$D\left(\left(\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ \hline & & & \\ B''_{21} & & & \\ \vdots & & C & \\ B''_{n1} & & & \end{array}\right)\right)$$

как функция от C удовлетворяет всем аксиомам для определителя матриц $(n-1)\times(n-1)$. Следовательно,

$$D(B'') = D(C),$$

где справа стоит D-функция (если она существует) для матриц $(n-1)\times(n-1)$.

Еще раз повторив предыдущие рассуждения, мы приходим к выводу, что или $D(C) = 0$ (если все элементы в первой строке равны нулю) или $D(C)$ можно выразить как D-функцию матриц $(n-2)\times(n-2)$. И так можно спускаться, пока не дойдем до матриц 1×1, для которых из аксиом (11.1) и (11.3) следует, что

$$D(a) = a(D(1)) = a.$$

Таким образом, мы доказали, что D-функция, если она существует, *единственна*, и получили правило для ее вычисления.

Например, пусть

$$A = \begin{pmatrix} 0 & 3 & -1 & 4 \\ 2 & 4 & 1 & 5 \\ -4 & 7 & 2 & 8 \\ 2 & 9 & 3 & 2 \end{pmatrix}.$$

Можно показать, что $D(A) = -D(B)$, где

$$B = \begin{pmatrix} -1 & 3 & 0 & 4 \\ 1 & 4 & 2 & 5 \\ 2 & 7 & -4 & 8 \\ 3 & 9 & 2 & 2 \end{pmatrix},$$

а дальше получаем, что $D(B) = -D(B'')$, где

$$B'' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 7 & 2 & 9 \\ -2 & 13 & -4 & 16 \\ -3 & 18 & 2 & 14 \end{pmatrix}.$$

Далее, $D(B'') = D(C)$, где

$$C = \begin{pmatrix} 7 & 2 & 9 \\ 13 & -4 & 16 \\ 18 & 2 & 14 \end{pmatrix}.$$

Наконец,

$$\begin{aligned} D(C) &= 2D\left(\begin{pmatrix} 7 & 1 & 9 \\ 13 & -2 & 16 \\ 18 & 1 & 14 \end{pmatrix}\right) \\ &= -2D\left(\begin{pmatrix} 1 & 0 & 9 \\ -2 & 27 & 16 \\ 1 & 11 & 14 \end{pmatrix}\right) \\ &= -2D\left(\begin{pmatrix} 1 & 0 & 0 \\ -2 & 27 & 34 \\ 1 & 11 & 5 \end{pmatrix}\right) \\ &= -2D\left(\begin{pmatrix} 27 & 34 \\ 11 & 5 \end{pmatrix}\right) \\ &= -2\cdot 27D\left(\begin{pmatrix} 1 & 34 \\ 11/27 & 5 \end{pmatrix}\right) \\ &= -2\cdot 27D\left(\begin{pmatrix} 1 & 0 \\ 11/27 & 5 - 34\cdot 11/27 \end{pmatrix}\right) \\ &= -2\cdot 27(5 - 34\cdot 11/27) = 478. \end{aligned}$$

В следующем параграфе мы дадим другое доказательство единственности $D(A)$ и другой рецепт его вычисления. Но нам еще надо показать, что $D(A)$, удовлетворяющий условиям (11.1), (11.2) и (11.3), существует.

Из приведенного способа вычисления определителя A можно получить важное следствие. Предположим, что матрица A имеет вид

$$A = \begin{pmatrix} L & M \\ 0 & N \end{pmatrix},$$

где L есть матрица $k \times k$, N — матрица $(n-k) \times (n-k)$ и M — матрица, у которой k строк и $(n-k)$ столбцов. Другими словами, предположим, что первые k столбцов матрицы A имеют нули в последних $n-k$ позициях. В этом случае первые k столбцов матрицы A линейно зависимы или линейно независимы тогда и только тогда, когда столбцы матрицы L линейно зависимы или линейно независимы, т. е. последние $n-k$ нулевых элементов не влияют на линейную зависимость или независимость этих столбцов. Если эти столбцы линейно зависимы, то

$$D(A) = 0 \quad \text{и} \quad D(L) = 0.$$

Если же они линейно независимы, то применяя вышеизложенный метод, можно использовать первые k столбцов и заменить матрицу M нулевой матрицей, не изменяя при этом элементы матрицы N. В этом случае

$$D\begin{pmatrix} L & M \\ 0 & N \end{pmatrix} = D(L)D(N). \tag{11.11}$$

Эта формула является обобщением на пространство n измерений формулы вычисления площади параллелограмма (рис. 11.1). В двумерном случае мы получаем

$$D\left(\begin{pmatrix} x & u \\ 0 & v \end{pmatrix}\right) = xv.$$

Это ориентированная площадь параллелограмма, построенного на векторах $\begin{pmatrix} x \\ 0 \end{pmatrix}$ и $\begin{pmatrix} u \\ v \end{pmatrix}$, причем она равна ориентированной площади параллелограмма, построенного на векторах $\begin{pmatrix} x \\ 0 \end{pmatrix}$ и $\begin{pmatrix} 0 \\ v \end{pmatrix}$.

В трехмерном случае

$$D\left(\begin{pmatrix} x & u & p \\ y & v & q \\ 0 & 0 & r \end{pmatrix}\right) = D\left(\begin{pmatrix} x & u \\ y & v \end{pmatrix}\right) D((r)),$$

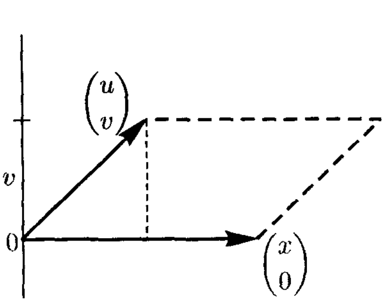


Рис. 11.1

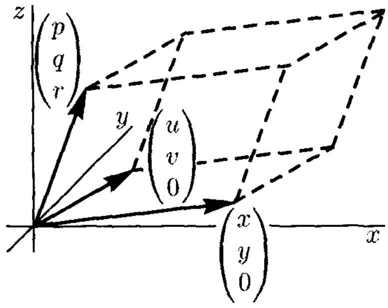


Рис. 11.2

а интерпретация та же самая (рис. 11.2).

11.2. Закон умножения и другие следствия аксиом

Рассмотрим еще некоторые следствия из свойства (11.8) о перестановке двух столбцов определителя. Пусть $(\nu_1, \ldots, \nu_n)$ — произвольная перестановка чисел $(1, \ldots, n)$. Возьмем определитель матрицы с переставленными столбцами

$$D((A_{\nu_1}, \ldots, A_{\nu_2}, \ldots, A_{\nu_n})),$$

и будем попарно переставлять их до тех пор, пока не вернемся к исходному порядку. На каждом шаге будем применять свойство (11.8) и придем к выводу, что

$$D((A_{\nu_1}, \ldots, A_{\nu_n})) = \pm D((A_1, \ldots, A_n)), \tag{11.12}$$

где знак + или − не зависит от конкретных значений матричных элементов матрицы A. Если (11.12) применить к единичной матрице, то получим

$$D((I_{\nu_1}, \ldots, I_{\nu_n})) = \pm 1$$

и, следовательно,

$$D((A_{\nu_1}, \ldots, A_{\nu_n})) = D((I_{\nu_1}, \ldots, I_{\nu_n}))D(A). \tag{11.13}$$

Пусть $B = (b_{ij})$ — другая матрица $n \times n$ и пусть

$$C = AB.$$

Столбцы матрицы C равны

$$C_k = b_{1k}A_1 + b_{2k}A_2 + \cdots + b_{nk}A_n.$$

Для вычисления $D(C)$ мы можем сначала применить (11.1) к первому столбцу C и получить сумму некоторых слагаемых; потом можно проделать то же самое со вторым столбцом, и так далее. Например, в трехмерном случае

$$\begin{aligned}
C_1 &= b_{11}A_1 + b_{12}A_2 + b_{13}A_3,\\
C_2 &= b_{12}A_1 + b_{22}A_2 + b_{32}A_3,\\
C_3 &= b_{13}A_1 + b_{23}A_2 + b_{33}A_3
\end{aligned}$$

Тогда

$$\begin{aligned}
&D(C_1, C_2, C_3)\\
&= b_{11}D(A_1, C_2, C_3) + b_{21}D(A_2, C_2, C_3) + b_{31}D(A_3, C_2, C_3)\\
&= b_{11}\{b_{12}D(A_1, A_1, C_3) + b_{22}D(A_1, A_2, C_3) + b_{32}D(A_1, A_3, C_3)\}\\
&\quad + b_{21}\{b_{12}D(A_2, A_1, C_3) + b_{22}D(A_2, A_2, C_3) + b_{32}D(A_2, A_3, C_3)\}\\
&\quad + b_{31}\{b_{12}D(A_3, A_1, C_3) + b_{22}D(A_3, A_2, C_3) + b_{32}D(A_3, A_3, C_3)\}.
\end{aligned}$$

Прежде чем двигаться дальше, можно исключить слагаемые с повторяющимися столбцами. Очевидно, что останутся только выражения типа $D((A_{\nu_1}, A_{\nu_2}, A_{\nu_3}))$, а они в силу (11.12) равны $\pm D(A)$. Таким образом, в общем случае мы видим, что

$$D(C) = D(A)\sum \pm b_{\nu_1 1}b_{\nu_2 2}\ldots b_{\nu_n n}, \tag{11.14}$$

где берется сумма по всем перестановкам и знак $\pm$ соответствует (11.12).

Предположим, что $A = \mathbb{I}$. Из (11.14) следует, что $C = B$ и

$$D(B) = \sum \pm b_{\nu_1 1}\ldots b_{\nu_n n}. \tag{11.15}$$

Таким образом, мы получили явный вид $D(B)$, что еще раз доказывает единственность определителя, если он существует. Если

(11.15) подставить в (11.14), то получается важный закон умножения определителей:

$$D(AB) = D(A)D(B). \tag{11.16}$$

Каждый член в (11.15) содержит только один элемент из каждой строки матрицы B. Поэтому

$$\begin{gathered} D(B) \text{ — \textit{линейная функция каждой строки}} \\ \text{\textit{при фиксированных других строках.}} \end{gathered} \tag{11.1$'$}$$

Выведем еще одно свойство определителя, связанное с его строками. Пусть для некоторого значения $1 \leq i < n$ матрица A имеет вид

$$\begin{aligned} A_k &= I_k, \quad k \neq i, i+1, \\ A_i &= I_i + I_{i+1}, \\ A_{i+1} &= 0. \end{aligned}$$

Например, для $n = 3$ и $i = 2$ у нас была бы матрица

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Заметим, что в матрице AB все строки, кроме $(i+1)$-й, совпадают со строками матрицы B, а $(i+1)$-я строка заменена на i-ю. В нашем примере мы получили бы

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 5 & 9 \\ 2 & 6 & 8 \\ 3 & 7 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 5 & 9 \\ 2 & 6 & 8 \\ 2 & 6 & 8 \end{pmatrix}.$$

Кроме того, $D(A) = 0$, потому что один столбец этой матрицы нулевой. С помощью подобного построения можно прийти к выводу, что

$$\text{\textit{если B имеет две одинаковых соседних строки, то}} \ D(B) = 0. \tag{11.2$'$}$$

Это значит, что $D(B^t)$ удовлетворяет аксиомам (11.1) и (11.2), потому что замена B на транспонированную матрицу B^t приводит к перестановке роли строк и столбцов. Поскольку $D(\mathbb{I}^t) = D(\mathbb{I}) = 1$, то функция $D(B^t)$ удовлетворяет аксиомам (11.1)–(11.3) и вследствие единственности определителя должна совпадать с $D(B)$. Другими словами,

$$D(B^t) = D(B). \tag{11.17}$$

11.3. Существование определителя

Теперь, наконец, мы должны доказать существование определителя, т. е. построить функцию матриц $n \times n$, которая удовлетворяет (11.1), (11.2) и (11.3)[1].

Для $n = 1$ положим по определению $D((a)) = a$.

Для $n = 2$ определим $D\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = ad - bc$. Легко проверить, что аксиомы (11.1)–(11.3) в этих случаях выполняются. Предположим, что существуют определители матриц размера $(n-1)\times(n-1)$. Пусть $A = (a_{ik})$ — матрица $n \times n$. Возьмем некоторый матричный элемент a_{ik} из i-й строки и k-го столбца. Вычеркнем из матрицы A i-ю строку и k-й столбец. Теперь рассмотрим определитель для получившейся матрицы $(n-1) \times (n-1)$. Этот определитель, умноженный на $(-1)^{i+k}$, называется *алгебраическим дополнением* элемента a_{ik} и обозначается A_{ik}. Знак $(-1)^{i+k}$ распределен в шахматном порядке, а именно:

$$\begin{matrix} + & - & + & - & \cdot & \cdot & \cdot \\ - & + & - & + & \cdot & \cdot & \cdot \\ + & - & + & - & \cdot & \cdot & \cdot \\ - & + & - & + & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix}$$

[1]На самом деле существование такой функции стало очевидным ранее, когда была получена формула (11.15) — явное выражение $D(B)$ через элементы матрицы B. В этом параграфе приводится несколько других важных понятий и формул: алгебраические дополнения и вычисление определителей с их помощью, присоединенная матрица, способ нахождения обратной матрицы. — *Прим. ред.*

Пусть i — некоторое фиксированное число от 1 до n. Рассмотрим следующую функцию D от матрицы A:

$$D = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}. \tag{11.18}$$

Ее структура — сумма произведений всех элементов i-й строки на их алгебраические дополнения.

Рассмотрим зависимость D от заданного столбца, скажем, от A_k. Для $\nu \neq k$ $A_{i\nu}$ зависит линейно от A_k, а $a_{i\nu}$ не зависит от него; для $\nu = k$ A_{ik} не зависит от A_k, но a_{ik} является элементом этого столбца, т. е. это слагаемое также линейно зависит от столбца A_k. Таким образом, аксиома (11.1) выполняется. Далее, предположим, что два соседних столбца A_k и A_{k+1} равны. Для $\nu \neq k, k+1$ мы имеем два одинаковых столбца в $A_{i\nu}$, т. е. $A_{i\nu} = 0$. При вычислении A_{ik} и $A_{i,k+1}$ используются одинаковые определители, но с противоположными знаками. Следовательно, $A_{ik} = -A_{i,k+1}$, а $a_{ik} = a_{i,k+1}$. Таким образом, $D = 0$ и выполняется аксиома (11.2). В частном случае $A_\nu = I_\nu$ $(\nu = 1, 2, \ldots, n)$ мы имеем $a_{i\nu} = 0$ для $\nu \neq i$ и $a_{ii} = 1$, $A_{ii} = 1$. Следовательно, $D = 1$, что означает выполнение аксиомы (11.3). Таким образом, мы доказали существование определителя матрицы $n \times n$ и выполнение формулы (11.18), так называемое разложение определителя по элементам i-й строки. Формулу (11.18) можно обобщить. В определителе заменим i-ю строку на j-ю и разложим определитель по этой новой строке. Для $i \neq j$ определитель равен нулю, а для $i = j$ он равен D, т. е.

$$a_{j1}A_{i1} + a_{j2}A_{i2} + \cdots + a_{jn}A_{in} = \begin{cases} D, & j = i, \\ 0, & j \neq i. \end{cases} \tag{11.19}$$

Если мы поменяем ролями строки и столбцы, то получим

$$a_{1h}A_{1k} + a_{2h}A_{2k} + \cdots + a_{nh}A_{nk} = \begin{cases} D, & h = k, \\ 0, & h \neq k. \end{cases} \tag{11.20}$$

Составим из алгебраических дополнений элементов a_{ij} матрицы A матрицу $B = (A_{ji})$. (Обратим внимание, что алгебраическое дополнение элемента a_{ij}, стоящего на пересечении i-й строки и j-го столбца, находится на пересечении j-й строки и i-го столбца.)

Матрица B называется присоединенной матрицей к матрице A. Из уравнения (11.20) следует, что

$$AB = BA = D(A)\mathbb{I}.$$

Мы уже доказали, что если A сингулярна (столбцы A линейно зависимы), то $D(A) = 0$. Если же $D(A) \neq 0$, то предыдущее уравнение дает формулу для A^{-1}. Таким образом, мы доказали, что

$$\begin{gathered}\textit{матрица } A \textit{ обратима} \\ \textit{тогда и только тогда, когда } D(A) \neq 0. \\ \textit{Если } D(A) \neq 0, \quad \textit{то } A^{-1} = \frac{1}{D(A)} B, \\ \textit{где } B \textit{ — матрица, присоединенная к матрице } A.\end{gathered} \tag{11.21}$$

(Для $n = 2$ это правило совпадает с методом, изложенным в главе 1.) При $n > 2$ это правило не очень удобно, лучше использовать алгоритм, изложенный в главе 10. И тем не менее, эта формула важна с теоретической точки зрения. Например, она показывает, что все элементы матрицы A^{-1} являются частным от деления полинома от матричных элементов A на определитель.

Резюме

A. Определители

Вы должны знать аксиомы для определителей и пользоваться этими аксиомами для вычисления определителей.

Вы должны уметь вычислять определители с помощью алгебраических дополнений.

Вы должны знать и уметь применять метод вычисления обратной матрицы для несингулярной квадратной матрицы.

Задачи

Напишите несколько матриц размера 3×3 и 4×4 и вычислите для них определители. Вы увидите, что уже для случая 4×4

выражения (11.15) и (11.18) становятся очень громоздкими (при вычислении по формуле (11.15) приходится выполнить $3 \cdot 4!$ умножений и $4! - 1$ сложений). А вот метод, описанный в параграфе 11.1, вполне приемлем. Ниже даются определители, для которых надо проверить ответы.

(a)

$$\operatorname{Det}\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 8 & 11 & 12 \\ 6 & 9 & 13 & 15 \\ 7 & 10 & 14 & 16 \end{pmatrix} = -2.$$

(b)

$$\operatorname{Det}\begin{pmatrix} 0 & 2 & 2 & 2 \\ 3 & 1 & 2 & 2 \\ 3 & 3 & 4 & 2 \\ 3 & 3 & 3 & 5 \end{pmatrix} = 12.$$

(c)

$$\operatorname{Det}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 3 & 3 & 3 \\ 2 & 2 & 3 & 3 \\ 2 & 2 & 2 & 3 \end{pmatrix} = 1.$$

(d)

$$\operatorname{Det}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 3 & 5 & 6 \\ 4 & 9 & 25 & 36 \\ 8 & 27 & 125 & 216 \end{pmatrix} = 72.$$

11.1. Обобщая пример (b), покажите, что

$$\operatorname{Det}\begin{pmatrix} r_1 & a & a & a \\ b & r_2 & a & a \\ b & b & r_3 & a \\ b & b & b & r_4 \end{pmatrix} = \frac{bf(a) - af(b)}{b - a}, \quad \text{если} \quad a \neq b,$$

где $f(x) = (r_1 - x)(r_2 - x)(r_3 - x)(r_4 - x)$.

Указание: рассмотрите определитель матрицы, выписанной ниже, его значение обозначим $F(x)$:

$$\begin{pmatrix} r_1 - x & a - x & a - x & a - x \\ b - x & r_2 - x & a - x & a - x \\ b - x & b - x & r_3 - x & a - x \\ b - x & b - x & b - x & r_4 - x \end{pmatrix}.$$

$F(x)$ линейно зависит от x, потому что мы можем вычесть первую строку из остальных, исключив x из всех строк, кроме первой. Следовательно,

$$F(x) = A + Bx,$$

где A и B — некоторые постоянные. При этом $F(a) = f(a)$ и $F(b) = f(b)$. Наш исходный определитель, очевидно, равен $F(0) = A$. Чему он будет равен при $a = b$?

11.2. Обобщая пример (c), покажите, что

$$\operatorname{Det}\begin{pmatrix} 1 & 1 & 1 & 1 \\ x & a & a & a \\ x & y & b & b \\ x & y & z & c \end{pmatrix} = (a - x)(b - y)(c - z).$$

11.3. Обобщая пример (d), покажите, что

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ x & y & z & w \\ x^2 & y^2 & z^2 & w^2 \\ x^3 & y^3 & z^3 & w^3 \end{pmatrix} = (y - x)(z - x)(w - x)(z - y)(w - y)(w - z).$$

Выпишите и докажите соответствующую формулу, если число 4 заменить на n.

11.4. Покажите, что

$$|\operatorname{Det}(A_1, \ldots, A_n)| \leq \|A_1\| \cdots \|A_n\|.$$

Когда будет равенство?

Указание: воспользуйтесь интерпретацией $|\operatorname{Det}|$ как объема.

11.5. Покажите, что если O — ортогональная матрица (т. е. $OO^t = \mathbb{I}$), то $\operatorname{Det} O = \pm 1$.

11.6. Матрица R будет матрицей *вращения*, если $RR^t = \mathbb{I}$ и $\operatorname{Det} R = +1$. Покажите, что вращение в пространстве нечетного числа измерений всегда сохраняет неподвижным по крайней мере один ненулевой вектор, т. е. число 1 является собственным значением матрицы R. Указание: рассмотрите определитель $\operatorname{Det}(R - \mathbb{I})$.

Рекомендуемая литература

В конце тома мы даем короткий перечень книг, не являющийся, конечно, библиографией. Это скорее дополнительная литература, полезная студентам, изучающим наш курс. Книгу LOOMIS AND STERNBERG можно считать параллельным учебником, хотя изложение материала более формально и абстрактно, с акцентом на математические доказательства. Содержание, по существу, такое же, как и в нашей книге, но гораздо больше внимания уделяется формальным определениям, строгой аргументации и математической изощренности. Во втором томе мы несколько раз будем отсылать читателя к этой книге за детальным доказательством некоторых ключевых теорем.

В данной книге одним из главных разделов математики является линейная алгебра. Поэтому мы рекомендуем классический учебник HALMOS, где имеется уклон к расширению конечномерной теории в направлении гильбертова пространства. Книга LANG обсуждает этот предмет с точки зрения абстрактной алгебры, в то время как в книге STRANG акцент делается на вычислительной технике и приложениях, которые мы здесь почти не рассматриваем. Чтобы получить полное представление о линейной алгебре, правильная стратегия — прочитать все три эти книги.

В главе 1 мы обсуждаем геометрию прямых. Далее, естественный шаг — проективная геометрия. Мы немного говорим о ней в добавлении к главе 1 и в задачах 1.16–1.20. В книге HARTSHORNE дается введение в проективную геометрию.

В конце главы 2 мы только слегка касаемся теории вероятностей. Этот важный раздел не получил серьезного обсуждения, что является одним из главных наших пробелов. Хорошее введение в теорию вероятностей без математических сложностей читатель может найти в трех томах HOEL, PORT AND STONE. Теория вероятнотей легко приводит к довольно изощренной математической «кухне», например, к теории меры и тонким вопросам Фурье-анализа. Достоинством этих книг является изложение важных идей без глубокого погружения в трудную математику. Книга MORAN уже сложнее для чтения, но на нее стоит потратить усилия. В книге KEMENY AND SNELL обсуждаются конечные цепи Маркова, и ее следует читать как продолжение главы 2. Книга DOYLE AND SNELL дает краткое введение и в теорию цепей Маркова, и в теорию сетей, которые мы будем изучать во втором томе.

В главе 3 мы даем начальные сведения о дифференциальных уравнениях. Учебник HIRSCH AND SMALE здесь будет вполне естественным, так как наши точки зрения довольно близки. Книга BRAUN излагает материал более стандартно. (Экспонента от матрицы появляется там только на с. 321!) Однако, ее следует прочитать, поскольку в ней есть много полезных деталей и приложений. Книга SIMMONS довольно традиционная, но в ней много интересной исторической информации. Две классические книги АРНОЛЬДА очень приятно читать. Во взаимодействии геометрии и анализа видна уверенная рука крупного специалиста в этом вопросе.

В главе 4 два параграфа посвящены теории относительности. На эту тему мы рекомендуем три книги. Во-первых, большая и сложная, богатая идеями книга MISNER, THORNE AND WHEELER. Во-вторых, короткая, захватывающая, с минимумом математики книга TAYLOR AND WHEELER. И, наконец, книга *Spacetime, Geometry, Cosmology*, написанная BURKE, обсуждает математические идеи, которые мы пытаемся объяснить в нашей книге, и дает прекрасное изложение физики относительности.

Другая книга BURKE вместе с книгами FLANDERS и SPIVAK могут быть рекомендованы как параллельное и дополнительное чтение.

В главе 6 и потом еще раз в параграфе 10.9 мы касаемся вопросов, относящихся к курсу дифференциальной топологии. Книга GUILLEMIN AND POLLACK написана как лекционный курс, в котором, помимо формального изложения теории, автор использует интуитивный подход и много рисунков. А вот книга BRÖCKER AND LANDER совсем другая. В ней минимум обсуждений, а главным образом точные формулировки теорем и доказательств. И, конечно, ее непросто читать. Эти книги посвящены несколько различным проблемам, и поэтому мы рекомендуем изучить обе.

АРНОЛЬД В. И. Обыкновенные дифференциальные уравнения. — М.: Наука, 1984 (3-е изд.).

АРНОЛЬД В. И. Дополнительные главы теории обыкновенных дифференциальных уравнений. — М.: Наука, 1978.

BRAUN M. Differential Equations and their Applications. An Introduction to Applied Mathematics. — New York: Springer-Verlag, 1993 (4th ed.).

BRÖCKER TH. AND LANDER L. Differentiable Germs and Catastrophes. — Cambridge, UK: Cambridge University Press, 1975. (London Mathematical Sociaty Lecture Note Series, 17.)

BURKE W. L. Spacetime, Geometry, Cosmology. A Series of Books in Astronomy. — Mill Valley, CA: University Science Books, 1980.

BURKE W. L. Applied Differential Geometry. — Cambridge, UK: Cambridge University Press, 1985.

DOYLE P. G. AND SNELL J. L. Random Walks and Electric Networks. — Washington, DC: Mathematical Association of America, 1985. (Carus Mathematical Monographs, 22.)

FEYNMAN R. P., LEIGHTON R. B. AND SANDS M. The Feynman Lectures on Physics. Vol. 1–3. — Reading, MA: Addison-Wesley, 1963–1965.
Русский перевод: ФЕЙНМАН Р. П., ЛЕЙТОН Р. Б. И СЭНДС М. Фейнмановские лекции по физике. Вып. 1–9. — М.: Мир, 1965–1967.

FLANDERS H. Differential Forms with Applications to the Physical Sciences. — New York: Dover, 1989 (2nd ed.). (Dover Books on Advanced Mathematics.)

GUILLEMIN V. W. AND POLLACK A. Differential Topology. — Englewood Cliffs, NJ: Prentice-Hall, 1974.

GUILLEMIN V. W. AND STERNBERG S. Symplectic Techniques in Physics. — Cambridge, UK: Cambridge University Press, 1984 (2nd ed.).

Halmos P. F. Finite-Dimensional Vector Spaces. — New York: Springer-Verlag, 1974 (reprint of the 1958 2nd edition). (Undergraduate Texts in Mathematics.)

Русский перевод: Халмош П. Конечномерные векторные пространства. — М.: Физматгиз, 1963; Ижевск: РХД, 2002.

Hartshorne R. Foundations of Projective Geometry. Lecture Notes, Harvard University, 1966/67. — New York: W. A. Benjamin, 1967.

Hirsch M. W. and Smale S. Differential Equations, Dynamical Systems and Linear Algebra. — New York: Academic Press, 1974. (Pure and Applied Mathematics, 60.)

Hoel P. G., Port S. C. and Stone C. J. Introduction to Probability; Introduction to Statistical Theory; Introduction to Stochastic Processes. — Boston, MA: Houghton Mifflin, 1971; 1971; 1972. (The Houghton Mifflin Series in Statistics.)

Kemeny J. G. and Snell J. L. Finite Markov Chains. — New York: Springer-Verlag, 1976 (reprint of the 1960 original). (Undergraduate Texts in Mathematics.)

Lang S. Linear Algebra. — New York: Springer-Verlag, 1987 (3rd ed.). (Undergraduate Texts in Mathematics.)

Loomis L. H. and Sternberg S. Advanced Calculus. — Boston, MA: Jones and Bartlett, 1990 (revised ed.).

Misner C. W., Thorne K. S. and Wheeler J. A. Gravitation. — San Francisco, CA: W. H. Freeman, 1973.

Русский перевод: Мизнер Ч., Торн К. и Уилер Дж. Гравитация. Т. 1–3. — М.: Мир, 1977.

Moran P. A. P. An Introduction to Probability Theory. — Oxford: Clarendon Press, 1968; corrected reprint 1984.

Simmons G. F. Differential Equations with Applications and Historical Notes. — New York: McGraw-Hill, 1972. (International Series in Pure and Applied Mathematics.)

Spivak M. Calculus on Manifolds. A Modern Approach to Classical Theorems of Advanced Calculus. — New York: W. A. Benjamin, 1965.

Strang G. Linear Algebra and its Applications. — New York: Academic Press, 1980 (2nd ed.).

Русский перевод: Стрэнг Г. Линейная алгебра и ее применения. — М.: Мир, 1980.

Taylor E. F. and Wheeler J. A. Spacetime Physics. — New York: W. H. Freeman, 1992 (2nd ed.).

Предметный указатель

Г

Д

Е

З

И

К

Л

М