

ЛИНЕЙНАЯ АЛГЕБРА  
И ЕЕ ПРИМЕНЕНИЯ

Г. Стрент



Г. Стрент

•  
ЛИНЕЙНАЯ  
АЛГЕБРА  
И ЕЕ  
ПРИМЕНЕНИЯ



LINEAR  
ALGEBRA  
and its  
applications

Gilbert Strang



Massachusetts Institute of Technology

ACADEMIC PRESS

NEW YORK SAN FRANCISCO LONDON

1976

Г. Стренг

ЛИНЕЙНАЯ  
АЛГЕБРА  
И ЕЕ  
ПРИМЕНЕНИЯ



Перевод с английского

Ю. А. КУЗНЕЦОВА

и

Д. М. ФАГЕ

под редакцией

Г. И. МАРЧУКА

ИЗДАТЕЛЬСТВО «МИР»

МОСКВА 1980

Книга отличается от традиционных руководств по линейной алгебре тем, что материал излагается в тесной связи с многочисленными приложениями. В виде отдельных глав представлены метод исключения Гаусса, ортогональные проекции, положительно определенные матрицы, линейное программирование и теория игр. Автор знаком советским читателям по переводу его (в соавторстве с Дж. Фиксом) «Теории метода конечных элементов» (М.: Мир, 1977).

Книга, несомненно, окажется полезной математикам-прикладникам различных специальностей; она заинтересует также и преподавателей, аспирантов и студентов университетов и вузов, преподающих или изучающих линейную алгебру и ее приложения.



*Редакция литературы по математическим наукам*

1702030000

С  $\frac{20203-031}{041(01)-80}$  31-1980

© 1976, by Academic Press, Inc.

© Перевод на русский язык, «Мир», 1980

## От редактора перевода

Традиционные курсы линейной алгебры, читаемые в высших учебных заведениях, и соответствующие учебные пособия, как правило, мало затрагивают прикладную сторону предмета. Но в то же время линейная алгебра служит основой всех методов вычислительной математики, являясь в этом смысле чисто прикладной наукой.

Предлагаемая вашему вниманию книга написана известным американским математиком Гильбертом Стренгом на основе курса лекций для студентов Массачусетского технологического института, который читался им с учетом именно этих обстоятельств, и это оказало существенное влияние как на стиль изложения материала, так и на его выбор. Например, в виде отдельных глав здесь представлены метод исключения Гаусса, положительно определенные матрицы и даже линейное программирование, и в то же время жорданова форма матрицы и линейные преобразования рассматриваются в виде кратких приложений. В книге рассматриваются также вопросы об ортогональном проектировании векторов на подпространства и дается представление о методе конечных элементов, который в настоящее время становится основным средством приближенного решения уравнений математической физики. Отдельная глава посвящена вычислениям с матрицами и, в частности, итерационным методам решения систем линейных алгебраических уравнений, играющим важную роль в вычислительной математике.

Каждая глава содержит большое число примеров и упражнений, которые также призваны способствовать развитию у читателя навыков в решении прикладных задач.

Написанная доступным языком, эта книга, несомненно, окажется полезной для широкого круга читателей: математиков-прикладников, аспирантов и студентов многих специальностей университетов и вузов. Она заинтересует также преподавателей курсов линейной алгебры—как с точки зрения методологии, так и с точки зрения максимальной приближенности теории к приложениям.

Перевод глав 1, 2, 6, 7 выполнен Ю. А. Кузнецовым, глав 3, 4, 5, 8 и приложений—Д. М. Фяге.

*Г. И. Марчук*

## Предисловие

Я считаю, что линейная алгебра преподается сейчас слишком абстрактно. Конечно, это утверждение спорно и, быть может, слишком спорно, чтобы быть верным. Но я убежден, что настоящее руководство должно объяснять существо линейной алгебры и развивать математическое мышление читателей — ведь этот предмет столь же фундаментален, как математический анализ, столь же полезен и имеет такие же богатые приложения. Кроме того, линейная алгебра доступнее анализа, и это обстоятельство слишком важно, чтобы им пренебрегать.

Разумеется, нынешнее состояние дел с линейной алгеброй вполне объяснимо. Ее преподавание дает прекрасную возможность иллюстрировать точность математических рассуждений и построения доказательств. Это достоинство я сознаю, ценю и надеюсь сохранить, и мне всегда было приятно читать лекции именно в таком стиле. Однако, когда я начал экспериментировать в Массачусетском технологическом институте с различными вариантами курса, я обнаружил еще одно его достоинство: преподавание линейной алгебры не только позволяет иллюстрировать единство двух важнейших черт математики — абстрактности и приложимости, но и постоянно побуждает подчеркивать это единство.

Так повелось, что большинство изучающих линейную алгебру вязнет в абстракциях и не доходит до приложений. И очень многие студенты, особенно нематематических отделений, вовсе не выбирают этот курс. Даже самые способные наши студенты приобретают тенденцию к постижению абстракций, но остаются беспомощными в вычислениях — например, они решают системы линейных уравнений по правилу Крамера, а собственные значения понимают только как корни характеристических уравнений. В силу всего этого возникает сильное желание сделать преподавание нашего предмета более полезным и более доступным.

Мы надеемся изложить курс линейной алгебры так, чтобы его изучение приобрело смысл для широких кругов студентов самых разных уровней. Это, конечно, не означает, что мы задумали написать своего рода поваренную книгу по алгебре — предмет заслуживает большего. Мы просто концентрируем внимание не на строгости изложения ради ее самой, а на сути понятий, всюду стараясь скорее объяснить, нежели доказать. Некоторые определения вводятся формально, но многие появляются в процессе обсуждения. Точно так же строги и точны лишь некоторые, а не все доказательства. Разумеется, в каждом случае имеется строгая теория, которая лежит в основе изложения; она должна быть разъяснена и подкреплена примерами.

При построении любого курса имеется специфическая трудность, которую нельзя отложить на более поздний срок: с чего начать курс? Большинство студентов начинают его слушать, уже имея некоторые представления о линейных уравнениях. Тем не менее мы убеждены, что изучение линейной алгебры должно начинаться с основной задачи о решении системы  $n$  уравнений с  $n$  неизвестными, причем решаться эта система должна простейшим и наиболее употребительным способом — методом исключения Гаусса (а не по правилу Крамера!). К счастью, несмотря на простоту этого метода, имеется ряд моментов, которые являются центральными для его понимания и новыми почти для каждого студента. Наиболее важно то, что метод исключения эквивалентен матричному разложению: матрица коэффициентов разлагается в произведение треугольных матриц. Это является прекрасным введением к матричным обозначениям и к правилу умножения матриц.

Другая трудность состоит в правильном выборе темпа изложения. Если предполагать, что операции с матрицами уже знакомы студенту, то материал первой главы нужно излагать не слишком медленно, поскольку следующая глава потребует от читателя значительных усилий. Ее цель состоит в том, чтобы объяснить смысл уравнения  $Ax = b$  глубже, чем позволяет метод исключения. Я считаю, что введение четырех основных подпространств — пространства столбцов матрицы  $A$ , пространства ее строк и их ортогональных дополнений (двух нуль-пространств) — дает эффективный способ построения примеров линейной зависимости и независимости, а также хорошо иллюстрирует идеи базиса, размерности и ранга. Кроме того, с помощью понятия ортогональности обычная геометрия трехмерного пространства естественным образом распространяется на  $n$ -мерный случай. И, разумеется, эти четыре основных подпространства служат ключом к пониманию уравнения  $Ax = b$ .

Главы 1—5 являются сердцевиной курса линейной алгебры.

В них содержится большое число примеров из физики, техники, теории вероятностей и статистики, экономики и биологии. (Здесь рассматривается, в частности, геометрия молекулы метана и даже намечается способ применения факторного анализа в психологии, который мои коллеги по МТИ отказываются излагать студентам!) В то же время ясно, что наша книга не претендует на описание всех возможных применений матриц. Это — всего лишь начальный курс линейной алгебры, и наша цель состоит не в изложении таких применений, а в подготовке к ним. И такая подготовка может быть успешной лишь в том случае, если удастся добиться хорошего понимания теории.

Теория в книге представлена достаточно подробно. После изучения в гл. 2 векторных пространств в гл. 3 мы изучаем проекции и скалярные произведения, в гл. 4 — определители и в гл. 5 — собственные значения. Я считаю, что инженеры и другие лица, интересующиеся приложениями, должны обратить особое внимание на гл. 5, где делается упор на использование диагонализации (включая спектральную теорему); жорданову форму матриц мы отнесли в приложение. Каждая глава оканчивается набором обзорных упражнений и построена таким образом, что последний параграф является дополнительным; это относится также и к § 3.4 о псевдообратных матрицах. Для семестровых или полусеместровых курсов преподаватель должен сделать выбор между положительно определенными матрицами (из гл. 6) или линейными программами (из гл. 8) в зависимости от интересов слушателей; я надеюсь, что § 8.1 и 8.5 дают краткое, но полезное введение в линейное программирование и теорию игр.

Заметим, что книга может послужить основой для трех различных курсов. Первый из них — вычислительная линейная алгебра — должен включать материал всей гл. 1, наиболее существенные факты из гл. 2 — 6 и, наконец, гл. 7 о вычислениях и § 8.2 о симплекс-методе. Второй курс — линейная алгебра для статистиков; в нем должны быть полнее изучены гл. 3 и 6. Третья возможность — рассматривать, как это делают экономисты, неравенства наряду с уравнениями, и тогда нужно возможно скорее перейти от уравнения  $Ax = b$  к линейному программированию и двойственности.

Мы надеемся на благосклонное внимание математиков, которые просто обучают основам линейной алгебры. Это — истинная цель нашей книги и хочется думать, что математиков не отпугнут многочисленные «подсчеты числа операций» и другие замечания вычислительного характера, особенно в гл. 1. С практической точки зрения важность таких замечаний очевидна. Но они имеют и серьезную теоретическую цель — способствовать более детальному изучению процесса исключения посредством фактического подсчета числа шагов. Обычно я прошу аудиторию на первой или второй лекции проделать такой подсчет, и результаты оказываются

ся совершенно непредсказуемыми. Однако нет нужды обсуждать этот или другие вычислительно-ориентированные вопросы в аудитории; любой учебник должен и дополнять, и суммировать материал лекций.

Итак, необходима книга, которая позволит читателям успешно овладеть приложениями и в то же время научит их математике, лежащей в основе этих приложений. Именно такую книгу я и попытался написать.

За помощь в ее написании мне хочется особенно поблагодарить Тома Слобко, который постоянно подбадривал меня, Урсулу, которая с такой милой добротой перепечатывала весь материал, и самое для меня драгоценное—мою семью. Есть еще и более ранний долг, которого я никогда не смогу оплатить,—это долг перед моими родителями. Я посвящаю им свою книгу в надежде, что они поймут, как много они сделали для меня. Спасибо им обоим.

*Гильберт Стрэнг*

# Глава I

## МЕТОД ИСКЛЮЧЕНИЯ ГАУССА

### § 1.1. ВВЕДЕНИЕ

Решение систем линейных уравнений — это центральная задача линейной алгебры. Наиболее важным и в то же время наиболее простым является случай, когда число неизвестных равно числу уравнений. Поэтому мы начнем с задачи, когда задано  $n$  уравнений с  $n$  неизвестными.

В курсе высшей алгебры рассматриваются два в каком-то смысле конкурирующих способа решения систем уравнений. Первым является *метод исключения*. Сначала некоторые кратные первого уравнения системы вычитаются из других уравнений, с тем чтобы устранить из этих уравнений первое неизвестное. В результате возникает меньшая система, состоящая из  $n-1$  уравнений с  $n-1$  неизвестными. Процесс повторяется, пока не останется только одно уравнение с одним неизвестным, которое можно решить непосредственно. Теперь нетрудно произвести обратный ход и определить все другие неизвестные в обратном порядке. Соответствующий пример мы скоро приведем. Второй, более сложный, путь дает идея *определителя*. Существует точная формула, называемая правилом Крамера, которая позволяет вычислить решение (значения неизвестных) как отношение двух определителей порядка  $n$ . Из примеров, приводимых в учебниках (человеческого терпения хватает, как правило, на случаи  $n=3$  или  $n=4$ , но не более), не всегда видно, который путь лучше.

На практике более сложные формулы, содержащие определители, оказываются для вычислителя бедствием, и потому для решения больших систем уравнений постоянно используются алгоритмы исключения. Таким образом, наша первая цель — понять алгоритм, который обычно называется *методом исключения Гаусса*.

Этот алгоритм обманчиво прост и, вероятно, в некоторых частных случаях уже хорошо знаком читателю. Однако имеются четыре аспекта, которые значительно глубже, чем простая техника исключения, и которые (вместе с самим алгоритмом) мы хотим обсудить в этой главе. Перечислим эти аспекты.

(1) Интерпретация метода исключения как разложения матрицы коэффициентов. Мы введем *матричные обозначения* для системы уравнений, записывая  $n$  неизвестных как вектор  $x$ , а  $n$  уравнений в матричной форме как  $Ax = b$ . Тогда метод исключения соответствует разложению матрицы  $A$  в произведение  $LU$  нижней треугольной матрицы  $L$  и верхней треугольной матрицы  $U$ . Это очень полезное замечание, которое представляет основу для дальнейших исследований.

Конечно, мы должны для этого ввести понятия матрицы и вектора, правила их умножения, определить для матрицы  $A$  транспонированную матрицу  $A^T$  и обратную матрицу  $A^{-1}$ .

(2) В большинстве случаев метод исключения осуществляется без каких-либо трудностей или модификаций. Но в некоторых исключительных случаях метод не срабатывает — либо из-за неправильного порядка расположения исходных уравнений, что легко исправляется их перестановкой, либо из-за того, что уравнения  $Ax = b$  не имеют единственного решения. В последнем случае или совсем не существует решения, или решений бесконечно много. Мы хотим выяснить, как в процессе исключения, если он не срабатывает, распознается каждая из этих ситуаций.

(3) Очень важно иметь приблизительный подсчет *числа арифметических действий*, требуемых для решения системы методом исключения. Во многих практических задачах решение вопроса о числе вводимых неизвестных, когда нам приходится балансировать между необходимой точностью математической модели и общим объемом вычислительной работы, регулируется именно подсчетом числа арифметических действий.

(4) Мы также хотим выяснить, хотя бы на *интуитивном уровне*, каким образом решение  $x$  может оказаться восприимчивым к *ошибкам округления*. Для некоторых задач решение оказывается восприимчивым, для других нет. Если первопричина восприимчивости становится понятной, то, как правило, легко догадаться, каким образом ее контролировать. Без такого контроля вычислительная машина может выполнить миллионы операций, округляя каждый результат до фиксированного числа знаков, а полученное «решение» окажется абсолютно бесполезным.

Окончательным результатом этой главы будет алгоритм исключения, который настолько эффективен, насколько это возможно. Это именно тот алгоритм, который постоянно используется на практике в многочисленных приложениях. И в то же время использование матричных обозначений (матрица коэффициентов, матрицы, осуществляющие каждый шаг исключения или перестановку строк, окончательные треугольные сомножители  $L$  и  $U$ ) является одним из основополагающих моментов всей теории.

## § 1.2. ПРИМЕР ПРИМЕНЕНИЯ МЕТОДА ИСКЛЮЧЕНИЯ ГАУССА

Понять проблему легче всего на примере. Рассмотрим систему при  $n = 3$

$$\begin{aligned}2u + v + w &= 1, \\4u + v &= -2, \\-2u + 2v + w &= 7.\end{aligned}\tag{1}$$

Наша задача — найти значения неизвестных  $u$ ,  $v$  и  $w$ , и мы применим для этого метод исключения Гаусса. (Гаусс признан величайшим из математиков, но, разумеется, не из-за этого открытия, на которое ему, вероятно, потребовалось минут десять. Но по иронии судьбы среди всех идей, связанных с его именем, наиболее часто упоминается рассматриваемая нами идея исключения.) Метод начинается с *вычитания кратных первого уравнения из других, чтобы исключить  $u$  из последних двух уравнений*. Нам требуется

- (а) вычесть первое уравнение, умноженное на 2, из второго;
- (б) вычесть первое уравнение, умноженное на  $-1$ , из третьего.

В результате получаем эквивалентную исходной систему уравнений

$$\begin{aligned}2u + v + w &= 1, \\-1v - 2w &= -4, \\3v + 2w &= 8.\end{aligned}\tag{2}$$

Коэффициент 2, который стоит при первом неизвестном  $u$  в первом уравнении, называется *ведущим элементом* первого шага исключения.

На втором шаге исключения мы не трогаем первое уравнение. Другие два уравнения содержат только два неизвестных  $v$  и  $w$ , и, следовательно, к ним можно применить ту же самую процедуру исключения. *Ведущим элементом для этого шага является  $-1$* , а кратные этого второго уравнения будут вычитаться из остальных уравнений (в нашем случае остается только одно третье уравнение), чтобы исключить второе неизвестное  $v$ . Таким образом, мы должны

- (с) вычесть второе уравнение, умноженное на  $-3$ , из третьего.

Теперь процесс исключения завершен (по крайней мере в «прямом» направлении). В результате мы получаем упрощенную систему

$$\begin{aligned}2u + v + w &= 1, \\-1v - 2w &= -4, \\-4w &= -4.\end{aligned}\tag{3}$$

Порядок действий при решении этой системы очевиден. Последнее уравнение дает  $w = 1$ ; подставляя это значение во второе уравнение, мы получаем  $v = 2$ ; наконец, первое уравнение дает  $u = -1$ . Этот простой процесс называется **обратной подстановкой**.

Легко понять, как эту идею исключения можно обобщить на случай  $n$  уравнений с  $n$  неизвестными при сколь угодно больших значениях  $n$ . На первом шаге мы используем кратные первого уравнения, чтобы аннулировать все коэффициенты, лежащие под первым ведущим элементом. На втором шаге аннулируются коэффициенты, лежащие под вторым ведущим элементом и т. д. Наконец, последнее уравнение будет содержать лишь последнее неизвестное. Обратная подстановка позволяет получить ответ в обратном порядке, начиная с последнего неизвестного и кончая первым.

**Упражнение 1.2.1.** Применить метод исключения и обратную подстановку для решения системы

$$\begin{aligned} 2u - 3v &= 3, \\ 4u - 5v + w &= 7, \\ 2u - v - 3w &= 5. \end{aligned}$$

Чему равны ведущие элементы? Выписать три операции, в которых кратное одной строки вычитается из другой.

**Упражнение 1.2.2.** Решить систему

$$\begin{aligned} 2u - v &= 0, \\ -u + 2v - w &= 0, \\ -v + 2w - z &= 0, \\ -w + 2z &= 5. \end{aligned}$$

Теперь мы хотим задать два вопроса. Они могут показаться несколько преждевременными, ведь мы еще только сформулировали рабочий алгоритм, однако ответы на них позволят осветить некоторые особенности данного метода. Первый вопрос: всегда ли процесс исключения приводит к решению, или, иначе говоря, *при каких обстоятельствах этот процесс может не сработать?* Ответ следующий: если ни один из ведущих элементов не равен нулю, то задача имеет единственное решение, и оно может быть найдено с помощью процессов исключения и обратной подстановки. Если же какой-либо из ведущих элементов окажется равным нулю, то следующий шаг процесса исключения становится неосуществимым.

Если, например, первый ведущий элемент оказался равным нулю, то исключение  $u$  из других уравнений будет невозможно. То же самое справедливо для каждого промежуточного шага. Заметим, что промежуточный ведущий элемент может стать равным нулю в течение процесса исключения (как ниже в упраж-

нении 1.2.3), в то время как для первоначальной системы коэффициент на этом месте был отличен от нуля. Грубо говоря, *мы не знаем, будет ли промежуточный ведущий элемент равен нулю, до того момента, пока не осуществим все предыдущие шаги процесса исключения.*

В большинстве случаев сложности, связанные с нулевым ведущим элементом, можно обойти и продолжить процесс исключения вплоть до нахождения единственного решения системы. В других случаях срыв процесса оказывается неизбежным, поскольку система либо не имеет решения, либо имеет их бесконечно много.

Анализ этой «аварийной» ситуации мы отложим на один из последующих разделов книги.

Второй вопрос — очень практический и, я бы сказал, финансовый. *Какое количество арифметических операций требует метод исключения для решения системы  $n$  уравнений с  $n$  неизвестными?* Если  $n$  большое и вместо нас исключение осуществляет вычислительная машина (соответствующую программу вы можете либо где-нибудь взять, либо написать сами), то, поскольку каждый шаг процесса нам известен, мы можем заранее оценить объем машинного времени, требуемого для решения задачи. Не обращая пока внимания на правые части уравнений, подсчитаем количество операций, осуществляемых только в левых частях. Это операции двух видов. Одна из них — это деление на ведущий элемент, чтобы найти число (пусть это будет  $l$ ), на которое надо умножить «ведущее» уравнение для последующего его вычитания из уравнения, лежащего ниже. Затем, когда мы осуществляем это вычитание одного уравнения из другого, мы непрерывно встречаемся с операцией «умножить-вычесть», т. е. члены ведущего уравнения умножаются на  $l$  и вычитаются затем из соответствующих членов уравнения, лежащего ниже.

Предположим, что мы договорились назвать каждое деление и каждое умножение-вычитание простой операцией. В самом начале, когда первое уравнение имеет  $n$  неизвестных, *для получения в первом столбце каждого из нулей требуется  $n$  операций:* одна, чтобы найти множитель  $l$ , и  $n-1$  других, чтобы вычислить новые элементы соответствующей строки. Так как ниже первой лежат  $n-1$  строк и, следовательно, требуется обратить в нуль  $n-1$  элементов первого столбца, находящихся ниже ведущего элемента, то *первый шаг процесса исключения требует  $n(n-1) = n^2 - n$  операций.* После этого шага первый столбец не изменяется. Заметим теперь, что последующие шаги реализуются быстрее, поскольку уравнения постепенно становятся короче; так, на втором шаге мы работаем с  $n-1$  уравнениями с  $n-1$  неизвестными. Когда процесс исключения подходит к  $k$  уравнениям, то для реализации следующего шага требуется только  $k(k-1) = k^2 - k$  операций, чтобы обратить в нуль элементы

столбца, находящегося под ведущим элементом (это эквивалентно первому шагу процесса, когда  $k=n$ ). Таким образом, общее число арифметических операций в левой части уравнений вычисляется по формуле

$$P = (n^2 - n) + \dots + (k^2 - k) + \dots + (1^2 - 1).$$

(Заметим, что последний шаг, когда мы имеем одно уравнение с одним неизвестным, не требует вычислительной работы, что соответствует  $1^2 - 1 = 0$ .) Окончательно

$$P = \sum_1^n k^2 - \sum_1^n k = \frac{1}{3} n \left( n + \frac{1}{2} \right) (n + 1) - \frac{1}{2} n (n + 1) = \frac{n^3 - n}{3}.$$

(Проверка легко осуществляется средствами математического анализа. Интеграл от  $x^2$  от 0 до  $n$  равен  $n^3/3$ , а интеграл от  $x$  от 0 до  $n$  равен  $n^2/2$ , что точно соответствует главным членам обеих сумм.) Если  $n$  оказывается достаточно большим, очень хорошей оценкой числа операций является величина  $P \approx n^3/3$ .

Обратная подстановка идет значительно быстрее. Последнее неизвестное находится с помощью одной операции (деление на последний ведущий элемент), предпоследнее неизвестное требует две операции (умножение-вычитание и затем деление) и т. д.;  $k$ -й шаг требует только  $k$  операций. Таким образом, обратная подстановка в целом требует

$$Q = \sum_{k=1}^n k = \frac{1}{2} n (n + 1) \approx \frac{n^2}{2}$$

операций.

Несколько лет назад почти все математики полагали, что эти числа действительно являются оптимальными, т. е. что систему порядка  $n$  общего вида невозможно решить при помощи меньшего числа умножений, чем  $n^3/3$ . (Были даже теоремы, утверждавшие это, но они были применимы не ко всем возможным методам.) Поразительно, но существовавшее предположение оказалось неверным, и сейчас обнаружен метод, который требует только  $Cn^{\log_2 7}$  операций! К счастью для метода исключения, константа  $C$  этого нового метода оказывается настолько большой, настолько больше он требует операций сложения по сравнению с методом исключения и настолько логически сложной оказывается программа его реализации, что он представляет скорее теоретический интерес, чем практический. По-видимому, до сих пор не известно, может ли показатель степени быть меньше, чем  $\log_2 7$ .

**Упражнение 1.2.3.** Применить метод исключения для решения системы

$$\begin{aligned}u + v + w &= -2, \\ 3u + 3v - w &= 6, \\ u - v + w &= -1.\end{aligned}$$

Когда возникает нулевой ведущий элемент, поменять местами это уравнение с лежащим ниже него и продолжить процесс. Каким числом нужно заменить коэффициент  $-1$  при  $v$  в третьем уравнении, чтобы процесс исключения нельзя было продолжить, т. е. чтобы метод исключения не сработал?

**Упражнение 1.2.4.** Для системы двух уравнений вида

$$\begin{aligned}au + bv &= 0, \\ cu + dv &= 1,\end{aligned}$$

выписать явно  $P=2$  конкретных операций, которые применяются к левой части.

**Упражнение 1.2.5.** При разумных предположениях о быстродействии вычислительной машины и стоимости машинного времени определить, систему какого наибольшего порядка можно решить за 1 доллар и за 1000 долларов? Например, используйте величину  $n^3/3$  для подсчета общего числа операций и сумму в 1000 долларов как цену одного часа работы вычислительной машины, осуществляющей один миллион операций в секунду.

**Упражнение 1.2.6** (необязательное). Обычное вычисление произведения двух комплексных чисел

$$(a + ib)(c + id) = (ac - bd) + i(bc + ad)$$

включает вычисление четырех произведений  $ac$ ,  $bd$ ,  $bc$  и  $ad$ . Не обращая внимания на мнимую единицу  $i$ , можно ли вычислить величины  $ac - bd$  и  $bc + ad$ , используя только три умножения? (Вы можете производить сложения без каких-либо ограничений, например, образовать сумму  $a + b$  перед тем, как производить умножение.)

**Упражнение 1.2.7.** Использовать метод исключения для решения системы

$$\begin{aligned}u + v + w &= 6, \\ u + 2v + 2w &= 11, \\ 2u + 3v - 4w &= 3.\end{aligned}$$

## § 1.3. МАТРИЧНЫЕ ОБОЗНАЧЕНИЯ И УМНОЖЕНИЕ МАТРИЦ

До сих пор в примерах с тремя уравнениями и тремя неизвестными мы могли выписывать все уравнения полностью. Более того, мы могли выписать каждый шаг процесса исключения, т. е. вычитание кратных одной строки из других, которое приводит систему уравнений к более простому виду. Для больших систем этот путь последовательного проведения исключений оказывается, однако, безнадежным; здесь требуется гораздо более краткая форма записи. Сейчас для описания первоначальной системы уравнений мы введем матричные обозначения и определим операцию умножения матриц, чтобы описать операции, упрощающие исходную систему.

Заметим, что в нашем примере

$$\begin{aligned} 2u + v + w &= 1, \\ 4u + v &= -2, \\ -2u + 2v + w &= 7 \end{aligned}$$

имеются три различных типа величин. Это неизвестные  $u$ ,  $v$ ,  $w$ , правые части 1,  $-2$ , 7 и, наконец, набор из девяти числовых коэффициентов левой части, один из которых оказался равным нулю. Для столбца чисел правой части (это так называемые *свободные члены уравнений*) мы вводим векторное обозначение

$$b = \begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix}.$$

Это трехмерный вектор-столбец. Чтобы представить его геометрически, мы можем считать три его компоненты координатами точки в трехмерном пространстве. Обратное, каждой точке в таком пространстве соответствует трехмерный вектор (который мы можем представлять себе как стрелку или отрезок прямой, начинающийся в начале координат и оканчивающийся в этой точке).

*Основными операциями над векторами являются сложение двух векторов и умножение вектора на число.* Геометрически вектор  $2b$  имеет то же самое направление, что и вектор  $b$ , но будет в два раза длиннее, вектор  $-2b$  имеет противоположное направление, а вектор  $b + c$  получится, если поместить начальную точку вектора  $c$  в конечную точку вектора  $b$ . Алгебраически это означает, что операции над векторами осуществляются покомпонентно:

$$\begin{aligned} 2b &= 2 \begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \\ 14 \end{bmatrix}, & -2b &= \begin{bmatrix} -2 \\ 4 \\ -14 \end{bmatrix}, \\ b + c &= \begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix} + \begin{bmatrix} 1 \\ -4 \\ -4 \end{bmatrix} = \begin{bmatrix} 2 \\ -6 \\ 3 \end{bmatrix}. \end{aligned}$$

Два вектора могут складываться только в том случае, когда они имеют одинаковую размерность, равную числу их компонент.

Три неизвестных в уравнении также можно записать в векторном виде:

$$\text{вектор неизвестных } x = \begin{bmatrix} u \\ v \\ w \end{bmatrix}; \quad \text{решение } x = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}.$$

Они вновь являются трехмерными вектор-столбцами. Для расположения девяти коэффициентов уравнений мы введем так называемую «матрицу», состоящую из трех строк и трех столбцов. Она называется **матрицей коэффициентов** и имеет вид

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix}.$$

Нужно заметить, что поскольку в нашем примере число неизвестных равно числу уравнений,  $A$  является *квадратной матрицей* порядка три. В более общем случае, когда имеется  $n$  уравнений с  $n$  неизвестными,  $A$  будет квадратной матрицей порядка  $n$ . Наконец, в самом общем случае, когда имеется  $m$  уравнений с  $n$  неизвестными, матрица коэффициентов имеет  $m$  строк и  $n$  столбцов и называется матрицей размера  $m \times n$  или  $(m \times n)$ -матрицей. Так же как и векторы, матрицы можно складывать между собой и умножать на числа, причем эти операции осуществляются покомпонентно. Операции над векторами являются частным случаем операций над матрицами, поскольку векторы можно рассматривать как матрицы, только с одним столбцом. Как и ранее, нужно отметить, что сложение двух матриц возможно только в том случае, когда они имеют одинаковый размер:

$$\begin{bmatrix} 2 & 1 \\ 3 & 0 \\ 0 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ -3 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 0 & 1 \\ 1 & 6 \end{bmatrix}, \quad 2 \begin{bmatrix} 2 & 1 \\ 3 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 6 & 0 \\ 0 & 8 \end{bmatrix}.$$

### УМНОЖЕНИЕ ВЕКТОРА НА МАТРИЦУ

Теперь мы начнем использовать введенные понятия. Мы хотим переписать систему (1) трех уравнений с тремя неизвестными в более простой матричной форме  $Ax = b$ . Это выглядит так:

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix}.$$

Правая часть не вызывает сомнений—это вектор-столбец со свободными членами системы (1) в качестве компонент, а левая часть состоит из вектора  $x$ , умноженного слева на матрицу  $A$ . Тот факт, что *это запись нашей системы*, уже подсказывает нам, как будет далее определено произведение вектора на матрицу. А именно, первая компонента вектора  $Ax$  вычисляется как «про-

изведение» первой строки матрицы  $A$  на вектор-столбец  $x$ :

$$[2 \quad 1 \quad 1] \begin{bmatrix} u \\ v \\ w \end{bmatrix} = [2u + v + w]. \quad (4)$$

Приравнивая полученное выражение единице, первой компоненте вектора  $b$ , мы получаем первое уравнение  $2u + v + w = 1$  нашей системы. Вторая компонента произведения  $Ax$ , равная  $4u + v$ , определяется второй строкой матрицы  $A$ , а третья компонента, равная  $-2u + 2v + w$ , — третьей строкой этой матрицы. Таким образом, матричное уравнение  $Ax = b$  полностью эквивалентно системе из трех уравнений, с которой мы начинали рассмотрение.

Операция (4) является основополагающей для умножения матриц. Исходными величинами для нее являются вектор-строка и вектор-столбец одинаковой длины, а результатом — обычное число. Это число называется *скалярным (внутренним) произведением двух векторов*. Иначе говоря, произведение матрицы размера  $1 \times n$  (вектор-строки) на матрицу размера  $n \times 1$  (вектор-столбец) является матрицей размера  $1 \times 1$  (матрицей порядка один):

$$[a_1 \quad a_2 \quad \dots \quad a_n] \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix} = [a_1b_1 + a_2b_2 + \dots + a_nb_n].$$

### Пример

$$[2 \quad 4 \quad 1] \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix} = [2].$$

Правило умножения вектора на матрицу непосредственно распространяется с рассмотренного нами случая матрицы порядка 3 на произвольный случай матрицы размера  $n \times n$  (порядка  $n$ ). При этом вектор обязательно должен иметь  $n$  компонент, т. е. его размерность должна совпадать с порядком матрицы. Если использовать буквы вместо конкретных величин для обозначения элементов матрицы  $A$  и компонент вектора  $x$ , то их произведе-

ние  $Ax$  записывается в виде

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \dots + a_{1n}x_n \\ a_{21}x_1 + \dots + a_{2n}x_n \\ \vdots + \dots + \vdots \\ a_{n1}x_1 + \dots + a_{nn}x_n \end{bmatrix}.$$

Обратим внимание на элементы  $a_{ij}$  матрицы  $A$ . Первый индекс элемента  $a_{ij}$  указывает номер строки, а второй — номер столбца, и, следовательно, этот элемент лежит на пересечении  $i$ -й строки и  $j$ -го столбца. Заметим также, что для записи  $i$ -й компоненты вектора  $Ax$  можно использовать сокращенное обозначение суммы

$$\sum_{j=1}^n a_{ij}x_j.$$

С такими обозначениями работать легче, чем с полностью выписанной суммой, но все же они приносят меньше пользы, чем сами матричные обозначения.

### Пример

$$A = \begin{bmatrix} 2 & 3 \\ 4 & 0 \end{bmatrix}, \quad x = \begin{bmatrix} 1 \\ 5 \end{bmatrix}.$$

$$Ax = \begin{bmatrix} (2)(1) + (3)(5) \\ (4)(1) + (0)(5) \end{bmatrix} = \begin{bmatrix} 17 \\ 4 \end{bmatrix}. \quad (5)$$

Здесь вектор  $Ax$  оказывается линейной комбинацией двух столбцов матрицы  $A$ , каждый из которых входит в эту комбинацию с коэффициентом, равным соответствующей компоненте вектора  $x$ . Иначе говоря, вектор  $Ax$  может быть вычислен целиком по формуле

$$Ax = \begin{bmatrix} 2 \\ 4 \end{bmatrix} (1) + \begin{bmatrix} 3 \\ 0 \end{bmatrix} (5) = \begin{bmatrix} 17 \\ 4 \end{bmatrix} \quad (6)$$

вместо приведенных ранее покомпонентных вычислений. Такой подход далее будет использоваться всюду, и потому мы рассмотрим его более детально. Итак, произведение  $Ax$  может быть вычислено либо с использованием отдельных элементов матрицы  $A$ , как в (5), либо *с использованием целых ее столбцов*:

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} = (-1) \begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Таким образом, произведение  $Ax$  является линейной комбинацией столбцов матрицы  $A$ .

В последнем рассмотренном примере выписанная комбинация столбцов матрицы  $A$  равна вектору

$$b = \begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix} \text{ и, следовательно, вектор } x = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

является решением системы  $Ax = b$ .

**Упражнение 1.3.1.** Вычислить, сначала используя отдельные элементы, а затем — целые столбцы матрицы  $A$ , произведение

$$Ax = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}.$$

**Упражнение 1.3.2.** Вычислить произведение

$$\begin{bmatrix} 1 & 2 \\ 3 & -3 \\ 0 & 4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ и } [-4 \ 1 \ 3] \begin{bmatrix} -4 \\ 1 \\ 3 \end{bmatrix}.$$

Если  $A$  является матрицей размера  $m \times n$ , а  $x$  — вектором размерности  $n$ , то какими будут размер и форма произведения  $Ax$ ?

**Упражнение 1.3.3.** Предположим, что, сравнивая начало и конец 1977 года, мы получили следующие данные:

а) из тех, кто в начале года проживали в Калифорнии, 20% выехали, а 80% остались в ней;

б) из тех, кто в начале года не проживали в Калифорнии, 10% переехали в Калифорнию, а 90% либо остались там, где они проживали, либо переехали в другой штат (не Калифорнию).

Следуя этим предположениям, записать соответствующую систему уравнений в матричном виде и ответить на следующие вопросы:

(i) Если в начале года 30 миллионов проживали в Калифорнии, а 200 миллионов проживали вне нее, то какая ситуация сложится в конце года?

(ii) Если в конце года в Калифорнии проживали 30 миллионов, а вне нее 200 миллионов, то какова была ситуация в начале года?

(iii) Каково должно быть распределение проживающих в начале года, чтобы оно не изменилось к концу года? Иначе говоря, каково должно быть соотношение в начале года между числом  $u$  проживающих в Калифорнии и числом  $v$  проживающих вне нее, чтобы к концу года значения  $u$  и  $v$  не изменились?

**Упражнение 1.3.4.** Начертить пару перпендикулярных осей и отметить точки  $x=2$ ,  $y=1$  и  $x=0$ ,  $y=3$ . Провести векторы из начала координат к этим точкам, а также к их сумме  $\begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix}$ . Построить соответствующий параллелограмм.

Дополним упражнение 1.3.3 двумя замечаниями. Во-первых, нужно подчеркнуть *линейность* задачи. Если числа в задаче увеличить (удвоить) соответственно до 60 и 400 миллионов, то ответ также удвоится. Далее, если мы решим дополнительно задачу для 40 и 250 миллионов, то решение для случая 70 и 450 миллионов будет в точности суммой полученных двух решений.

Второе замечание касается вопроса (ii) этого упражнения, где надо вычислить числа  $u$  и  $v$  людей, находившихся соответственно внутри Калифорнии и вне нее в начале года. Очевидно, что вам были нужны два уравнения и вы их, вероятно, уже построили: первое уравнение имеет вид  $0.8u + 0.1v = 30$  и гарантирует проживание 30 миллионов в конце года в Калифорнии, а второе должно гарантировать проживание 200 миллионов вне нее. Наше замечание следующее: идея производить операции «постолбцово» позволяет рассмотреть эти уравнения одновременно. Правой частью этой системы будет вектор  $b = \begin{bmatrix} 30 \\ 200 \end{bmatrix}$ , указывающий число проживающих в Калифорнии и вне нее в конце года. Вектор  $\begin{bmatrix} 0.8u \\ 0.2u \end{bmatrix}$  указывает распределение к концу года  $u$  людей, проживавших в начале года в Калифорнии, а вектор  $\begin{bmatrix} 0.1v \\ 0.9v \end{bmatrix}$  — распределение к этому же времени  $v$  людей, проживавших вне Калифорнии. Сумма этих векторов должна быть равна вектору

$$u \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix} + v \begin{bmatrix} 0.1 \\ 0.9 \end{bmatrix} = \begin{bmatrix} 30 \\ 200 \end{bmatrix}.$$

Таким образом, с «постолбцовых» позиций мы вычисляем такие числа  $u$  и  $v$ , чтобы комбинация указанных столбцов с этими коэффициентами в левой части давала в правой части вектор  $b$ .

В вопросе (iii) требуется подобрать столбец  $\begin{bmatrix} u \\ v \end{bmatrix}$  так, чтобы правая часть была равна ему же, т. е. чтобы год заканчивался так же, как и начинался.

#### МАТРИЧНЫЙ ВИД ОДНОГО ШАГА ИСКЛЮЧЕНИЯ

К настоящему моменту мы получили короткую запись  $Ax = b$  для первоначальной системы уравнений. Что же можно сказать относительно операций, осуществляемых в процессе исключения? В нашем примере первый шаг заключался в вычитании первого уравнения, умноженного на два, из второго. Для правой части это означает, что удвоенная первая компонента  $b$  вычитается из второй. Мы утверждаем, что *тот же самый результат дости-*

гается умножением вектора  $b$  на матрицу специального вида

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Это проверяется с помощью правила умножения вектора на матрицу:

$$Eb = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix} = \begin{bmatrix} 1 \\ -4 \\ 7 \end{bmatrix}.$$

Первая и третья компоненты вектора  $b$  (1 и 7) остаются неизменными (это связано со специальным выбором первой и третьей строк матрицы  $E$ ). Новая вторая компонента равна  $-4$  (она появляется в уравнениях (2) после первого шага исключения).

Чтобы сохранить равенства, мы должны применить операцию умножения на  $E$  к обеим частям системы  $Ax = b$ . Иначе говоря, мы должны также умножить вектор  $Ax$  слева на матрицу  $E$ . Эта операция снова вычитает удвоенную первую компоненту из второй, оставляя неизменными первую и третью. После осуществления операции умножения системы на  $E$  новая более простая система (эквивалентная исходной) имеет вид  $E(Ax) = Eb$ . Эта система проще, так как в матрице  $EA$  возник нулевой элемент под первым ведущим элементом, и эквивалентна исходной, поскольку мы можем восстановить исходную систему, добавив ко второму уравнению новой системы удвоенное первое уравнение. Поэтому обе системы имеют одно и то же решение  $x$ . Это обстоятельство оправдывает применение метода исключения Гаусса, так как от шага к шагу решение не изменяется, а найти его становится все легче.

#### УМНОЖЕНИЕ МАТРИЦ

Теперь мы хотим спросить, какова же будет матрица коэффициентов новой системы  $E(Ax) = Eb$ ? Простой шаг метода исключения Гаусса, осуществляемый с помощью матрицы  $E$ , преобразует первоначальную матрицу коэффициентов

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix}$$

в новую матрицу

$$EA = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -1 & -2 \\ -2 & 2 & 1 \end{bmatrix}.$$

Как и ранее, первая и третья строки матрицы не изменяются, в то время как удвоенная первая строка вычитается из второй. Новая матрица коэффициентов действует на  $x$  так же, как если бы сначала умножить его на матрицу  $A$ , а затем  $Ax$  умножить на  $E$ . Эта новая матрица называется произведением двух исходных матриц и обозначается через  $EA$ . Здесь нужно отметить, что обозначение сохраняет порядок, в котором появлялись эти матрицы. Повторим изложенное:

1А. Произведением  $EA$  матриц  $E$  и  $A$  называется такая матрица, умножение произвольного вектора на которую эквивалентно умножению этого вектора сначала на матрицу  $A$ , а затем на матрицу  $E$ . Иначе говоря, для любого вектора  $x$

$$(EA)x = E(Ax). \quad (7)$$

Это правило определяет законы умножения матриц, которые иллюстрируются рис. 1.1 и соотношением (8).

Предположим, что нам заданы две, возможно прямоугольные, матрицы  $E$  и  $A$ . Как использовать сформулированное правило для вычисления их произведения  $EA$ ? Во-первых, форма матриц  $E$  и  $A$  должна быть такой, чтобы их произведение имело смысл. Если они являются квадратными, как в нашем примере, то они обязательно должны иметь одинаковые порядки. Если же матрицы  $E$  и  $A$  прямоугольные, то они *не обязательно* должны иметь одинаковые размеры, но число столбцов матрицы  $E$  должно быть равно числу строк матрицы  $A$ . Иначе говоря, если  $E$  — матрица размера  $l \times t$  и  $A$  — матрица размера  $t \times n$ , т. е.  $E$  имеет  $t$  столбцов, а  $A$  имеет  $t$  строк, то их перемножение возможно. Возьмем некоторый вектор  $x$  размерности  $n$ , затем вычислим вектор  $Ax$  размерности  $t$  и, наконец, вектор  $EAx$ , имеющий  $l$  компонент. (В выражении  $EAx$  мы опустили скобки, поскольку правило (7) выбрано именно так, чтобы сделать их ненужными.) Матрица  $EA$  имеет  $l$  строк, как и матрица  $E$ , и  $n$  столбцов, как и матрица  $A$ , т. е. произведение матриц размера  $l \times t$  и  $t \times n$  является матрицей размера  $l \times n$ .

Теперь мы должны показать, какой же будет матрица  $EA$  в действительности. Обычный способ — это описать каждый отдельный элемент матрицы  $EA$ , например элемент, лежащий в  $i$ -й строке и  $j$ -м столбце: этот элемент является скалярным произведением  $i$ -й строки матрицы  $E$  и  $j$ -го столбца матрицы  $A$ . Так, скалярное произведение последней строки из  $E$  и последнего столбца из  $A$ , отмеченных на рис. 1.1, дает последний элемент матрицы  $EA$ :

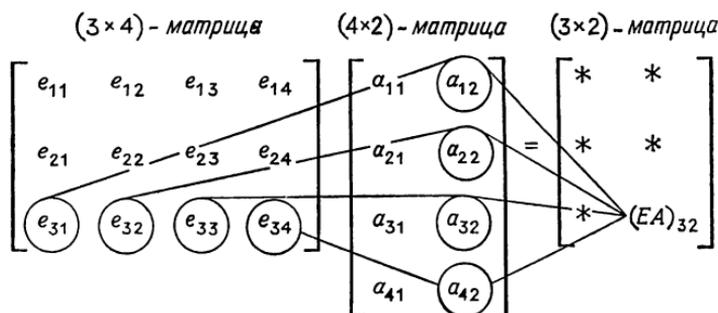


Рис. 1.1. Иллюстрация умножения матриц.

$$(EA)_{32} = [e_{31} \ e_{32} \ e_{33} \ e_{34}] \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ a_{42} \end{bmatrix} = e_{31}a_{12} + e_{32}a_{22} + e_{33}a_{32} + e_{34}a_{42}. \quad (8)$$

**Пример 1.**

$$\begin{bmatrix} 2 & 3 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 5 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 17 & 1 & 0 \\ 4 & 8 & 0 \end{bmatrix}.$$

Например, элемент  $17 = (2)(1) + (3)(5)$  является скалярным произведением первой строки матрицы  $E$  и первого столбца матрицы  $A$ , а элемент  $8 = (4)(2) + (0)(-1)$  — скалярным произведением второй строки из  $E$  и второго столбца из  $A$ . Третий столбец матрицы  $A$  нулевой и, следовательно, третий столбец матрицы  $EA$  оказывается нулевым.

**Пример 2.**

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 7 & 8 \\ 2 & 3 \end{bmatrix}.$$

Здесь матрица  $E$  осуществляет перестановку строк матрицы  $A$ .

**Упражнение 1.3.5.** Вычислить произведение матриц

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{и} \quad A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix}.$$

**Упражнение 1.3.6.** Вычислить произведение матриц

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -5 & 3 & 1 \end{bmatrix} \quad \text{и} \quad A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix}.$$

**Упражнение 1.3.7.** Перемножить матрицы  $E = [1 \ -4]$  размера  $1 \times 2$  и  $A = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$  размера  $2 \times 1$ . Объяснить, почему для вычисления матрицы  $EA$ , где

$E$  — матрица размера  $l \times m$ , а  $A$  — матрица размера  $m \times n$ , требуется  $l \times n$  отдельных умножений.

**Упражнение 1.3.8.** Перемножить матрицы

$$E = \begin{bmatrix} 8 & -3 \\ -5 & 2 \\ 1 & 0 \end{bmatrix} \quad \text{и} \quad A = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

размеров  $3 \times 2$  и  $2 \times 1$  соответственно.

Заметим, что последнее упражнение заключается в умножении вектора на матрицу. *Правило вычисления произведения матриц  $EA$  полностью соответствует правилу вычисления произведения  $Ax$ .*

Более того, так же как произведение  $Ax$ , произведение матриц  $EA$  можно вычислить, используя сразу целые столбцы. Например, первый столбец произведения  $EA$  является произведением первого столбца матрицы  $A$  на матрицу  $E$ . Действительно, взгляните снова на пример

$$\begin{bmatrix} 2 & 3 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 5 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 17 & 1 & 0 \\ 4 & 8 & 0 \end{bmatrix},$$

где, вычисляя произведения столбец за столбцом, имеем

$$\begin{bmatrix} 2 & 3 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 17 \\ 4 \end{bmatrix}, \quad \begin{bmatrix} 2 & 3 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \end{bmatrix}, \quad \begin{bmatrix} 2 & 3 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Приведенный пример иллюстрирует важное положение:  *$j$ -й столбец матрицы  $EA$  зависит только от  $j$ -го столбца матрицы  $A$ , но не от других ее столбцов.* Кроме того, этот столбец матрицы  $EA$  является линейной комбинацией всех столбцов матрицы  $E$  с коэффициентами  $a_{ij}$  — элементами  $j$ -го столбца матрицы  $A$ . Таким образом, правило перемножения двух матриц действительно является расширением введенного ранее правила перемножения матрицы и вектор-столбца с единственным отличием, что вместо одного столбца мы имеем дело последовательно с несколькими столбцами.

Проведенное рассмотрение очень полезно для понимания одного из наиболее важных свойств умножения матриц. Предположим, что заданы три матрицы  $A$ ,  $B$  и  $C$ , возможно прямоугольные, и предположим, что размеры матриц позволяют последовательно их перемножить, т. е. число столбцов матриц  $A$  и  $B$  равно числу строк матриц  $B$  и  $C$  соответственно. Тогда отмеченное свойство формулируется следующим образом.

|| **1B.** Умножение матриц ассоциативно:  $(AB)C = A(BC)$ .

Если  $C$  окажется вектором (иначе говоря, матрицей только с одним столбцом), то это свойство совпадает с (7), которое было

основой введенного нами правила умножения матриц. Если же  $C$  имеет несколько столбцов  $c_1, \dots, c_n$ , то мы должны последовательно применить это правило несколько раз. С одной стороны, первым столбцом матрицы  $(AB)C$  является вектор-столбец  $(AB)c_1$ , а, с другой стороны, первым столбцом матрицы  $BC$  будет вектор-столбец  $Bc_1$  и, следовательно, первым столбцом матрицы  $A(BC)$  является вектор-столбец  $A(Bc_1)$ . Так как, согласно (7),  $(AB)c_1 = A(Bc_1)$ , то первые столбцы матриц  $(AB)C$  и  $A(BC)$  совпадают, и нам остается провести подобное же рассмотрение для остальных столбцов.

**Упражнение 1.3.9.** Проверить закон ассоциативности для матриц

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix}.$$

Наша цель — установление связи между методом исключения Гаусса и умножением матриц, но для этого нам потребуется еще два свойства, одним из которых операция матричного умножения обладает, а другим *нет*.

|| **1С.** Операции над матрицами дистрибутивны:

$$A(B+C) = AB+AC, \quad (B+C)D = BD+CD.$$

Конечно, размеры этих матриц должны быть соответствующим образом согласованы:  $B$  и  $C$  должны иметь одинаковые размеры, чтобы их сумма имела смысл, а матрицы  $A$  и  $D$  — иметь нужные размеры для умножения слева и справа.

Свойство, которым умножение матриц не обладает, несколько более интересно:

|| **1D.** Умножение матриц не коммутативно, т. е., как правило,

$$FE \neq EF.$$

**Пример.** Предположим, что  $E$  — введенная ранее матрица, умножение на которую вычитало удвоенное первое уравнение из второго:

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

а  $F$  — матрица, с которой мы встречаемся на последнем шаге процесса исключения, когда второе уравнение, умноженное на  $-3$ , вычитается из третьего (утроенное второе уравнение складывается

с третьим),

$$F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix}.$$

Сравним произведение

$$FE = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -6 & 3 & 1 \end{bmatrix} \quad \text{с} \quad EF = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix}.$$

Очевидно, что результат умножения зависит от порядка сомножителей. В первом случае, когда мы применяем к системе сначала  $E$ , а затем  $F$ , второе уравнение изменяется на удвоенное первое *раньше*, чем оно используется для изменения третьего уравнения. За счет этого на месте  $(3, 1)$  появился элемент  $-6$ . Такой порядок и имел место в процессе исключения. Если же, наоборот, сначала применяется матрица  $F$ , то элемент  $-6$  не возникает — при таком порядке первое уравнение не оказывает какого-либо влияния на третье. Таким образом,  $FE \neq EF$ .

#### ЕДИНИЧНАЯ МАТРИЦА

Существует одна очень важная матрица порядка  $n$ , которая коммутирует с любой матрицей такого же порядка. В действительности при умножении на нее другая матрица не меняется, так как ее действие аналогично умножению на единицу. Матрица, обладающая таким свойством, называется *единичной*; все ее диагональные элементы равны единице, а остальные элементы — нулю. Для случая  $n = 4$  она имеет вид

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Легко проверить, что  $IA = AI = A$  для любой матрицы  $A$  такого же порядка.

**Упражнение 1.3.10.** Показать, что среди всех матриц  $A$  вида  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$  с двумя матрицами

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{и} \quad C = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

одновременно коммутируют лишь матрицы, которые кратны единичной, т. е. такие, что  $a = d$  и  $b = c = 0$ . Проверить, что такие матрицы коммутируют с лю-

бой матрицей порядка 2 и, таким образом, это единственные матрицы с таким свойством.

**Упражнение 1.3.11.** Привести примеры таких матриц порядка 2, чтобы

- (a)  $A^2 = -I$  и  $A$  имела только вещественные элементы,  
 (b)  $B^2 = 0$ , хотя  $B \neq 0$ ,  
 (c)  $CD = -DC$ , причем  $CD \neq 0$ ,  
 (d)  $EF = 0$ , хотя ни одна из матриц  $E, F$  не имеет нулевых элементов.

**Упражнение 1.3.12.** Справедливы ли следующие утверждения (если утверждение не верно, то привести контрпример)?

- (a) Если первый и третий столбцы матрицы  $B$  одинаковы, то первый и третий столбцы матрицы  $AB$  также одинаковы.  
 (b) Если первая и третья строки матрицы  $B$  одинаковы, то первая и третья строки матрицы  $AB$  также одинаковы.  
 (c) Если первая и третья строки матрицы  $A$  одинаковы, то первая и третья строки матрицы  $AB$  также одинаковы;  
 (d)  $(AB)^2 = A^2B^2$ .

**Упражнение 1.3.13.** Первая строка произведения  $AB$  является линейной комбинацией всех строк матрицы  $B$ . Каковы коэффициенты в этой комбинации и какова первая строка матрицы  $AB$ , если

$$A = \begin{bmatrix} 2 & 1 & 4 \\ 0 & -1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}?$$

## § 1.4. ЭКВИВАЛЕНТНОСТЬ МЕТОДА ИСКЛЮЧЕНИЯ ГАУССА И РАЗЛОЖЕНИЯ НА ТРЕУГОЛЬНЫЕ МАТРИЦЫ

Вернемся опять к методу исключения и посмотрим, как соответствующие операции выражаются через матрицы. Нашей отправной точкой была система  $Ax = b$ :

$$Ax = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix}.$$

Затем было три шага исключения:

- (i) вычитание удвоенного первого уравнения из второго;  
 (ii) вычитание первого уравнения, умноженного на  $-1$ , из третьего;  
 (iii) вычитание второго уравнения, умноженного на  $-3$ , из третьего.

В результате этих исключений получалась эквивалентная, но более простая система с новой матрицей коэффициентов, которую мы обозначили через  $U$ :

$$Ux = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ -4 \\ 4 \end{bmatrix}. \quad (9)$$

Полученная матрица коэффициентов является **верхней треугольной**—все ее элементы, лежащие ниже главной диагонали, равны нулю.

Правая часть, являющаяся также новым вектором  $c$ , получена из первоначального вектора  $b$  с помощью тех же самых шагов, которые использовались при преобразовании  $A$  в  $U$ . Просуммируем этапы процесса исключения Гаусса:

- (a) начинаем с матрицы  $A$  и вектора  $b$ ;
- (b) реализуем шаги исключения (i), (ii), (iii);
- (c) заканчиваем матрицей  $U$  и вектором  $c$ .

Самый последний шаг заключается в решении системы  $Ux = c$  с помощью обратной подстановки, но мы пока отложим этот вопрос в сторону, сосредоточив наше внимание на связи между матрицами  $A$  и  $U$ .

Матрица  $E$ , которая реализует шаг (i), вычитая удвоенную первую строку из второй, уже была введена нами в предыдущем разделе. Мы обозначим ее через  $E_{21}$ , указывая индексами, что она изменяет вторую строку с помощью кратного первой строки, а также подчеркивая, что результатом умножения на нее будет появление у матрицы коэффициентов нулевого элемента на месте (2, 1). Напомним, что поскольку первая строка умножалась на 2, то матрица, вычитающая из второй строки первую с этим коэффициентом, имеет вид

$$E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Аналогично шаги исключения (ii) и (iii) могут быть описаны с помощью матриц. Точнее говоря, на этих шагах мы умножаем на матрицы

$$E_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{и} \quad E_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix}.$$

Эти матрицы называются **элементарными матрицами**, и легко видеть, как они определяются и действуют в общем случае. Пусть требуется вычесть  $j$ -е уравнение с множителем  $l_{ij}$  из  $i$ -го уравнения. Тогда для построения элементарной матрицы мы берем единичную матрицу  $I$  и заменяем ее нулевой элемент в  $i$ -й строке и  $j$ -м столбце на величину  $-l_{ij}$ . Заметим, что элементарные матрицы являются **нижними треугольными** с единицами на главной диагонали.

Преобразование матрицы  $A$  в  $U$  с помощью трех матричных операций можно выразить соотношением

$$E_{32}E_{31}E_{21}A = U. \quad (10)$$

Аналогично, поскольку с помощью тех же операций преобразуется свободный член,

$$E_{32}E_{31}E_{21}b = c. \quad (11)$$

Мы могли бы, если бы захотели, перемножить все эти матрицы  $E$ , чтобы получить одну матрицу, преобразующую  $A$  и  $b$  в  $U$  и  $c$ :

$$E_{32}E_{31}E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -5 & 3 & 1 \end{bmatrix}. \quad (12)$$

Заметим, что она опять оказывается нижней треугольной матрицей, как это уже было в упражнении 1.3.6. Закон ассоциативности оказывается здесь очень полезным: мы можем вычислить произведение всех матриц  $E$  и умножить на него матрицу  $A$ , получив при этом тот же результат, как если бы умножали  $A$  последовательно на матрицы  $E$ . Обратим внимание на элемент  $-5$  в нижнем левом углу. Он не является одним из множителей в процессе исключения, но комбинирование шагов (i) и (iii) дает  $-6$ , а шаг (ii) добавляет  $+1$ .

Теперь мы пришли к важному вопросу: как можно вернуться назад от матрицы  $U$  к  $A$ ? **Как можно обратить шаги исключения?**

Простой шаг, скажем шаг (i), обратить нетрудно. Вместо вычитания мы *добавим* удвоенную первую строку ко второй (не удвоенную вторую строку к первой!). Тогда элементарная матрица  $E_{21}$  заменяется другой элементарной матрицей с величиной  $+2$  на том же месте, которое перед этим занимала величина  $-2$ ,

$$E_{21}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Произведение  $E_{21}$  и  $E_{21}^{-1}$ , взятое в любом порядке, является единичной матрицей: одна операция сокращается с другой. В терминах матриц каждая из этих матриц является *обратной к другой*:

$$E_{21}^{-1}E_{21} = I, \quad E_{21}E_{21}^{-1} = I. \quad (13)$$

Аналогично вторая и третья элементарные матрицы могут быть обращены добавлением того, что вычиталось на шагах (ii) и (iii):

$$E_{31}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \quad E_{32}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix}.$$

Таким образом, мы построили матрицы, которые обращают каждый отдельный из шагов исключения (i), (ii) и (iii). Наша задача — обратить весь процесс и посмотреть, какая матрица преобразует  $U$  в  $A$ . Заметим, что *поскольку шаг (iii) был последним в процессе преобразования  $A$  в  $U$ , то он должен быть обращен первым, когда мы идем в обратном направлении.* Это общее правило: обращение осуществляется в обратном порядке по отношению к первоначальной последовательности операций. (Мы еще вернемся к этому вопросу в следующем параграфе, когда обращения будут обсуждаться более систематически.) Вторым обращается шаг (ii) и последним — шаг (iii). В результате, начиная с матрицы  $U$ , мы приходим назад к матрице  $A$ :

$$A = E_{21}^{-1} E_{31}^{-1} E_{32}^{-1} U. \quad (14)$$

Читатель может мысленно подставить  $U = E_{32} E_{31} E_{21} A$  в это уравнение и увидеть, что используемый порядок расположения матриц правильный, поскольку все произведения типа  $E_{32}^{-1} E_{32}$  равны единичной матрице.

Вычислим теперь матрицу  $L$ , которая преобразует  $U$  в  $A$ . Она является произведением трех матриц, которые обращают отдельные шаги, и уже участвует в уравнении (14):

$$L = E_{21}^{-1} E_{31}^{-1} E_{32}^{-1}, \quad \text{так что} \quad A = LU. \quad (15)$$

Построенная матрица  $L$  и матрица  $E_{32} E_{31} E_{21}$ , преобразующая  $A$  в  $U$ , обратны друг другу, т. е.

$$L = (E_{32} E_{31} E_{21})^{-1}, \quad \text{или} \quad L^{-1} = E_{32} E_{31} E_{21}. \quad (16)$$

Мы получили ключевые формулы, но самую важную особенность матрицы  $L$  можно заметить, только произведя перемножение:

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -3 & 1 \end{bmatrix}. \quad (17)$$

Очевидно, что  $L$  — нижняя треугольная матрица с единицами на главной диагонали. Ее особенность состоит в том, что *элементы ниже главной диагонали — это в точности множители 2,  $-1$  и  $-3$ , используемые на трех шагах исключения.* Обычно, когда перемножаются три матрицы, нельзя прямо выписать результат, используя отдельные элементы этих трех матриц. Но в настоящем случае матрицы расположены в таком идеальном порядке, что их произведение может быть выписано сразу. Если в вычислительной машине хранятся все  $l_{ij}$ -множители, на которые умножается  $j$ -я строка перед вычитанием из  $i$ -й строки, чтобы получить нуль в позиции  $(i, j)$ , то эти величины не только дают полную характеристику процесса исключения Гаусса, но и **непосредственно**

являются элементами матрицы  $L$ , преобразующей матрицу  $U$  обратно в  $A$ . Напомним также, что эта же матрица преобразует вектор  $c$  обратно в  $b$ , т. е.  $Lc = b$ .

Мы напомним, что этого не происходит в произведении матриц  $E_{32}E_{31}E_{21}$ , вычисленном в (12). При перемножении в этом порядке происходит сокращение некоторых членов, в результате чего в позиции (3, -1) возникает  $-5$ .

**Упражнение 1.4.1.** Умножить матрицу из (12) на  $L$ , чтобы проверить, что это действительно матрица  $L^{-1}$ , обратная к  $L$ . Заметим, что только  $L$ , но не обратная к ней, может быть выписана по виду сомножителей.

Постараемся понять, почему для системы любого порядка перемножение сомножителей при вычислении матрицы  $L$  происходит без каких-либо сокращений участвующих величин. Посмотрим, каков будет порядок сомножителей в случае матрицы порядка 4:

$$L = E_{21}^{-1}E_{31}^{-1}E_{41}^{-1}E_{32}^{-1}E_{42}^{-1}E_{43}^{-1}I.$$

Единичная матрица добавлена справа для удобства. Она не нарушает равенства и при этом позволяет рассмотреть все операции в применении к ее строкам. Первая операция  $E_{43}^{-1}$  добавляет третью строку  $[0 \ 0 \ 1 \ 0]$  к четвертой строке с множителем  $l_{43}$ , что соответствует помещению этого множителя в позицию (4,3) вместо нуля. Затем операции  $E_{42}^{-1}$  и  $E_{32}^{-1}$  добавляют вторую строку  $[0 \ 1 \ 0 \ 0]$  с соответствующими множителями к четвертой и третьей строкам, что означает помещение величин  $l_{42}$  и  $l_{32}$  в позиции (4, 2) и (3, 2) без изменения  $l_{43}$ . Наконец, первый столбец матрицы  $L$  заполняется с помощью трех операций  $E_{41}^{-1}$ ,  $E_{31}^{-1}$  и  $E_{21}^{-1}$ .

$E_{43}^{-1}$ , ...,  $E_{21}^{-1}$  являются теми самыми операциями (расположенными в том же порядке), которые преобразуют систему  $Ux = c$  с нижней треугольной матрицей к первоначальной системе  $Ax = b$ . Они прибавляют кратные строки, которые вычитались, когда строилась матрица  $U$ . Поскольку общее действие этих операций эквивалентно умножению на матрицу  $L$ , то мы легко можем дать матричную формулировку метода последовательного исключения Гаусса.

**1Е.** Если все ведущие элементы отличны от нуля, то матрица  $A$  может быть представлена в виде произведения  $LU$  нижней треугольной матрицы  $L$  и верхней треугольной матрицы  $U$ . Элементы главной диагонали матрицы  $L$  равны единице, а поддиагональными элементами являются множители  $l_{ij}$ , с которыми  $j$ -я строка вычитается из  $i$ -й в процессе исключения. Матрица коэффициентов  $U$  получается по окончании процесса исключения (до начала процесса обратной подстановки); ее диагональными элементами являются ведущие элементы.

**Пример.**

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{ приводится к } U = \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix} \text{ с } L = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}.$$

**Упражнение 1.4.2.** Применяя метод исключения, вычислить сомножители  $L$  и  $U$  для матриц

$$A = \begin{bmatrix} 2 & 1 \\ 8 & 7 \end{bmatrix} \quad \text{и} \quad A = \begin{bmatrix} 1 & 0 \\ 8 & 1 \end{bmatrix}.$$

**Упражнение 1.4.3.** Построить разложение  $A = LU$  и выписать систему  $Ux = c$  с верхней треугольной матрицей  $U$  для системы

$$Ax = \begin{bmatrix} 2 & 3 & 3 \\ 0 & 5 & 7 \\ 6 & 9 & 8 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 5 \end{bmatrix}.$$

Первоначальная система уравнений  $Ax = b$  (или  $LUx = b$ ) преобразовывалась с помощью исключений к системе  $Ux = c$ . С матричной точки зрения этот процесс эквивалентен умножению системы  $Ax = b$  на матрицу  $L^{-1}$ , а значит,  $c = L^{-1}b$ . (Нужно отметить, что эту матрицу мы в явном виде не получали и в действительности при вычислениях *никогда получать не должны*. Исключение осуществлялось шаг за шагом, когда еще и речи не было о матрице  $L^{-1}$ .)

На заключительном этапе система  $Ux = c$  с верхней треугольной матрицей  $U$  решается с помощью обратной подстановки: из последнего уравнения находим  $x_n = c_n/u_{nn}$ , затем из следующего от конца уравнения находим  $x_{n-1}$  и т. д. В матричной формулировке решение можно записать любым из следующих способов:

$$x = U^{-1}c, \quad x = U^{-1}L^{-1}b, \quad x = A^{-1}b. \quad (18)$$

Опять заметим, что матрица  $U^{-1}$  не образуется явно и не должна никогда образовываться. Ее вычисление было бы просто потерей времени, поскольку, даже если мы уже имеем эту матрицу, умножение на нее вектора  $c$  требует  $n^2/2$  действий, т. е. столько же, сколько требует обратная подстановка. Такое замечание применимо почти ко всем приложениям матрицы  $A^{-1}$ ; мы хотим получить решение  $x = A^{-1}b$  рассматриваемой системы, но нам не нужны элементы матрицы  $A^{-1}$ .

Это не значит, что если нам надо будет решить систему с новой правой частью (скажем,  $b'$ ), то придется повторять весь процесс исключения. В задачах проектирования часто возникает необходимость решать системы с одной и той же матрицей  $A$  и различными правыми частями. Естественно, очень важно при их решении избежать повторения исключений, которые каждый раз требуют  $n^3/3$  действий. Вместо этого вычислительная машина просто просматривает уже имеющуюся информацию о шагах

исключения, используя ее для преобразования новой правой части  $b'$ , но не матрицы  $A$ , так как здесь в результате получается та же самая матрица  $U$ .

Суммируем изложенное. Пусть множители  $L$  и  $U$  один раз вычислены. Тогда решение  $x'$  системы с новой правой частью  $b'$  можно найти только за  $n^2$  действий:  $n^2/2$  действий в прямом направлении для вычисления вектора  $c'$  и  $n^2/2$  действий при обратной подстановке для вычисления вектора  $x'$ .

**Пример.** Матрица  $A$  нашего первоначального уравнения имела следующее  $LU$ -разложение (матрица  $U$  берется из (9), а  $L$  — из (17)):

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \end{bmatrix} = LU.$$

Предположим теперь, что матрица  $A$  та же самая, а новой является лишь правая часть  $b'$ , например:

$$\begin{aligned} 2u' + v' + w' &= 8, \\ 4u' + v' &= 11, \\ -2u' + 2v' + w' &= 3. \end{aligned} \quad (19)$$

Тогда, зная  $L$  и  $U$ , сначала решаем систему  $Lc' = b'$ :

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -3 & 1 \end{bmatrix} \begin{bmatrix} c'_1 \\ c'_2 \\ c'_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 11 \\ 3 \end{bmatrix}, \quad \text{откуда} \quad c' = \begin{bmatrix} 8 \\ -5 \\ -4 \end{bmatrix}.$$

Решение  $c'$  точно совпадает с правой частью системы, которую мы получили бы из (19), проведя процесс исключения; но учитывая, что этот процесс уже однажды был проведен и матрица  $L$  известна, мы находим вектор  $c'$  много быстрее. Далее, проводя обратную подстановку в системе  $Ux' = c'$ , которая имеет вид

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} u' \\ v' \\ w' \end{bmatrix} = \begin{bmatrix} 8 \\ -5 \\ -4 \end{bmatrix},$$

находим ее решение

$$x' = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}.$$

Таким образом, если  $LU$ -разложение уже известно, то задача сводится к решению двух систем с треугольными матрицами, которые решаются с помощью прямой и обратной подстановок.

Сделаем два дополнительных замечания относительно сомножителей  $L$  и  $U$ .

(1) В одном отношении  $LU$ -разложение «несимметрично»: на главной диагонали матрицы  $U$  расположены ведущие элементы, в то время как в  $L$  на этих местах стоят единицы. Эту «несимметричность» легко устранить, выделяя из матрицы  $U$  диагональную матрицу  $D$  с ведущими элементами  $d_1, \dots, d_n$  по диагонали:

$$U = \begin{bmatrix} d_1 & & & & \\ & d_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & d_n \end{bmatrix} \begin{bmatrix} 1 & u_{12}/d_1 & u_{13}/d_1 & \cdot \\ & 1 & u_{23}/d_2 & \cdot \\ & & & \cdot & \cdot \\ & & & & 1 \end{bmatrix}.$$

Теперь треугольное разложение матрицы  $A$  можно записать в виде  $A = LDU$ . Здесь  $L$  — нижняя треугольная матрица с единицами на главной диагонали,  $U$  — верхняя треугольная матрица с единицами на главной диагонали и  $D$  — диагональная матрица с ведущими элементами  $d_i$  на диагонали. (Использование той же самой буквы  $U$  для новой верхней треугольной матрицы общепринято, хотя иногда и вносит путаницу.) Теперь в разложении  $A = LDU$  матрицы  $L$  и  $U$  в некотором смысле симметричны. Разложение для матрицы  $A$  из нашего примера с той же самой матрицей  $L$  и только что определенной матрицей  $U$  имеет вид

$$\begin{aligned} A &= \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & & \\ & -1 & \\ & & -4 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} = LDU. \end{aligned}$$

(2) В результате нашего описания каждого шага процесса исключения могло сложиться впечатление, что в другом порядке осуществлять вычисления невозможно. Это не так. Некоторая свобода действий существует и имеется «алгоритм Кроута», который осуществляет вычисления в несколько другом порядке. Конечно, это не абсолютная свобода, так как операции, проводимые со строками, легко могут уничтожить на некотором шаге нули, которые были получены на одном из предыдущих шагов. И, кроме того, нет никакой свободы в выборе окончательных матриц  $L$ ,  $D$  и  $U$ . Вот наше основное положение:

**1F.** Если  $A=L_1D_1U_1$  и  $A=L_2D_2U_2$ , где матрицы  $L$  нижние треугольные с единицами на главной диагонали, матрицы  $U$  верхние треугольные с единицами на главной диагонали и  $D$ —диагональные матрицы с ненулевыми диагональными элементами, то  $L_1=L_2$ ,  $D_1=D_2$  и  $U_1=U_2$ . Таким образом,  $LDU$ -разложение однозначно определяется матрицей  $A$ .

**Доказательство.** Пусть  $L_1D_1U_1=L_2D_2U_2$ . Мы воспользуемся тем, что матрица  $L_1^{-1}$  обладает теми же свойствами (нижняя треугольная, единицы на главной диагонали), что и матрица  $L_1$ , более того, обе они являются произведениями элементарных матриц. Аналогичным образом, существует верхняя треугольная матрица  $U_2^{-1}$  с единицами на главной диагонали, для которой  $U_2U_2^{-1}=I$ . Ясно, что любая диагональная матрица  $D_1$  с ненулевыми диагональными элементами имеет обратную, которая также диагональна и

$$\begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix}^{-1} = \begin{bmatrix} 1/d_1 & & & \\ & 1/d_2 & & \\ & & \ddots & \\ & & & 1/d_n \end{bmatrix}.$$

Теперь, умножая исходное равенство слева на матрицы  $L_1^{-1}$  и  $D_1^{-1}$  и справа на матрицу  $U_2^{-1}$ , получаем

$$U_1U_2^{-1}=D_1^{-1}L_1^{-1}L_2D_2.$$

Левая часть этого равенства—произведение двух верхних треугольных матриц с единицами на главной диагонали и является матрицей такого же типа. С другой стороны, правая часть—нижняя треугольная матрица. Отсюда вытекает, что обе части последнего равенства будут единичными матрицами, поскольку это единственная матрица, которая одновременно является нижней треугольной и верхней треугольной и имеет единицы на главной диагонали. Таким образом,  $U_1U_2^{-1}=I$ , и умножая это равенство справа на матрицу  $U_2$ , получаем  $U_1=U_2$ .

Аналогично показывается, что  $D_1=D_2$  и  $L_1=L_2$ .

**Упражнение 1.4.4.** Предполагая, что все ведущие элементы отличны от нуля, построить  $LDU$ -разложение матрицы  $A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$  порядка 2 общего вида.

**Упражнение 1.4.5.** Вычислить сомножители  $LDU$ -разложения матрицы

$$A=\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

Каков промежуточный вектор  $c$  процесса исключения и решение системы  $Ax=b$  для

$$b = \begin{bmatrix} 6 \\ 2 \\ -6 \end{bmatrix} ?$$

**Упражнение 1.4.6.** Пусть заданы две системы уравнений с одной и той же матрицей  $A$  порядка  $n=150$ . Почему решение второй системы (после того как решена первая) займет в 50 раз меньше времени, чем решение первой?

**Упражнение 1.4.7.** Решить систему  $Ax=b$ , зная

$$L = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}.$$

Провести процесс исключения—это все равно, что решить систему  $Lc=b$ , а обратная подстановка осуществляется для системы  $Ux=c$ .

## § 1.5. ПЕРЕСТАНОВКИ СТРОК, ОБРАЩЕНИЯ И ОШИБКИ ОКРУГЛЕНИЯ

Теперь мы вплотную подошли к проблеме, которой до сих пор избегали,—возможности появления нулевого ведущего элемента. Эта проблема может возникнуть уже в самом начале процесса исключения, если верхний левый элемент  $a_{11}$  оказывается нулем и, следовательно, бесполезен в качестве ведущего элемента. Возьмем простейший пример

$$\begin{bmatrix} 0 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \quad (20)$$

Здесь трудность очевидна—никакое кратное первой строки не может быть использовано для аннулирования числа 3 в позиции (2, 1).

Выход из положения столь же ясно виден: переставить эти два уравнения, поскольку тогда число 3 перемещается в левую верхнюю позицию и становится ведущим элементом. В этом простейшем случае матрица сразу становится верхней треугольной, т. е. одновременно является матрицей  $U$  и, следовательно, нижняя треугольная матрица  $L$  является единичной матрицей. Система

$$\begin{aligned} 3u + 4v &= b_2, \\ 2v &= b_1 \end{aligned}$$

может быть непосредственно решена с помощью обратной подстановки.

Переупорядочение строк можно описать в терминах матриц, если мы найдем матрицу  $P$ , осуществляющую перестановку двух

строк. Такой *матрицей перестановки* является матрица

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

умножение на нее матрицы  $A$  переставляет строки последней:

$$PA = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 0 & 2 \end{bmatrix}.$$

Конечно, такую же операцию  $P$  осуществляет с вектором  $b$ , переставляя его компоненты  $b_1$  и  $b_2$ .

Рассмотрим более общий случай. Предположим, что после проведения нескольких шагов исключения возник нулевой ведущий элемент. (Позднее, на стр. 205, мы сможем дать точную формулу для ведущих элементов. Однако метод исключения является наилучшим способом нахождения этой формулы и выяснения, действительно ли ведущий элемент равен нулю.) Мы, например, можем предположить, что первые  $k-1$  ведущих элементов ненулевые и мы уже получили нули ниже этих ведущих элементов. Что происходит, если  $k$ -й ведущий элемент равен нулю?

*Ситуация с нулевым ведущим элементом допускает две возможности: трудность может оказаться легко устранимой или серьезной.* Либо одно из последующих уравнений можно переставить с  $k$ -м так, чтобы появился ненулевой ведущий элемент, либо ни одно из последующих уравнений не в состоянии помочь. Иначе говоря, мы просматриваем элементы  $k$ -го столбца, расположенные ниже нулевого ведущего элемента, и если некоторый из них, скажем элемент  $l$ -й строки, ненулевой, то после перестановки  $l$ -й и  $k$ -й строк процесс исключения может быть продолжен. Если же все элементы  $k$ -го столбца, начиная с ведущего элемента и ниже, равны нулю, то матрица является **вырожденной** (это утверждение относится и к первоначальной матрице  $A$ , с которой начинался процесс исключений). Исключение точно указывает эту вырожденность: если посмотреть на оставшуюся часть системы, то легко увидеть, что слева осталось слишком мало неизвестных по сравнению с числом оставшихся уравнений. Мы должны найти неизвестные с номерами от  $k$  до  $n$  из уравнений с  $k$ -го по  $n$ -е, но поскольку все элементы  $k$ -го столбца оказываются нулями, найти  $k$ -е неизвестное невозможно. Эта ситуация будет обсуждаться в следующей главе: система либо имеет бесконечно много решений, либо (что более вероятно) не имеет решений совсем.

Заметим, что если все ведущие элементы, кроме последнего, ненулевые, то мы, конечно, имеем дело с вырожденной матрицей, поскольку ниже  $n$ -го ведущего элемента смотреть уже не на что.

Примером успешного выбора ведущего элемента служит по-

следовательность (невырожденный случай)

$$\begin{bmatrix} 1 & 3 & 2 \\ 2 & 6 & 9 \\ 2 & 8 & 8 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 3 & 2 \\ 0 & 0 & 5 \\ 0 & 2 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 3 & 2 \\ 0 & 2 & 4 \\ 0 & 0 & 5 \end{bmatrix}.$$

Примером невозможности выбора ненулевого ведущего элемента является последовательность (вырожденный случай):

$$\begin{bmatrix} 1 & 3 & 2 \\ 2 & 6 & 9 \\ 3 & 9 & 8 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 3 & 2 \\ 0 & 0 & 5 \\ 0 & 0 & 2 \end{bmatrix}.$$

**Упражнение 1.5.1.** Решить, производя, когда это необходимо, выбор ведущего элемента, систему

$$\begin{aligned} u + 4v + 2w &= -2, \\ 2u - 8v + 3w &= 32, \\ v + w &= 1. \end{aligned}$$

**Упражнение 1.5.2.** Провести исключение для вырожденной системы

$$\begin{aligned} u + v &= b_1, \\ 3u + 3v &= b_2 \end{aligned}$$

и определить, для каких векторов  $b$  решение существует.

В настоящем параграфе мы остановимся только на случае невырожденной матрицы, т. е. когда ненулевой ведущий элемент может быть получен путем перестановки строк. (Конечно, позже может опять возникнуть нулевой ведущий элемент, что потребует других перестановок строк, и такая ситуация может возникать несколько раз, хотя это и маловероятно. Заметим также, что вычислительной машине нет необходимости тратить время на перестановку строк, поскольку учесть каждую перестановку можно путем изменения порядковых номеров уравнений.) Наш вопрос: *что случается с LU- (или LDU-) разложением, когда происходят перестановки строк?*

Несомненно, что результатом процесса исключения, как и раньше, будет некоторая верхняя треугольная матрица  $U$ . Это цель процесса. Только теперь этот процесс включает не только операции с матрицами  $E_{ij}$ , которые вычитают кратное одной строки из другой, но и операции с матрицами перестановки  $P_{kl}$ . Умножение справа на матрицу  $P_{kl}$  осуществляет перестановку  $k$ -й и  $l$ -й строк исходной матрицы. (Определить вид матрицы  $P_{kl}$  можно с помощью изящного приема. Умножим на  $P_{kl}$  справа единичную матрицу  $I$ . Тогда матрица  $P_{kl}I = P_{kl}$  — матрица, по-

лущающаяся из единичной перестановкой  $k$ -й и  $l$ -строк.) Например,

$$P_{24} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Очень легко вычислить обратную матрицу  $P_{kl}^{-1}$ , которая осуществляет обратную к перестановке  $P_{kl}$  перестановку. Для этого нам нужно переставить те же самые строки и, следовательно,  $P_{kl}^{-1} = P_{kl}$ .

Теперь матрица, преобразующая матрицу  $U$  в  $A$ , может быть вычислена, как и прежде: она является произведением матриц  $E_{ij}^{-1}$  и  $P_{kl}^{-1}$ , обращающих каждый отдельный шаг исключения. Однако в случае, когда возникают нулевые ведущие элементы и соответственно требуются перестановки строк, из-за участия матриц перестановки произведение уже не будет нижней треугольной матрицей  $L$ . Иначе говоря, первоначальную матрицу  $A$  нельзя представить в виде произведения нижней треугольной матрицы  $L$  и верхней треугольной матрицы  $U$  с ненулевыми элементами на главных диагоналях. Такого разложения уже не получается.

Тем не менее есть простое соображение, которое позволяет нам почти полностью вернуть разложение  $A = LU$ . Предположим, что у нас есть перечень всех перестановок строк, которые необходимо осуществить в ходе процесса исключения, и мы производим их все до начала исключения. Иначе говоря, мы заменяем исходную матрицу  $A$  матрицей  $PA$  с переставленными строками, где  $P$  — матрица перестановки, имеющая в каждой строке и каждом столбце только один ненулевой элемент, равный единице. Очевидно, что  $P$  является произведением элементарных матриц перестановки  $P_{kl}$  и объединяет все перестановки воедино.

Теперь упомянутое соображение формулируется следующим образом: процесс исключения можно осуществлять для матрицы  $PA$ , строки которой располагаются в нужном порядке заранее. При этом ведущие элементы оказываются теми же самыми, которые получались, когда перестановки проводились по мере необходимости в ходе самого процесса исключения. Другими словами, матрица  $PA$  имеет обычное  $LU$ - (или  $LDU$ -) разложение с ненулевыми ведущими элементами. Имеется лишь одно небольшое изменение, связанное с предварительным проведением всех перестановок. Предположим, что кратное первой строки вычитается из второй, а затем для получения ненулевого ведущего элемента вторая и третья строки переставляются. Тогда для случая матрицы  $PA$ , поскольку вторая и третья строки переставлены у нее заранее, кратное первой строки уже будет вычитаться из третьей строки.

**Пример.** Основная последовательность процесса исключения имеет вид

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 2 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow{E_{21}} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & -4 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow{P_{23}} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & -4 \end{bmatrix} = U.$$

Если же перестановка строк проведена заранее с  $P = P_{23}$ , то

$$PA = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 2 & 4 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & -4 \end{bmatrix} = LU.$$

Здесь удвоенная первая строка матрицы  $PA$  вычитается из ее *третьей* строки (операция  $E_{31}$  должна быть заменена операцией  $E_{31}$ ), и именно этот шаг обращается матрицей  $L$ .

В полном виде теорию исключения Гаусса, которая была главной целью настоящей главы, можно сформулировать следующим образом.

**1G.** Если матрица  $A$  невырожденная, то существует такая матрица перестановки  $P$ , что матрица  $PA$  допускает разложение  $PA = LU$  (или  $PA = LDU$ , что предпочтительнее) с ненулевыми ведущими элементами. В этом случае существует единственное решение системы  $Ax = b$ , которое можно найти с помощью процесса исключения при использовании перестановок строк. В случае вырожденной матрицы никакое переупорядочение строк не позволяет избежать нулевых ведущих элементов.

**Упражнение 1.5.3.** Найти  $LDU$ -разложение матрицы  $PA$  для

$$A = \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}.$$

**Упражнение 1.5.4.** При какой матрице перестановки  $P$  будет возможно разложение  $PA = LDU$ , если

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 5 \end{bmatrix}?$$

**Упражнение 1.5.5.** Определить, будут следующие две системы вырожденными или невырожденными и имеют ли они решение:

$$\begin{array}{ll} v - w = 2, & v - w = 0, \\ u - v = 2, & u - v = 0, \\ u - w = 2; & u - w = 0. \end{array}$$

## ОБРАЩЕНИЯ

В случае вырожденной матрицы излагаемая теория будет оставаться неполной вплоть до § 2.4. Более того, даже невырожденный случай, который мы связываем с наличием ненулевых ведущих элементов, очень нуждается в обсуждении с позиций, не зависящих от метода исключения. Но мы уже ввели обозначения  $E_{ij}^{-1}$ ,  $U^{-1}$  и  $A^{-1}$ , и нам кажется, что пора установить некоторые общие правила обращения.

Первое из них утверждает, что если для матрицы  $A$  существует «левая обратная» матрица  $B$  и «правая обратная» матрица  $C$ , то они должны совпадать.

|| **1Н.** Если  $BA=I$  и  $AC=I$ , то  $B=C$ .

Доказательство. Так как умножение матриц ассоциативно, то из равенства  $B(AC)=(BA)C$  получаем  $BI=IC$  и, следовательно,  $B=C$ .

Для прямоугольных матриц одно обращение (например, левостороннее) может существовать без другого, но для квадратных матриц это невозможно. Невырожденными матрицами  $A$  порядка  $n$  являются только такие, которые имеют двусторонние обратные:

$$AA^{-1}=A^{-1}A=I.$$

Это свойство было проверено для элементарных матриц  $E_{ij}$  и обратных к ним, после чего матрица  $L^{-1}$  была построена с помощью правила обращения произведений:

|| **11.** Если матрицы  $A$  и  $B$  обратимы, то матрица  $AB$  также обратима и обратная к ней находится как произведение матриц  $A^{-1}$  и  $B^{-1}$ , взятых в обратном порядке:

$$(AB)^{-1}=B^{-1}A^{-1}. \quad (21)$$

Доказательство. Воспользовавшись опять законом ассоциативности, получим

$$\begin{aligned} (AB)(B^{-1}A^{-1}) &= ABB^{-1}A^{-1} = AIA^{-1} = AA^{-1} = I, \\ (B^{-1}A^{-1})(AB) &= B^{-1}A^{-1}AB = B^{-1}IB = B^{-1}B = I. \end{aligned}$$

Распространяя установленное правило на произведение трех матриц, легко видеть, что обратная к матрице  $A=LDU$  матрица будет тройным произведением  $A^{-1}=U^{-1}D^{-1}L^{-1}$ . Как мы уже отмечали ранее, в действительности вычисление такой обратной матрицы требуется очень редко: если надо найти только вектор  $A^{-1}b$ , то вычислять отдельно  $A^{-1}$ —это лишняя трата времени. Но все же интересно знать, сколько для этого требуется времени и как это можно сделать. Хороший способ—это расписать по

столбцам равенство  $AA^{-1} = I$ : если обозначить  $j$ -й столбец матрицы  $A^{-1}$  через  $x_j$ , а  $j$ -й столбец единичной матрицы обозначить через  $e_j$ , то получим  $Ax_j = e_j$ . Это приводит к  $n$  системам уравнений с  $n$  различными правыми частями  $e_1, \dots, e_n$ . Поскольку здесь матрица коэффициентов  $A$  одна и та же для всех систем, то процесс исключения для матрицы  $A$  проводится только один раз. Практически же самой простой схемой является **метод Гаусса — Жордана**, когда исключение проводится для всех  $n$  систем одновременно.

**Пример**

$$\begin{aligned} \left[ \begin{array}{c|ccc} A & e_1 & e_2 & e_3 \end{array} \right] &= \left[ \begin{array}{cccccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 4 & 1 & 0 & 0 & 1 & 0 \\ -2 & 2 & 1 & 0 & 0 & 1 \end{array} \right] \rightarrow \left[ \begin{array}{cccccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & -2 & -2 & 1 & 0 \\ 0 & 3 & 2 & 1 & 0 & 1 \end{array} \right] \rightarrow \\ &\rightarrow \left[ \begin{array}{cccccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & -2 & -2 & 1 & 0 \\ 0 & 0 & -4 & -5 & 3 & 1 \end{array} \right]. \end{aligned}$$

Это завершает процесс исключения. В последней матрице первые три столбца задают известную нам матрицу  $U$ . Другие три столбца являются тремя правыми частями систем после их подготовки к процедуре обратной подстановки и в то же время являются столбцами матрицы  $L^{-1}$ , о которой мы много раз говорили. Иначе говоря, эти столбцы являются результатом применения элементарных операций  $E_{21}, E_{31}, E_{32}$  (или, что эквивалентно, матрицы  $L^{-1} = E_{32}E_{31}E_{21}$ ) к единичной матрице. Полученные после реализации для всех трех систем обратной подстановки решения  $x_1, x_2, x_3$  и будут являться искомыми столбцами матрицы  $A^{-1}$ . Метод Гаусса — Жордана находит эти столбцы одновременно, как бы продолжая процесс исключения с той единственной разницей, что теперь нужно вычитать кратные строки из строк, *расположенных выше*. Начинаем мы с получения нуля выше ведущего элемента во втором столбце:

$$\begin{aligned} [A \ I] &\rightarrow \left[ \begin{array}{cccccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & -2 & -2 & 1 & 0 \\ 0 & 0 & -4 & -5 & 3 & 1 \end{array} \right] \rightarrow \left[ \begin{array}{cccccc} 2 & 0 & -1 & -1 & -1 & 0 \\ 0 & -1 & -2 & -2 & 1 & 0 \\ 0 & 0 & -4 & -5 & 3 & 1 \end{array} \right] \rightarrow \\ &\rightarrow \left[ \begin{array}{cccccc} 2 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ 0 & -1 & 0 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & -4 & -5 & 3 & 1 \end{array} \right] \rightarrow \left[ \begin{array}{cccccc} 1 & 0 & 0 & \frac{1}{8} & \frac{1}{8} & -\frac{1}{8} \\ 0 & 1 & 0 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & \frac{5}{4} & -\frac{3}{4} & -\frac{1}{4} \end{array} \right] = \\ &= [I \ A^{-1}]. \end{aligned}$$

На последнем шаге мы делим на ведущие элементы. Теперь матрица коэффициентов в левой половине единичная, а три решаемые системы приведены к виду

$$I x_1 = \begin{bmatrix} \frac{1}{8} \\ -\frac{1}{2} \\ \frac{5}{4} \end{bmatrix}, \quad I x_2 = \begin{bmatrix} \frac{1}{8} \\ \frac{1}{2} \\ -\frac{3}{4} \end{bmatrix}, \quad I x_3 = \begin{bmatrix} -\frac{1}{8} \\ \frac{1}{2} \\ -\frac{1}{4} \end{bmatrix}. \quad (22)$$

Это и есть столбцы матрицы  $A^{-1}$ . Таким образом, метод Гаусса — Жордана осуществляет абсолютное исключение для матрицы  $A$ , давая нули как выше ведущих элементов, так и ниже них и одновременно применяя те же самые операции к столбцам единичной матрицы. При этом матрица  $A$  в левой половине нашей матрицы размера  $3 \times 6$  переходит в матрицу  $I$ , в то время как матрица  $I$  в ее правой половине преобразуется в матрицу  $A^{-1}$ .

*Замечание.* Любопытно подсчитать число действий, необходимых для нахождения матрицы  $A^{-1}$ . Обычный подсчет для каждой новой правой части дает величину  $n^2$  — половина для действий в прямом направлении и половина для обратной подстановки. Для  $n$  различных правых частей это дает величину  $n^3$ , и, добавляя к ней  $n^3/3$  действий, требуемых для преобразования самой матрицы  $A$ , получаем общее количество действий  $4n^3/3$ .

В действительности эта оценка несколько завышена, поскольку правые части  $e_j$  имеют специальный вид. В первой половине наших вычислений (исключении) в векторе  $e_j$  нужно изменять только те компоненты, которые расположены ниже единицы, стоящей на  $j$ -м месте. Так, для изменения  $e_1$  требуется  $n^2/2$  действий, а для изменения  $e_j$  достаточно только  $(n-j)^2/2$  действий. (В процессе исключения с вектором  $e_j$  ничего не происходит до момента, пока мы не подойдем к  $j$ -му ведущему элементу. Но тогда оставшаяся часть содержит только  $n-j$  компонент.) Иначе говоря, процесс преобразования матрицы  $I$  в  $L^{-1}$  оперирует только с нижними треугольными матрицами, сохраняя нули над диагональю. Поэтому первая половина вычислений требует

$$\frac{n^2}{2} + \dots + \frac{(n-j)^2}{2} + \dots + \frac{1}{2} \approx \frac{n^3}{6} \quad (23)$$

действий. К этой величине нужно прибавить  $n^3/3$  действий, требуемых для проведения один раз процесса исключения для матрицы  $A$ , и  $n(n^2/2)$  действий, которые требуются на последнем этапе для нахождения  $x_j$  обратными подстановками независимо от того, осуществляются ли они отдельно для каждой правой части или одновременно, как в методе Гаусса — Жордана. Таким обра-

зом, общее число действий, требуемых для вычисления матрицы  $A^{-1}$ , равно

$$\frac{n^3}{6} + \frac{n^3}{3} + n \left( \frac{n^2}{2} \right) = n^3.$$

Эта величина необыкновенно мала. Действительно, поскольку перемножение матриц уже требует  $n^3$  действий, то для вычисления матрицы  $A^2$  требуется столько же действий, сколько для нахождения матрицы  $A^{-1}$ . Этот факт кажется почти невероятным. (Вычисление  $A^3$  требует в два раза больше действий, чем  $A^{-1}$ .) Тем не менее, если в вычислении явного вида матрицы  $A^{-1}$  нет необходимости, делать этого не надо.

**Упражнение 1.5.6.** Найти обратную к матрице  $AB^{-1}C$ .

**Упражнение 1.5.7.** Найти обратные к матрицам

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{и} \quad B = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

**Упражнение 1.5.8.** Показать, что матрица  $\begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix}$  не имеет обратной, пытаясь решить систему

$$\begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

**Упражнение 1.5.9.** Описать, как действует на произвольную матрицу  $A$  умножение на матрицу  $E$ , и выписать матрицу  $E^{-1}$ , если

$$E = \begin{bmatrix} 1 & 0 & 8 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

**Упражнение 1.5.10.** Используя метод Гаусса—Жордана одновременного решения систем  $Ax_j = e_j$ , найти обратные к матрицам

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{и} \quad A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

### ОШИБКИ ОКРУГЛЕНИЯ

В теоретической части рассмотрение невырожденного случая завершено. Перестановки строк могут оказаться необходимыми, чтобы избежать нулевых ведущих элементов, и они позволяют нам добиться этого, если матрица невырожденная. Тогда на заключительном этапе решение системы  $Ax = b$  можно найти с помощью обратной подстановки. На практике, однако, случается, что столь же необходимыми оказываются другие перестановки строк, иначе вычисленное решение может оказаться бессмысленным.

Напомним, что для системы средних размеров, скажем порядка ста, для реализации исключения требуется около миллиона действий и при выполнении каждого действия мы должны ожидать ошибку округления. При обычных вычислениях с плавающей точкой (например, в десятичной системе счисления) каждое число  $n$  описывается мантиссой  $m$  и показателем степени  $c$ , т. е.  $n = m10^c$  с  $\frac{1}{10} \leq |m| < 1$ . Число цифр в  $m$  равно длине машинного слова. Без какого-либо анализа ошибок для вычислений с плавающей точкой мы можем указать на очевидный факт: если в арифметике с плавающей точкой складываются два числа и их экспоненты  $c$  различаются, например, на два, то последние две цифры в меньшем числе будут потеряны:

$$0.345 + 0.00123 \rightarrow 0.346.$$

Основной вопрос здесь следующий: какой вклад вносят все эти отдельные ошибки округления в конечное решение?

Это нелегкая задача. Ее решение было начато Джоном фон Нейманом. Он был одним из ведущих математиков в то время, когда благодаря вычислительным машинам стало возможным осуществлять миллионы операций. В действительности объединение усилий Гаусса и фон Неймана придало историческую известность простому алгоритму исключения, хотя даже фон Нейман дал лишь очень сложную оценку ошибок округления; правильный путь решения вопроса был найден Уилкинсоном, книги которого стали теперь классическими.

Два простых примера, взятых из книг Нобла и Форсайта — Молера, будут иллюстрировать три самых важных момента в оценке ошибок округления. Это примеры систем с матрицами

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix} \quad \text{и} \quad A' = \begin{bmatrix} 0.0001 & 1 \\ 1 & 1 \end{bmatrix}.$$

Первый момент:

11. *Одни матрицы чрезвычайно чувствительны к малым изменениям, а другие нет. Матрица  $A$  плохо обусловлена (т. е. чувствительна к изменениям), а матрица  $A'$  хорошо обусловлена.*

Если мы очень незначительно изменим матрицу  $A'$  или вектор  $b$  в правой части, то решение системы  $A'x = b$  также изменится незначительно. С другой стороны, рассмотрим две системы с матрицей  $A$  и очень близкими правыми частями:

$$\begin{aligned} u + v &= 2, & u + v &= 2, \\ u + 1.0001v &= 2.0001; & u + 1.0001v &= 2.0002. \end{aligned}$$

Решением первой системы является вектор с компонентами  $u = 1$  и  $v = 1$ , а решением второй — вектор с компонентами  $u = 0$  и  $v = 2$ . Изменение пятой значащей цифры в компоненте вектора  $b$  привело к изменению первой значащей цифры в компонентах решения системы, и не существует численного метода, который мог бы устранить эту чувствительность к малым возмущениям. Плохая обусловленность может быть при вычислениях перенесена из одного места в другое, но не может быть устранена: истинное решение очень чувствительно, и вычисленное решение не может обладать меньшей чувствительностью.

Второй момент:

|| **1К.** *Даже хорошо обусловленная матрица может быть испорчена плохим алгоритмом.*

К сожалению, нужно сказать, что даже для матрицы  $A'$  прямой ход исключения Гаусса относится к плохим алгоритмам. Предположим, что величина 0.0001 является первым ведущим элементом, и первую строку нужно умножить на 10000 и вычесть ее из второй строки. Нижний правый элемент становится —9999, но за счет округления до трех значащих цифр он берется равным —10000. Итак, округление приводит к исчезновению единицы, которая была здесь первоначально.

Рассмотрим конкретный пример

$$\begin{aligned} 0.0001u + v &= 1, \\ u + v &= 2. \end{aligned} \tag{24}$$

После исключения второе уравнение имеет вид

$$-9999v = -9998,$$

откуда  $v = 0.99990$ . В действительности, округления приводят к уравнению  $-10000v = -10000$ , откуда  $v = 1$ . До сих пор изменение второго уравнения не привело к плохому решению: три первые значащие цифры приближенного решения  $v$  совпадают с точными. Однако при обратной подстановке первое уравнение с истинным значением  $v$  имеет вид

$$0.0001u + 0.9999 = 1,$$

откуда получаем  $u = 1$ . С другой стороны, если вместо  $v$  подставить вычисленное с тремя точными знаками решение  $v = 1$ , то получим уравнение

$$0.0001u + 1 = 1$$

и, следовательно, вычисляемое значение  $u = 0$ . Вычисленное значение  $u$  абсолютно ошибочно. Отсюда вытекает, что, несмотря на

хорошую обусловленность матрицы  $A'$ , процесс исключения очень неустойчив. Действительно, множители  $L$ ,  $D$  и  $U$  (как точные, так и приближенные) имеют совершенно другие масштабы, чем исходная матрица:

$$A' = \begin{bmatrix} 1 & 0 \\ 10000 & 1 \end{bmatrix} \begin{bmatrix} 0.0001 & 0 \\ 0 & -9999 \end{bmatrix} \begin{bmatrix} 1 & 10000 \\ 0 & 1 \end{bmatrix}.$$

Таким образом, маленький ведущий элемент 0.0001 привел к неустойчивости.

Качественно, матрица  $A$  является почти вырожденной, а  $A'$  нет. Если мы заменим последний элемент в  $A$  единицей или сделаем еще где-нибудь соответствующее маленькое изменение элементов этой матрицы, то она станет вырожденной. Не так обстоит дело с матрицей  $A'$ . Близость к вырожденности — это почти то же, что и плохая обусловленность. Различие между ними мы попытаемся разъяснить во втором параграфе гл. 7. Но причины, почему хорошая обусловленность  $A'$  была испорчена алгоритмом исключения, и средство против этого видны уже теперь. Это третий момент, который мы хотели отметить.

**11.** *Так же как нулевой ведущий элемент требует теоретического изменения алгоритма исключения, очень маленький ведущий элемент требует его практического изменения. А именно, если не известно заранее, что обычный алгоритм исключения работает хорошо, вычислительная машина должна сравнить каждый ведущий элемент со всеми другими возможными ведущими элементами этого столбца. Затем она должна переставить строки матрицы так, чтобы наибольший из них по модулю стал новым ведущим элементом. Такой алгоритм называется **частичным выбором ведущего элемента**.*

В матрице  $A'$  один из возможных ведущих элементов 0.0001 надо сравнить с элементом ниже него, который равен единице, что сразу приводит к перестановке строк. В матричных обозначениях это, как и ранее, означает умножение на матрицу перестановки. Теперь разложение новой матрицы  $A'' = PA'$  имеет вид

$$A'' = \begin{bmatrix} 1 & 1 \\ 0.0001 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0.0001 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.9999 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Ведущими элементами здесь являются величины одного масштаба: единица и 0.9999, в то время как раньше вместо них были величины 0.0001 и  $-9999$ .

Частичный выбор ведущего элемента отличен от гораздо более надежной стратегии *полного выбора ведущего элемента*, когда для нахождения возможно большего ведущего элемента просматривается как  $k$ -й столбец, так и все последующие столбцы. При

полном выборе ведущего элемента для его перемещения в нужную позицию необходимо осуществлять перестановку не только строк, но и столбцов. (Иначе говоря, дополнительно нужна перенумерация неизвестных, или, что эквивалентно, умножение анализируемой матрицы справа на матрицу перестановок.) За такую надежность приходится дорого платить, поскольку просмотр и сравнение элементов всех оставшихся столбцов требует чрезвычайно больших затрат времени. В то же время частичного выбора ведущего элемента, как правило, бывает достаточно.

Сейчас мы окончательно подошли к основному алгоритму численной линейной алгебры — **исключению с частичным выбором ведущего элемента**. Конечно, для дальнейшего совершенствования этого алгоритма еще остается целый ряд возможностей, таких, как масштабирование строк и столбцов матрицы, если это оказывается необходимым. Но по существу читатель уже знает, как вычислительная машина поступает с системой линейных уравнений. В отличие от «теоретического» описания задачи — *найти матрицу  $A^{-1}$  и умножить на нее вектор  $b$*  — наше изложение потребовало от читателя времени (и терпения). Было бы хорошо, если бы существовал более легкий путь, как в действительности находить вектор  $x$ , но думаю, что его не существует.

**Упражнение 1.5.11** (рекомендуемое). Вычислить обратную к матрице Гильберта

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}$$

двумя способами, используя обычный метод исключения Гаусса — Жордана: (i) с помощью точных вычислений и (ii) округляя каждое число до трех значащих цифр. *Замечание:* в данном случае выбор ведущего элемента не помогает, поскольку матрица  $A$  является плохо обусловленной и это никак нельзя исправить.

**Упражнение 1.5.12.** Сравнить ведущие элементы обычного исключения и исключения с частичным выбором ведущего элемента для матрицы

$$A = \begin{bmatrix} 0.001 & 0 \\ 1 & 10000 \end{bmatrix}.$$

(Это пример ситуации, когда перед исключением необходимо масштабирование.)

**Упражнение 1.5.13.** Объяснить, почему при частичном выборе ведущего элемента все элементы  $l_{ij}$  матрицы  $L$  удовлетворяют неравенству  $|l_{ij}| \leq 1$ . Показать, что если элементы исходной матрицы  $A$  удовлетворяют неравенствам  $|a_{ij}| \leq 1$ , то после первого этапа исключения (получения нулей в первом столбце) все элементы будут ограничены по модулю величиной 2, а после  $k$ -го этапа — величиной  $2^k$ . Можно ли построить пример такой матрицы  $A$  порядка три с элементами  $|a_{ij}| \leq 1$ , для которой  $|l_{ij}| \leq 1$ , а последний ведущий элемент равен четырем?

## § 1.6. ЛЕНТОЧНЫЕ МАТРИЦЫ, СИММЕТРИЧЕСКИЕ МАТРИЦЫ И ИХ ПРИМЕНЕНИЯ

В настоящем параграфе мы преследуем две цели. Первая из них — показать на конкретной задаче, каким образом большие системы линейных уравнений возникают на практике. До сих пор в книге не было каких-либо упоминаний о приложениях, так как описание большой и вполне реальной задачи из строительной техники или экономики увело бы нас далеко от основного направления. Но здесь мы выберем другую естественную и важную прикладную задачу, рассмотрение которой не потребует большой подготовки.

Другая цель — проиллюстрировать (на примере той же самой задачи) специальные свойства, которыми часто обладают матрицы коэффициентов. Редко можно встретить матрицу, которая выглядела бы так, как будто она построена случайным образом. Почти всегда матрица имеет некоторые особенности структуры, заметные, как правило, с первого взгляда. Такими особенностями, например, являются симметрия или большое число нулевых элементов. В последнем случае, так как разреженная матрица содержит значительно меньше чем  $n^2$  единиц информации, то вычислительной работы также оказывается много меньше, чем для плотной матрицы. Мы, в частности, будем рассматривать специальные свойства *ленточных матриц*, все ненулевые элементы которых сосредоточены около главной диагонали, и изучать, как эти свойства отражаются на процессе исключения.

Наш пример возникает при преобразовании непрерывной задачи в дискретную. Непрерывная задача имеет бесконечно много неизвестных и ее нельзя решить точно на вычислительной машине. Следовательно, ее нужно приближенно заменить некоторой дискретной задачей. При этом, как правило, оказывается, что чем больше неизвестных останется в задаче, тем выше будет ее точность и больше стоимость решения. В качестве весьма простой, но в то же время очень типичной непрерывной задачи, мы выберем дифференциальное уравнение

$$-\frac{d^2u}{dx^2} = f(x), \quad 0 \leq x \leq 1. \quad (25)$$

Это линейное уравнение относительно неизвестной функции  $u$  со свободным членом  $f$ . Левая часть уравнения допускает неопределенность, поскольку к любому ее решению  $u$  можно добавить произвольную функцию вида  $C + Dx$  и результат снова будет решением этой задачи (вторая производная функции  $C + Dx$  тождественно равна нулю). Неопределенность, связанная с произвольностью констант  $C$  и  $D$ , устраняется заданием на каждом из концов

интервала «краевых условий»:

$$u(0) = 0, \quad u(1) = 0. \quad (26)$$

В результате мы приходим к так называемой *двухточечной краевой задаче*, которая описывает не кратковременное, а стационарное явление, например распределение температуры стержня с фиксированными нулевыми значениями на концах и распределенными источниками тепла интенсивности  $f(x)$ .

Напомним, что наша цель — получить задачу, которая дискретна (конечномерна), или, другими словами, является задачей линейной алгебры. По этим соображениям мы не можем допустить более чем конечное количество информации о функции  $f$  и возьмем лишь ее значения в равномерно расположенных точках  $x = h, x = 2h, \dots, x = nh$ . Нашей задачей будет вычисление приближенных значений  $u_1, \dots, u_n$  точного решения  $u$  в перечисленных точках при условии, что на концах отрезка  $x = 0$  и  $x = 1 = (n + 1)h$  уже заданы точные краевые условия  $u_0 = 0$  и  $u_{n+1} = 0$ .

Прежде всего возникает вопрос, как заменить производную  $d^2u/dx^2$ ? Поскольку каждая производная является пределом разностных отношений, то ее можно аппроксимировать одним из этих отношений при фиксированном значении шага  $h$  (или  $\Delta x$ ). Для первой производной имеется три возможности:

$$\begin{aligned} \frac{du}{dx} &\approx \frac{u(x+h) - u(x)}{h}, & \frac{du}{dx} &\approx \frac{u(x) - u(x-h)}{h}, \\ \frac{du}{dx} &\approx \frac{u(x+h) - u(x-h)}{2h}. \end{aligned} \quad (27)$$

Оказывается, что последнее отношение, в силу своей симметрии относительно  $x$ , является самым лучшим. Для второй же производной существует единственная возможная комбинация, которая использует значения только в точках  $x$  и  $x \pm h$ :

$$\frac{d^2u}{dx^2} \approx \frac{u(x+h) - 2u(x) + u(x-h)}{h^2}. \quad (28)$$

Эта аппроксимация также обладает тем преимуществом, что она симметрична относительно величины  $x$ . Напомним, что правая часть (28) стремится к точному значению  $d^2u/dx^2$  при  $h \rightarrow 0$ , но мы останавливаемся на некотором конкретном значении  $h$ .

Заменим в стандартных узлах  $x = jh$  сетки дифференциальное уравнение  $-d^2u/dx^2 = f(x)$  его дискретным аналогом (28) и умножим его на  $h^2$ :

$$-u_{j+1} + 2u_j - u_{j-1} = h^2 f(jh). \quad (29)$$

В результате получим  $n$  уравнений такого вида, по одному для каждого из значений  $j = 1, \dots, n$ . Первое и последнее из этих уравнений включают величины  $u_0$  и  $u_{n+1}$ , которые не являются

неизвестными, так как их значения определены краевыми условиями. Они переносятся в правые части уравнений и добавляются к свободным членам (в тех случаях, когда не известно заранее, что они равны нулю). Равенство (29) легко интерпретировать как условие стационарного режима, при котором притоки  $(u_j - u_{j+1})$  справа и  $(u_j - u_{j-1})$  слева уравновешиваются стоком  $h^2 f(jh)$  в центре.

Структура  $n$  уравнений (29) становится виднее, если выписать матричное уравнение  $Au = b$  для конкретного значения  $h = 1/6$  ( $n = 5$ ):

$$\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} = h^2 \begin{bmatrix} f(h) \\ f(2h) \\ f(3h) \\ f(4h) \\ f(5h) \end{bmatrix}. \quad (30)$$

С этого момента мы будем иметь дело с уравнением (30), и для нас уже несущественно, откуда оно возникло. Какое значение имеет тот факт, что мы построили класс матриц коэффициентов, далеко не произвольного вида, порядок  $n$  которых может быть очень большим? Матрица  $A$  обладает рядом специальных свойств, одно из которых представляет для нас наибольший интерес: *она является трехдиагональной*. Все ее ненулевые элементы лежат на главной диагонали и на двух прилегающих к ней диагоналях. Иначе говоря, элементы матрицы равны нулю, если разность между номерами их строк и столбцов по модулю больше единицы, т. е.  $a_{ij} = 0$  для всех  $i, j$ , таких, что  $|i - j| > 1$ . Наличие этих нулей вносит огромное упрощение в процесс исключения Гаусса.

Матрица  $A$  обладает еще двумя свойствами, которые мы не решаемся обсуждать здесь, но не потому, что они не важны, а именно из-за того, что они слишком важны. Тем не менее мы можем сказать, что это за свойства и какое влияние они оказывают на алгоритм исключения.

(1) Матрица  $A$  *симметрическая*. Это означает, что каждый элемент  $a_{ij}$  с одной стороны от главной диагонали равен своему «зеркальному отражению»  $a_{ji}$  с другой стороны от главной диагонали, т. е.  $a_{ij} = a_{ji}$  для всех значений  $i, j$ . Другими словами, первый столбец матрицы  $A$  совпадает с ее первой строкой, второй столбец — со второй строкой и т. д. Одно следствие этого свойства очевидно: в вычислительной машине достаточно хранить только нижнюю треугольную часть  $A$ , поскольку ее элементы, стоящие над главной диагональю, можно сразу восстановить по соответствующим элементам, расположенным ниже главной диагонали.

Самым важным, однако, является то обстоятельство, что процесс исключения сохраняет эту симметрию (см. упражнения). Следовательно, *при исключении отсутствует необходимость в вычислении изменений, происходящих над главной диагональю*. Иначе говоря, если мы сохраняем только результаты, получающиеся ниже главной диагонали и, конечно, ведущие элементы, то верхняя треугольная часть восстанавливается по ним автоматически. Таким образом, симметричность матрицы позволяет фактически вдвое уменьшить как требуемую память, так и время, затрачиваемое на вычисления.

Мы обратим внимание на другой результат подобного сорта, для чего впервые введем индекс  $T$  для обозначения операции *транспонирования* матрицы. Пусть  $A$  — некоторая матрица. Тогда замена всех строк матрицы  $A$  на столбцы с теми же номерами дает транспонированную к ней матрицу  $A^T$ . Симметрия матрицы означает, что  $A^T = A$ , а следствием этого свойства является такое утверждение:

**1М.** Если матрица  $A$  симметрическая и для нее осуществимо разложение  $A = LDU$ , то столбцы нижней треугольной матрицы  $L$  одновременно являются строками матрицы  $U$  и наоборот. Иначе говоря, матрица  $L$  является зеркальным отражением (или транспонированной к)  $U$ :  $L = U^T$  и  $U = L^T$ .

**Доказательство.** Начнем с разложения  $A = LDU$  и постараемся выяснить, что произойдет, если мы поменяем местами строки со столбцами. В левой части останется матрица  $A$ , поскольку предполагается, что она симметрическая. В правой части процесс транспонирования приведет к перемножению отдельно транспонированных матриц  $L$ ,  $D$  и  $U$ , но в другом порядке (аналогично обращению). Иначе говоря, мы приходим к формуле  $A^T = U^T D^T L^T$ . Первый множитель  $U^T$  формируется путем замены строк матрицы  $U$  на столбцы и, следовательно, является нижней треугольной матрицей. Другие сомножители  $D^T$  и  $L^T$  соответственно являются диагональной и верхней треугольной матрицами. Таким образом, мы видим, что получено новое « $LDU$ -разложение» матрицы  $A^T = A$ . Но сомножители такого разложения любой матрицы определяются однозначно (см. 1F) и, следовательно,  $L = U^T$ ,  $D = D^T$  и  $U = L^T$ , что и завершает доказательство.

Мы еще вернемся к операции транспонирования в § 3.1, но это доказательство было очень неожиданным введением к очень простой идее.

Перейдем ко второму свойству нашей конечно-разностной матрицы.

(2) Матрица  $A$  положительно определена. Это свойство определяется (только для симметрических матриц) в гл. 6 и данное там определение не имеет ничего общего с методом исключения. Но позже мы покажем, что оно эквивалентно другому определению, которое тесно связано с процессом исключения, а именно: *симметрическая матрица  $A$  будет положительно определенной тогда и только тогда, когда все ее ведущие элементы  $d_i$  положительны*. Это означает, что нам не потребуется осуществлять перестановки строк, чтобы избежать нулевого ведущего элемента, и исключение можно проводить в обычном порядке без нарушения симметричности, которое в противном случае могло произойти.

Мы знаем, что на практике нужно задавать еще один вопрос относительно ведущих элементов. Они ненулевые, но будут ли они достаточно большими? Ответ на него является другим следствием положительной определенности, с вычислительной точки зрения более важным: для практических вычислений с положительно определенными матрицами перестановки строк не требуются. Множители  $L$ ,  $D$  и  $U$  не могут иметь больших расхождений с матрицей  $A$  в масштабах. Мы сможем в этом убедиться на примере разложения нашей матрицы.

*Замечание.* Сказанное не означает, что здесь нет опасности ошибок округления. Плохо обусловленная матрица

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix}$$

из предыдущего раздела была симметрической и положительно определенной. Такой же была и матрица Гильберта из упражнений. Положительная определенность означает только то, что исключения Гаусса без изменений строк численно так же устойчивы, как и сама матрица.

Оставляя в стороне перечисленные основные свойства матрицы  $A$ , вернемся к тому факту, что она является трехдиагональной. Какой эффект дает это в процессе исключения? Предположим, что мы провели первый этап процесса исключения и получили нули под первым ведущим элементом:

$$\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & -1 & & & \\ 0 & \frac{3}{2} & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}$$

По сравнению с матрицей общего вида порядка пять здесь имеется два основных упрощения.

(а) Ниже ведущего элемента был *только один ненулевой элемент*, а в общем случае кратное первой строки нужно вычитать из каждой другой строки;

(б) Эта одна операция исключения была выполнена с очень короткими строками. После того как нужный множитель  $l_{21} = -1/2$  был найден, потребовалось *только одно действие умножение-вычитание* справа от ведущего элемента.

Таким образом, первый этап исключений был очень упрощен наличием нулей в первой строке и первом столбце. Более того, *трехдиагональная структура матрицы сохраняется в течение процесса исключений* (при отсутствии перестановки строк!).

(с) На втором этапе исключения, так же как и на любом последующем, упрощения (а) и (б) сохраняются.

Мы можем представить окончательный результат несколькими способами. Наибольшая наглядность достигается рассмотрением *LDU-разложения* матрицы  $A$ :

$$A = \begin{bmatrix} 1 & & & & & \\ -\frac{1}{2} & 1 & & & & \\ & -\frac{2}{3} & 1 & & & \\ & & -\frac{3}{4} & 1 & & \\ & & & -\frac{4}{5} & 1 & \\ & & & & -\frac{5}{6} & 1 \end{bmatrix} \begin{bmatrix} \frac{2}{1} & & & & & \\ & \frac{3}{2} & & & & \\ & & \frac{4}{3} & & & \\ & & & \frac{5}{4} & & \\ & & & & \frac{6}{5} & \\ & & & & & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & & & & \\ & 1 & -\frac{2}{3} & & & \\ & & 1 & -\frac{3}{4} & & \\ & & & 1 & -\frac{4}{5} & \\ & & & & 1 & -\frac{5}{6} \\ & & & & & 1 \end{bmatrix}.$$

Замечания (а)—(с) можно сформулировать следующим образом: *сомножители  $L$  и  $U$  разложения трехдиагональной матрицы являются двухдиагональными матрицами*. Эти сомножители имеют близкую к  $A$  структуру расположения нулевых элементов. Заметим также, что  $L$  и  $U$  транспонированы друг к другу, как и должно было быть в силу симметричности, и ведущие элементы  $d_i$  все положительны<sup>1)</sup>. Очевидно, что при стремлении  $n$  к бесконечности ведущие элементы сходятся к  $+1$ . Такие матрицы доставляют вычислительной машине большое удовольствие.

Перечисленные упрощения приводят к полному пересмотру обычного подсчета числа действий. Каждый шаг исключения требует только двух действий — деления на ведущий элемент и умножения-вычитания из диагонального элемента. Таких шагов исключения около  $n$ . Следовательно, *вместо  $n^3/3$  действий требуется только  $2n$* , т. е. объем вычислений является величиной меньшего

<sup>1)</sup> Позже будет показано, что произведение ведущих элементов совпадает с *определителем* матрицы  $A$ . В данном случае  $\det A = 6$ .

порядка. То же самое справедливо для обратной подстановки — вместо  $n^2/2$  действий опять требуется только  $2n$ . Таким образом, число действий для системы с трехдиагональной матрицей пропорционально  $n$ , но не более высокой степени  $n$ . Системы  $Ax=b$  с трехдиагональными матрицами решаются почти мгновенно.

Рассмотрим более общий случай, когда  $A$  — ленточная матрица, т. е. все ее ненулевые элементы лежат внутри ленты, располагающейся параллельно главной диагонали. *Ширина* ленты или, точнее говоря, *половина ширины ленты*,  $\omega$  определяется следующим образом. Это такое наименьшее из целых чисел, что  $a_{ij} = 0$  для всех  $|i-j| \geq \omega$  (см. рис. 1.2). Отсюда следует, что ленточ-

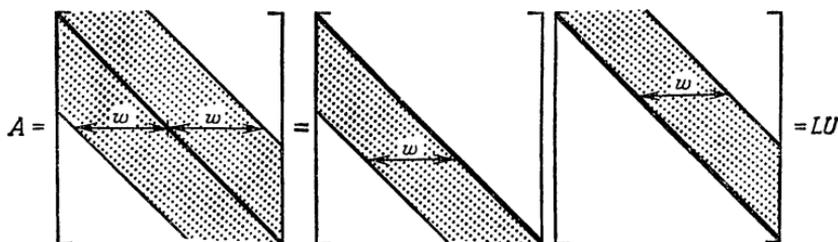


Рис. 1.2. Ленточная матрица и ее разложение.

ная матрица имеет не более  $2\omega - 1$  ненулевых диагоналей — главную диагональ и по  $\omega - 1$  дополнительных диагоналей по каждую сторону от нее. Ширина ленты  $\omega = 1$  для диагональной матрицы,  $\omega = 2$  для трехдиагональной и  $\omega = n$  для плотной матрицы.

Для любой ленточной матрицы справедливы замечания (а) — (с). На первом этапе исключения требуется получить только  $\omega - 1$  последовательных нулей ниже ведущего элемента, поскольку все элементы столбца, располагающиеся ниже, уже равны нулю. Более того, так как первая строка имеет не более  $\omega$  ненулевых элементов, то для получения каждого нуля достаточно  $\omega$  действий. Следовательно, первый этап исключения требует  $\omega(\omega - 1)$  действий, причем после его реализации мы опять имеем ленточную матрицу с шириной ленты  $\omega$ . Так как всего осуществляется около  $n$  этапов, то применение исключений для ленточной матрицы требует около  $\omega^2 n$  действий. Опять число действий оказывается пропорциональным величине  $n$  и, как мы видим, одновременно пропорционально квадрату величины  $\omega$ . Так как при стремлении  $\omega$  к  $n$  матрица становится плотной, то и число действий опять становится пропорциональным  $n^3$ . Более точный подсчет связан с тем обстоятельством, что, когда исключения достигают нижнего правого угла матрицы  $A$ , ширина ленты становится меньше. Точное число делений и умножений-вычитаний, требуемых для построения  $L$ ,  $D$  и  $U$  (без предположения симметричности  $A$ ),

вычисляется по формуле  $P = \frac{1}{3} \omega(\omega - 1)(3n - 2\omega + 1)$ . Для плотной матрицы, когда  $\omega = n$ , она принимает вид  $P = \frac{1}{3} n(n-1)(n+1)$ , т. е. согласуется с уже известной формулой<sup>1)</sup>. Резюмируем изложенное: ненулевые элементы треугольных сомножителей  $L$  и  $U$  разложения матрицы  $A$  лежат внутри той же самой ленты, а процессы исключения и обратной подстановки реализуются очень быстро.

Это наш последний подсчет числа действий, и мы должны подчеркнуть один важный момент. Для матрицы типа рассмотренного нами примера конечно-разностной матрицы  $A$  обратная матрица является плотной. Следовательно, при решении системы  $Ax = b$  гораздо лучше знать матрицы  $L$  и  $U$ , чем матрицу  $A^{-1}$ . Умножение вектора  $b$  на матрицу  $A^{-1}$  требует  $n^2$  действий, в то время как  $4n$  действий достаточно для решения систем  $Lc = b$  и  $Ux = c$ , т. е. проведения исключения и обратной подстановки, что позволяет получить решение  $x = U^{-1}c = U^{-1}L^{-1}b = A^{-1}b$ .

Мы надеемся, что последний пример полезен в двух отношениях: он углубляет понимание процесса исключения (который, как мы полагаем, теперь совершенно ясен!) и дает пример системы большого порядка, которая действительно встречается на практике. В следующей главе мы вернемся к изучению «теоретической» структуры линейной системы  $Ax = b$ , а именно к вопросу существования и единственности ее решения  $x$ .

**Упражнение 1.6.1.** Модифицировать пример, рассмотренный в настоящем разделе, заменив элемент  $a_{11} = 2$  на  $a_{11} = 1$ , и найти  $LDU$ -разложение новой трехдиагональной матрицы.

**Упражнение 1.6.2.** Для симметрической матрицы

$$A = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$$

применить процесс исключения для получения нулей в первом столбце под ведущим элементом  $a$  и показать, что после этого матрица порядка два в нижнем правом углу остается симметрической. Таким же способом можно показать, что симметричность сохраняется после каждого этапа исключения.

**Упражнение 1.6.3.** Найти матрицу  $A$  порядка 5, возникающую при аппроксимации задачи

$$-\frac{d^2u}{dx^2} = f(x), \quad \frac{du}{dx}(0) = \frac{du}{dx}(1) = 0,$$

<sup>1)</sup> Приятно отметить, что  $P$  является целым числом, поскольку хотя бы одно из расположенных рядом чисел  $n-1$ ,  $n$  и  $n+1$  должно делиться на три.

при замене краевых условий соотношениями  $u_0 = u_1$  и  $u_6 = u_5$ . Показать, что умножение этой матрицы на вектор  $[1 \ 1 \ 1 \ 1 \ 1]^T$  дает нулевой вектор и, следовательно, матрица  $A$  вырожденная. Аналогично показать, что если функция  $u(x)$  является решением непрерывной задачи, то функция  $u(x) + 1$  также является ее решением, т. е. указанные два условия не устраняют неопределенности в члене  $C + Dx$  и решение задачи не единственно.

**Упражнение 1.6.4.** При  $h = 1/4$  и  $f(x) = 4\pi^2 \sin 2\pi x$  разностные уравнения (29) могут быть представлены в виде системы

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \frac{\pi^2}{4} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

Вычислить значения  $u_i$  и найти их погрешность по сравнению с точным решением  $u = \sin 2\pi x$  в точках  $x = 1/4$ ,  $x = 1/2$  и  $x = 3/4$ .

## ОБЗОРНЫЕ УПРАЖНЕНИЯ

1.1. Для матриц

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \text{ и } B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

вычислить матрицы  $A + B$  и  $AB$ .

1.2. Для тех же самых матриц вычислить  $A^{-1}$ ,  $B^{-1}$  и  $(AB)^{-1}$ .

1.3. С помощью исключения и обратной подстановки решить систему

$$\begin{aligned} 2u - 3v &= 8, \\ 4u - 5v + w &= 15, \\ 2u &+ 4w = 1. \end{aligned}$$

1.4. Для матрицы коэффициентов предыдущей системы построить разложение  $A = LU$ .

1.5. Пусть задана система трех уравнений. Какая матрица  $E$  осуществляет вычитание второго уравнения из третьего?

1.6. Какая матрица  $P$  порядка три осуществляет перестановку первого уравнения и третьего?

1.7. Какая матрица порядка три умножает второе уравнение на  $-1$  и оставляет другие два уравнения без изменения?

1.8. С помощью метода исключения определить, имеет ли решение система

$$\begin{aligned} u + v + w &= 0, \\ u + 2v + 3w &= 0, \\ 3u + 5v + 7w &= 1. \end{aligned}$$

1.9. С помощью метода исключения и перестановок строк решить систему

$$\begin{aligned} u + v - w &= 2, \\ 3u + 3v + w &= 2, \\ u &+ w = 0. \end{aligned}$$

1.10. Построить  $LU$ -разложение матрицы

$$A = \begin{bmatrix} 1 & 4 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix}.$$

1.11. Найти обратную к матрице

$$\begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix}.$$

1.12. Является ли матрица порядка три обязательно обратимой, если ни одна из ее строк не пропорциональна ни одной другой? Если это не так, то привести соответствующий пример.

1.13. С помощью метода Гаусса—Жордана найти обратную к матрице

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

1.14. Объяснить, почему матрица, обратная к треугольной матрице, тоже является треугольной.

1.15. Если  $E$ —матрица порядка два, умножение на которую прибавляет первое уравнение ко второму, то какой вид имеет матрица  $E^{50}$ ? Выписать матрицы  $E$ ,  $E^{50}$  и  $50E$ .

1.16. Выписать систему с матрицей порядка два, которая имеет бесконечно много решений.

## Глава 2

# ТЕОРИЯ СИСТЕМ ЛИНЕЙНЫХ УРАВНЕНИЙ

### § 2.1. ВЕКТОРНЫЕ ПРОСТРАНСТВА И ПОДПРОСТРАНСТВА

В предыдущей главе описано, как проводить процесс исключения (по одному элементу) для упрощения линейной системы  $Ax=b$ . К счастью, этот алгоритм не только облегчает вычислительные решения  $x$ , но и одновременно дает ответ на более теоретические вопросы о его существовании и единственности. Чтобы изучить прямоугольные матрицы столь же хорошо, как и квадратные, нам необходимо посвятить еще один раздел технике исключения. Тогда у нас будут все необходимые нам понятия. Но, концентрируя внимание на отдельных элементах, мы приходим к несколько одностороннему восприятию линейных систем; поэтому главная цель настоящей главы — достичь разностороннего и глубокого понимания проблемы. Нам потребуется понятие **векторного пространства**. Но мы хотим ввести его не с помощью обычной системы аксиом, а на примере трех уравнений с двумя неизвестными:

$$\begin{bmatrix} 1 & 0 \\ 5 & 4 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}. \quad (1)$$

Если бы неизвестных было больше, чем уравнений, мы могли бы надеяться найти одно или даже бесконечно много решений (хотя это и не всегда так). В данном же случае число уравнений больше чем число неизвестных ( $m > n$ ), и можно ожидать, что здесь, как правило, не будет существовать решения. Система (1) будет разрешима только для некоторых правых частей, т. е. для очень «тонкого» подмножества множества всевозможных трехмерных векторов  $b$ . Наша задача — выделить это подмножество.

Способ его описания настолько прост, что позволяет полностью его обозреть.

**2А.** Система  $Ax=b$  разрешима тогда и только тогда, когда вектор  $b$  может быть представлен в виде линейной комбинации столбцов матрицы  $A$ .

Это описание—не что иное, как переформулировка системы  $Ax=b$  в следующем виде:

$$u \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix} + v \begin{bmatrix} 0 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}. \quad (2)$$

Здесь те же самые три уравнения с двумя неизвестными. Но теперь задача выглядит следующим образом: найти такие веса  $u$  и  $v$ , чтобы умножая первый столбец матрицы  $A$  на  $u$ , а второй столбец на  $v$  и складывая полученные векторы, в результате получить вектор  $b$ . Система (1) разрешима, когда такие веса существуют. При этом пара  $(u, v)$  представляет собой решение  $x^1$ .

Таким образом, подмножество правых частей  $b$ , для которых система (1) разрешима, есть **множество всех линейных комбинаций столбцов матрицы  $A$** . В качестве  $b$  можно взять, например, первый столбец этой матрицы; тогда  $u=1$  и  $v=0$ . Другая возможность—второй столбец этой матрицы:  $u=0$  и  $v=1$ . Третий вариант—правая часть  $b$  равна нулю; здесь веса таковы:  $u=0$ ,  $v=0$  (и при таких тривиальных весах получается  $b=0$  вне зависимости от матрицы коэффициентов).

Теперь мы дадим геометрическое описание этого результата: *система  $Ax=b$  разрешима тогда и только тогда, когда вектор  $b$  лежит в плоскости, являющейся линейной оболочкой вектор-столбцов матрицы  $A$*  (рис. 2.1). Эти векторы являются отрезками прямых, соединяющими начало координат  $(0, 0, 0)$  с точками  $(1, 5, 2)$  и  $(0, 4, 4)$  соответственно. Сама плоскость определяется этими двумя прямыми. Она не ограничивается векторами с положительными компонентами: например, весам  $u=-1$ ,  $v=1$  соответствует  $b=(-1, -1, 2)$ .

Построенная плоскость иллюстрирует одно из наиболее важных понятий линейной алгебры—понятие векторного пространства. Под **векторным пространством** мы подразумеваем набор векторов, удовлетворяющих следующим двум условиям:

- (i) сумма  $b+b'$  двух любых векторов  $b, b'$  из векторного пространства принадлежит этому пространству;
- (ii) произведение любого вектора  $b$  из векторного пространства на любой скаляр  $c$  также принадлежит этому пространству.

<sup>1)</sup> Мы используем здесь слово «веса», как синоним слова «коэффициенты», поэтому веса не обязательно положительны.

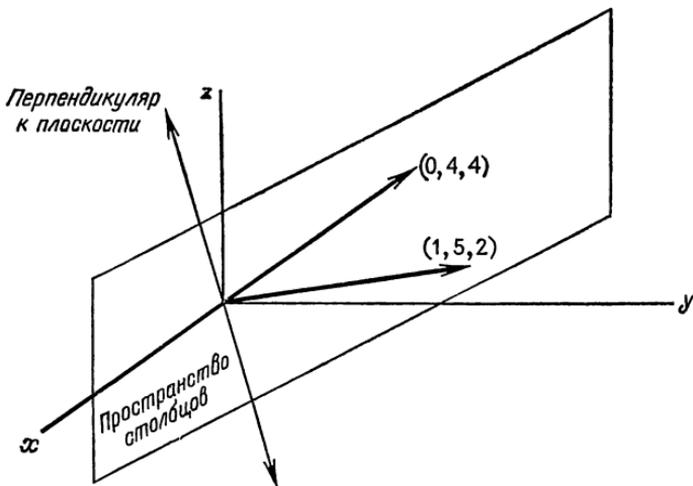


Рис. 2.1. Пространство столбцов матрицы  $A$ —плоскость в трехмерном пространстве.

Другими словами, множество векторов, лежащих в нашей плоскости, замкнуто относительно сложения и умножения на скаляр<sup>1)</sup>. Плоскость является подмножеством множества всех трехкомпонентных векторов  $b = (b_1, b_2, b_3)$ , а это большее множество в свою очередь является векторным пространством, а именно всем трехмерным пространством. Наша плоскость фактически является **подпространством**, т. е. таким подмножеством векторного пространства, которое само является векторным пространством. Мы называем ее **пространством столбцов** матрицы  $A$ , потому что она порождается этими столбцами.

Легко выделить два крайних варианта пространства столбцов матрицы. Первый возникает при нулевой матрице  $A = 0$  порядка  $n$ . Пространство столбцов этой матрицы содержит только один вектор, а именно вектор  $(0, 0, \dots, 0)$ . Тем не менее это (не пустое!) векторное пространство, так как сложение и умножение на скаляр не выводят за его пределы. Это *наименьшее из возможных векторных пространств*. Другой крайний вариант дает единичная матрица  $A = I$  порядка  $n$ . Пространством ее столбцов является все  $n$ -мерное пространство, так как каждый вектор-столбец с  $n$  компонентами может быть представлен в виде линейной комбинации столбцов единичной матрицы. Мы предполагаем, что эти компоненты являются вещественными числами, хотя могли бы рассматривать матрицы  $A$  и векторы  $x$  и  $b$  с комплексными компонентами, причем никаких новых проблем здесь не возникло бы.

<sup>1)</sup> Слово *замкнуто* означает, что эти операции не выводят нас за пределы плоскости. Мы не интересуемся операцией перемножения двух векторов.

Пространство всех вещественных  $n$ -мерных векторов будем обозначать символом  $\mathbf{R}^n$ . Плоскость, изображенная на рис. 2.1, являющаяся подпространством пространства  $\mathbf{R}^3$ .

Различие между *подмножеством* и *подпространством* лучше всего пояснить примерами. Некоторые мы приведем сейчас, а остальные позже. При этом в каждом случае должен быть решен вопрос, выполнены ли условия (i) и (ii).

**Пример 1.** Пусть  $S$  состоит из всех векторов, кратных заданному вектору  $v$ , скажем вектору  $(1, 4, -2)$ . Эти векторы заполняют *прямую в трехмерном пространстве*  $\mathbf{R}^3$ , которая проходит через начало координат, так как нулевой вектор  $0 = 0 \cdot v$  является кратным вектору  $v$ . Такое подмножество  $S$  оказывается подпространством, потому что сумма двух любых векторов, лежащих на этой прямой, как и любое кратное такого вектора, также лежит на этой прямой.

**Пример 2.** Пусть  $S$  состоит только из двух векторов  $(0, 0, 0)$  и  $(1, 0, 0)$ . Тогда условие (ii) не выполнено, потому что произведение  $5 \cdot (1, 0, 0)$  не принадлежит  $S$ . Не выполнено также и условие (i): мы можем взять и в качестве  $b$ , и в качестве  $b'$  вектор  $(1, 0, 0)$ , и тогда их сумма  $b + b'$  множеству  $S$  не принадлежит.

**Пример 3.** Пусть  $S$  состоит из всех векторов, лежащих на прямой, параллельной оси  $x$ , т. е.  $S$  — множество всех векторов, у которых компонента  $x$  произвольна, а компоненты  $y$  и  $z$  фиксированы, скажем  $y = 3$  и  $z = 4$ . Это подмножество тоже не удовлетворяет условиям (i) и (ii), хотя и является прямой. Вектор, равный сумме двух векторов, лежащих на этой прямой, имеет компоненты  $y = 6$ ,  $z = 8$  и, следовательно, лежит на другой прямой. Более того, нулевой вектор не принадлежит  $S$ , а из условия (ii) следует, что *каждое подпространство должно содержать нулевой вектор*.

Закончим это краткое введение. С геометрической точки зрения здесь трудно было что-либо пропустить, поскольку основные факты отражены на рис. 2.1:

(i) множество векторов  $b$ , для которых система  $Ax = b$  разрешима, образует *двумерное пространство столбцов*; два столбца матрицы  $A$  порождают плоскость, причем менее чем два вектора породить эту плоскость не могут;

(ii) векторы, проходящие через начало координат и перпендикулярные этой плоскости, лежат на одной прямой; эти векторы также образуют подпространство, которое одномерно;

(iii) каждый вектор из трехмерного пространства можно разложить в сумму двух векторов, один из которых лежит в этой

плоскости, а второй перпендикулярен ей, и это разложение единственно<sup>1)</sup>).

Из геометрических соображений эти факты более или менее очевидны. Но мы еще не имеем *алгебраического* способа их выражения или проверки, или обобщения на случай системы  $m$  уравнений с  $n$  неизвестными (другими словами, на  $n$  столбцов и  $m$ -мерное пространство). Это одна из двух целей, которые мы преследуем.

Вторая цель настоящей главы «двойственна» к первой. Мы интересуемся не только правыми частями  $b$ , для которых разрешима система (1), но и множеством всех решений  $x$  для каждого такого вектора  $b$ . Правая часть  $b=0$  всегда допускает решение  $x=0$ , но может существовать бесконечно много других решений. (Мы докажем, что в случае, когда неизвестных больше, чем уравнений ( $n > m$ ), система  $Ax=0$  *должна иметь* решения, отличные от нулевого решения  $x=0$ .) *Множество решений системы  $Ax=0$  образует векторное пространство, называемое нуль-пространством матрицы  $A$ .* Оно является подпространством пространства  $\mathbf{R}^n$ , в то время как пространство столбцов матрицы  $A$  было подпространством пространства  $\mathbf{R}^m$ .

Нетрудно решить систему  $Ax=0$  и, таким образом, найти нуль-пространство матрицы  $A$  для приведенного выше примера системы

$$\begin{bmatrix} 1 & 0 \\ 5 & 4 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Первое уравнение дает  $u=0$ , а из второго следует, что  $v=0$ . Поэтому в данном случае нуль-пространство состоит из единственной точки, а именно из нулевого вектора; только одна комбинация столбцов дает вектор  $b=0$ , это комбинация с весами  $u=0$ ,  $v=0$ .

Ситуация изменится, если мы добавим третий столбец, являющийся линейной комбинацией двух других:

$$B = \begin{bmatrix} 1 & 0 & 1 \\ 5 & 4 & 9 \\ 2 & 4 & 6 \end{bmatrix}.$$

Пространство столбцов матрицы  $B$  совпадает с пространством столбцов матрицы  $A$ , потому что новый столбец лежит в плоскости, изображенной на рис. 2.1. Но нуль-пространство этой новой матрицы  $B$  содержит вектор с компонентами  $1, 1, -1$ , а также

<sup>1)</sup> Третий пункт менее ясен, чем остальные, и мы вернемся к нему в § 2.5.

любой вектор, кратный этому вектору:

$$\begin{bmatrix} 1 & 0 & 1 \\ 5 & 4 & 9 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} c \\ c \\ -c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Следовательно, нуль-пространство матрицы  $B$  представляет собой прямую, содержащую все точки  $x=c$ ,  $y=c$ ,  $z=-c$ , где  $c$  пробегает значения от  $-\infty$  до  $\infty$ . Эта прямая проходит через начало координат, как и положено подпространству. Кроме того, перпендикулярное к этому одномерному нуль-пространству пространство (плоскость) оказывается тесно связанным со строками матрицы  $B$  и является особенно важным.

Подведем итог: нам нужно для любой системы  $Ax=b$  уметь находить все правые части  $b$ , для которых эта система разрешима, и все решения системы  $Ax=0$ . Это значит, что мы должны уметь находить размерности введенных выше подпространств и множества векторов, порождающие их. Мы также надеемся, что к концу наших исследований будем ясно представлять себе все четыре тесно связанные между собой подпространства—пространство столбцов матрицы  $A$ , ее нуль-пространство и ортогональные к ним подпространства.

**Упражнение 2.1.1.** Показать, что условия (i) и (ii) в определении векторного пространства независимы, установив, что

(а) подмножество двумерного пространства может быть замкнутым относительно операции сложения и даже вычитания, но не замкнутым относительно операции умножения на скаляры;

(б) подмножество двумерного пространства может быть замкнутым относительно операции умножения на скаляры, но не замкнутым относительно операции сложения векторов.

**Упражнение 2.1.2.** Какие из следующих подмножеств являются подпространствами пространства  $\mathbb{R}^3$ ?

(а) Множество векторов из  $\mathbb{R}^3$  с первой компонентой  $b_1=0$ .

(б) Множество векторов  $b$  с  $b_1=1$ .

(с) Множество векторов  $b$  с  $b_1b_2=0$  (это множество является объединением двух подпространств: плоскости  $b_1=0$  и плоскости  $b_2=0$ ).

(д) Множество, состоящее из одного вектора  $b=(0, 0, 0)$ .

(е) Всевозможные линейные комбинации двух векторов  $u=(1, 1, 0)$  и  $v=(2, 0, 1)$ .

(ф) Множество векторов  $b=(b_1, b_2, b_3)$ , компоненты которых удовлетворяют условию  $b_3-b_2+3b_1=0$ .

**Упражнение 2.1.3.** Проверить, что нуль-пространство любой матрицы  $A$  является подпространством, т. е. что множество решений системы  $Ax=0$  замкнуто относительно сложения и умножения на скаляры. Привести контр-пример для случая  $b \neq 0$ .

## § 2.2. РЕШЕНИЕ $m$ УРАВНЕНИЙ С $n$ НЕИЗВЕСТНЫМИ

К настоящему моменту процесс исключения для квадратных матриц можно считать достаточно хорошо изученным. Поэтому, чтобы проиллюстрировать новые ситуации, возникающие в случае прямоугольной матрицы, нам будет достаточно одного примера. Сами исключения проводятся без значительных изменений, но, когда дело доходит до получения решения путем обратной подстановки, возникают некоторые различия.

Прежде чем переходить к этому примеру, мы проиллюстрируем различные возникающие тут ситуации на примере простейшего скалярного уравнения  $ax=b$ . Эта «система» содержит только одно уравнение с одним неизвестным. Очевидно, что здесь имеется три возможности:

(i) Если  $a \neq 0$ , то для любого  $b$  существует решение  $x=b/a$ , и это решение единственно. Это **невыврожденный** случай (обратимая матрица порядка один).

(ii) Если  $a=0$  и  $b=0$ , то существует бесконечно много решений, потому что любое число  $x$  удовлетворяет уравнению  $0 \cdot x=0$ . Это случай **недоопределенной** системы, т. е. решение существует, но не единственно.

(iii) Если  $a=0$ , а  $b \neq 0$ , то решения не существует. Это случай **несовместной** системы.

Выберем теперь менее тривиальный пример. Забыв на время о правой части  $b$ , будем оперировать только с матрицей

$$A = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 5 \\ -1 & -3 & 3 & 0 \end{bmatrix}$$

порядка  $3 \times 4$ . Так как  $a_{11}=1$ , то с помощью обычных элементарных операций можно получить нули в первом столбце ниже элемента  $a_{11}$ :

$$A \rightarrow \begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 6 & 2 \end{bmatrix}.$$

Второй элемент во второй строке теперь стал нулевым (элемент  $b_{22}$ ), поэтому мы ищем во втором столбце ниже него ненулевой элемент, намереваясь поменять местами строки. В данном случае все элементы второго столбца, стоящие ниже  $b_{22}$ , равны нулю. Если бы первоначальная матрица была квадратной, то это означало бы, что она вырожденная. В случае прямоугольной матрицы могут встретиться любые неожиданности, и у нас нет оснований

прекращать процесс исключения. Единственное, что мы можем сделать — это *перейти к следующему столбцу*, где ведущий элемент не равен нулю. Вычитая удвоенную вторую строку из третьей, получим матрицу

$$U = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Строго говоря, мы должны были бы теперь заняться четвертым столбцом. Однако там мы вновь сталкиваемся с нулем на месте ведущего элемента и ничего больше сделать не можем. Поэтому первый этап исключения завершен.

Окончательная матрица  $U$  имеет верхнюю треугольную форму или, более точно, *верхнюю трапецидальную*, т. е. все ее ненулевые элементы  $u_{ij}$  лежат выше главной диагонали,  $i \leq j$ . Легко видеть, что ненулевые элементы преобразованной матрицы занимают область, имеющую форму «ступенек» или «уступов», изображенную на рис. 2.2. Ненулевые ведущие элементы взяты

$$U = \begin{bmatrix} \textcircled{*} & * & * & * & * & * & * & * & * \\ 0 & \textcircled{*} & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \textcircled{*} & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \textcircled{*} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Рис 2.2. Ненулевые элементы типичной ступенчатой матрицы  $U$ .

в кружок. Элементы же, обозначенные просто символом  $*$ , могут быть как нулевыми, так и ненулевыми.

Проанализируем рис. 2.2:

(i) Ненулевые строки расположены выше нулевых (в противном случае надо было произвести перестановку строк) и каждый ведущий элемент является первым ненулевым элементом в своей строке.

(ii) В каждом столбце все элементы, расположенные ниже ведущего, равны нулю.

(iii) Каждый ведущий элемент лежит правее ведущих элементов предыдущих строк.

**Упражнение 2.2.1.** Ступенчатый вид, изображенный на рис. 2.2, возникает в тех случаях, когда некоторые элементы столбцов — кандидаты на роль ведущих элементов — оказываются равными нулю (такие, как в 3 и 5—8 столб.

цах). Изобразить ступенчатую форму  $5 \times 9$ -матрицы, у которой нет нулевых ведущих элементов. Показать на примере, что сумма двух матриц ступенчатого вида может уже не иметь ступенчатый вид даже в случае матриц порядка 2, поскольку ведущие элементы могут уничтожить друг друга.

Так как мы начали с матрицы  $A$  и закончили матрицей  $U$ , нетерпеливый читатель обязательно спросит: «Связаны ли эти матрицы посредством нижней треугольной матрицы  $L$  соотношением  $A=LU$ , как раньше?» Вероятно, сомневаться в этом нет причин, поскольку шаги исключения не изменились. На каждом шаге по-прежнему производится вычитание одной строки, умноженной на соответствующее число, из строки, расположенной ниже. Более того, обращение каждого шага выполняется так же, как и раньше, путем прибавления того, что вычиталось. Эти обратные шаги следуют в определенном порядке, что позволяет нам записать их в виде матрицы  $L$ :

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 2 & 1 \end{bmatrix}.$$

Читатель легко может проверить, что  $A=LU$ . Нужно также отметить, что матрица  $L$  не прямоугольная, а квадратная. Порядок  $m=3$  этой матрицы равен числу строк в матрицах  $A$  и  $U$ .

В нашем примере не потребовалось осуществлять перестановки строк. Однако в общем случае это необходимо. Как и в § 1.5, это приводит к матрицам перестановки  $P$ . Если все строки матрицы  $A$  переставить до начала исключения, то это даст тот же эффект, что и в соответствующем месте первой главы. Далее, так как мы теперь договорились переходить к следующему столбцу, когда в данном столбце нет ненулевого ведущего элемента, у нас отпала необходимость в предположении, что матрица невырожденная. Этот факт представляет собой основную теорему:

**2В.** Для любой матрицы  $A$  размера  $m \times n$  существуют такие матрица перестановки  $P$ , нижняя треугольная матрица  $L$  с единицами на диагонали и верхняя трапециевидальная матрица  $U$ , ненулевые элементы которой лежат в области, имеющей ступенчатый вид, что  $PA=LU$ .

Теперь наша цель — полностью описать множество всех решений системы  $Ax=b$  (если они существуют).

Начнем со случая однородного уравнения с  $b=0$ . Тогда, поскольку элементарные операции со строками не оказывают никакого влияния на нули в правой части системы  $Ax=0$ , последняя

редуцируется к системе  $Ux=0$ :

$$\begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Неизвестные  $u$ ,  $v$ ,  $w$  и  $y$  распадаются на две группы. Одна группа состоит из **базисных переменных**, которые соответствуют **столбцам с ненулевыми ведущими элементами** (в нашем примере первый и третий столбцы содержат ненулевые ведущие элементы, поэтому  $u$  и  $w$  — базисные переменные). Другая группа образуется **свободными переменными**, соответствующими **столбцам без ведущих элементов**, т. е. второму и четвертому столбцам. Поэтому  $v$  и  $y$  являются свободными переменными.

Чтобы найти наиболее общее решение системы  $Ux=0$  (или, что эквивалентно, системы  $Ax=0$ ), мы можем придавать свободным переменным произвольные значения. Предположим, что мы обозначили эти значения просто через  $v$  и  $y$ . Тогда базисные переменные полностью определяются и могут быть вычислены через свободные переменные с помощью обратной подстановки. Система  $Ux=0$  эквивалентна следующим двум уравнениям:

$$\begin{aligned} 3w + y &= 0, \quad \text{или} \quad w = -\frac{1}{3}y, \\ u + 3v + 3w + 2y &= 0, \quad \text{или} \quad u = -3v - y. \end{aligned}$$

Таким образом, мы имеем «двойную бесконечность» решений нашей системы, зависящих от свободных и независимых параметров  $v$  и  $y$ . Общее решение выражается через эти параметры соотношением

$$x = \begin{bmatrix} -3v - y \\ v \\ -\frac{1}{3}y \\ y \end{bmatrix} = v \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} -1 \\ 0 \\ -\frac{1}{3} \\ 1 \end{bmatrix}. \quad (3)$$

Посмотрим еще раз на последнюю формулу общего решения системы  $Ax=0$ . Вектор  $(-3, 1, 0, 0)$  дает решение, когда свободные переменные имеют значения  $v=1$  и  $y=0$ , а вектор  $(-1, 0, -1/3, 1)$  является решением с  $v=0$  и  $y=1$ . Все остальные решения являются линейными комбинациями этих двух. Ситуация здесь такая же, как и в начале этой главы, где было установлено, что вектор  $b$ , для которого система  $Ax=b$  разрешима, является линейной комбинацией столбцов матрицы  $A$ . В данном случае, правда, мы имеем дело не со столбцами матрицы  $A$ ,

а с ее нуль-векторами, т. е. с решениями системы  $Ax=0$ . Эти векторы не совпадают уже по длине со столбцами матрицы  $A$ : они имеют не  $m$ , а  $n$  компонент, т. е. их длина равна длине строк матрицы  $A$ .

Геометрически получается следующая картина: в четырехмерном пространстве решения системы  $Ax=0$  образуют двумерное подпространство — **нуль-пространство** матрицы  $A$ . Это нуль-пространство может быть описано как «плоскость», которая порождается двумя векторами  $(-3, 1, 0, 0)$  и  $(-1, 0, -1/3, 1)$ . (Читатель не должен требовать четырехмерной картинки.) Линейные комбинации этих двух векторов образуют множество, которое замкнуто относительно сложения и умножения на скаляры, так как эти операции попросту приводят к другим комбинациям тех же самых двух векторов, и все эти комбинации составляют нуль-пространство исходной матрицы.

В настоящий момент уместно привести одну крайне важную теорему. Предположим, что исходная матрица имеет больше столбцов, чем строк, т. е.  $n > m$ . Тогда, так как в процессе исключения не может возникнуть более чем  $m$  ненулевых ведущих элементов (для этого у нас не хватит строк), то *число свободных переменных не может быть меньше, чем  $n-m$* . Более того, число свободных переменных может быть даже больше, если, как в нашем последнем примере, некоторые строки матрицы  $U$  станут нулевыми. Этим свободным переменным могут присваиваться различные значения, что и приводит нас к следующему заключению:

|| **2С.** *Каждая однородная система  $Ax=0$ , имеющая больше неизвестных, чем уравнений ( $n > m$ ), имеет хотя бы одно ненулевое решение, т. е. имеет некоторое решение  $x$ , отличное от нуля.*

Таких решений в действительности должно быть бесконечно много, так как любое кратное  $cx$  решения  $x$  также будет удовлетворять исходной системе. И если имеются дополнительные свободные переменные, то нуль-пространство становится больше, чем прямая  $n$ -мерного пространства. Точнее говоря, *нуль-пространство является подпространством, размерность которого равна числу свободных переменных.*

Неоднородный случай,  $b \neq 0$ , в корне отличается от однородного. Вернемся к первоначальному примеру  $Ax=b$  и применим к обеим частям системы операции, ведущие от матрицы  $A$  к матрице  $U$ , т. е. от системы  $Ax=b$  к системе  $Ux=c$ :

$$\begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \\ y \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 - 2b_1 \\ b_3 - 2b_2 + 5b_1 \end{bmatrix}. \quad (4)$$

Вектор  $c$ , стоящий в правой части, получился в результате применения метода исключения, И так же, как и в предыдущей главе,  $c = L^{-1}b$ .

Совсем не ясно, имеет система уравнений (4) решения или нет. Третье уравнение вызывает здесь серьезное подозрение: его левая часть равна нулю и система может быть совместной, только когда  $b_3 - 2b_2 + 5b_1 = 0$ . Другими словами, множество векторов  $b$ , для которых система  $Ax = 0$  разрешима, не совпадает со всем трехмерным пространством, несмотря на то что неизвестных здесь больше, чем уравнений. Мы уже знаем (см. стр. 63) другой путь исследования возникшего вопроса: система  $Ax = b$  совместна тогда и только тогда, когда  $b$  принадлежит пространству столбцов матрицы  $A$ . Это подпространство порождается четырьмя столбцами

$$\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ 6 \\ -3 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ 9 \\ 3 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ 5 \\ 0 \end{bmatrix}$$

матрицы  $A$  (но не  $U!$ ). Очевидно, что, хотя здесь и четыре вектора, их линейная оболочка заполняет только плоскость в трехмерном пространстве, поскольку второй столбец равен утроенному первому, а четвертый столбец равен сумме первого столбца и третьего с коэффициентом  $1/3$ . (Отметим, что именно эти зависимые столбцы не имеют ведущих элементов.) Пространство столбцов может быть теперь описано двумя совершенно различными способами. С одной стороны, это плоскость, порожденная столбцами 1 и 3; другие столбцы принадлежат этой плоскости и не дают ничего нового. С другой стороны, эта плоскость должна состоять из всех векторов  $(b_1, b_2, b_3)$ , которые удовлетворяют условию  $b_3 - 2b_2 + 5b_1 = 0$ . Это ограничение должно быть наложено на  $b$ , если мы хотим, чтобы система (4) была разрешима. Читатель может убедиться, что эти две плоскости совпадают, показав, что все четыре приведенные выше вектора удовлетворяют этому ограничению. Из геометрических соображений видно, что вектор  $(5, -2, 1)$  перпендикулярен к построенной плоскости и, следовательно, перпендикулярен ко всем столбцам матрицы  $A$ .

Если теперь мы предположим, что вектор  $b$  лежит в определенной выше плоскости, т. е. принадлежит пространству столбцов матрицы  $A$ , то решение системы  $Ax = b$  легко найти. Последнее уравнение в системе (4) становится равенством  $0 = 0$ . Свободным переменным  $v$  и  $y$  мы можем придавать, как и раньше, произвольные значения. Тогда базисные переменные определяются с помощью обратной подстановки:

$$3w + y = b_2 - 2b_1, \quad \text{или} \quad w = -\frac{1}{3}y + \frac{1}{3}(b_2 - 2b_1),$$

$$u + 3v + 3w + 2y = b_1, \quad \text{или} \quad u = -3v - y + 3b_1 - b_2.$$

Мы опять получаем «двойную бесконечность» решений. Выражая компоненты  $u$  и  $w$  через величины  $v$  и  $y$ , можно выписать общее решение в следующем виде:

$$x = v \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} -1 \\ 0 \\ -\frac{1}{3} \\ 1 \end{bmatrix} + \begin{bmatrix} 3b_1 - b_2 \\ 0 \\ \frac{1}{3}(b_2 - 2b_1) \\ 0 \end{bmatrix}. \quad (5)$$

Сравнивая это выражение с общим решением (3), полученным для однородной системы, мы видим, что общее решение (5) отличается от него только дополнительным слагаемым  $(3b_1 - b_2, 0, (b_2 - 2b_1)/3, 0)$ . Этот вектор — *частное решение* системы  $Ax = b$ , а общее решение  $x$  этой системы представляет собой сумму этого частного решения и общего решения однородной системы  $Ax = 0$ . Геометрически изложенное выглядит следующим образом: решения снова принадлежат плоскости в четырехмерном пространстве, но они уже не образуют подпространства, так как эта плоскость не проходит через начало координат, поскольку при  $b \neq 0$  вектор  $x = 0$  не является решением системы  $Ax = b$ . Эта плоскость, конечно, *параллельна* нуль-пространству, полученному раньше, но она смещена относительно него на вектор, дающий частное решение. На самом деле мы и не должны были ожидать, что решения системы  $Ax = b$  образуют подпространство, потому что сумма двух таких решений  $x^1$  и  $x^2$  является решением новой системы  $Ax = 2b$ . Иначе говоря, множество решений образует подпространство только в случае однородной системы.

Отметим, что найденное нами частное решение системы  $Ax = b$  является только одной из точек плоскости всех решений. Любая другая точка этой плоскости с тем же успехом может быть выбрана в качестве частного решения; то частное решение, которое мы выбрали сначала, соответствует выбору значений  $v = 0, y = 0$  для свободных переменных.

Подведем итог распространению метода исключения Гаусса на прямоугольные матрицы.

**2D.** Предположим, что прямоугольная матрица размера  $m \times n$  приведена с помощью элементарных операций и перестановки строк к матрице  $U$  ступенчатого вида. Далее, пусть матрица  $U$  имеет  $r$  ненулевых ведущих элементов, т. е. последние  $m - r$  строк матрицы  $U$  являются нулевыми. Тогда  $r$  переменных будут базисными, а  $n - r$  переменных — свободными, причем они соответствуют столбцам матрицы  $U$  с ведущими элементами и без ведущих элементов соответственно. Нуль-пространство, образованное решениями системы  $Ax = 0$ , имеет

$n - r$  свободных переменных в качестве независимых параметров. Если  $r = n$ , то свободные переменные отсутствуют, а нуль-пространство содержит только один вектор  $x = 0$ .

Решение существует для любой правой части  $b$  тогда и только тогда, когда  $r = m$ ; в этом случае матрица  $U$  не имеет нулевых строк и система  $Ux = c$  может быть решена с помощью обратной подстановки. В случае  $r < m$  матрица  $U$  будет иметь  $m - r$  нулевых строк, и, следовательно, для разрешимости системы  $Ax = b$  на вектор  $b$  необходимо наложить  $m - r$  ограничений, которые присутствуют в явном виде в последних  $m - r$  строках системы  $Ux = c$ . Если существует одно решение системы  $Ax = b$ , то любое другое решение этой же системы отличается от первого на вектор из нуль-пространства матрицы  $A$ .

Число  $r$  называется **рангом** матрицы  $A$ .

**Упражнение 2.2.2.** Построить систему как можно меньшего порядка, имеющую больше неизвестных, чем уравнений, и не имеющую решения.

**Упражнение 2.2.3.** Вычислить  $LU$ -разложение для матрицы

$$A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix}.$$

Определить множество базисных переменных, множество свободных переменных и найти общее решение системы  $Ax = 0$ . Записать его в форме, аналогичной (3). Каков ранг матрицы  $A$ ?

**Упражнение 2.2.4.** Для матрицы

$$A = \begin{bmatrix} 0 & 1 & 4 & 0 \\ 0 & 2 & 8 & 0 \end{bmatrix},$$

определить ступенчатый вид матрицы  $U$ , базисные и свободные переменные и общее решение системы  $Ax = 0$ . После этого применить метод исключения к системе  $Ax = b$ , где вектор  $b$  имеет компоненты  $b_1$  и  $b_2$ . Найти условие разрешимости системы  $Ax = b$  и ее общее решение в форме, аналогичной (5). Каков ранг матрицы  $A$ ?

**Упражнение 2.2.5.** Прodelать все то же самое, что и в предыдущем упражнении, для матрицы

$$A = \begin{bmatrix} 0 & 0 \\ 1 & 2 \\ 4 & 8 \\ 0 & 0 \end{bmatrix},$$

транспонированной к матрице  $A$  из этого упражнения, и вектора  $b$  с компонентами  $b_1, b_2, b_3, b_4$ .

**Упражнение 2.2.6.** Выписать общее решение системы

$$\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 5 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

в виде суммы частного решения и общего решения однородной системы, т. е. в форме (5).

**Упражнение 2.2.7.** Описать множество допустимых правых частей  $b$  для системы

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix},$$

найдя ограничения на  $b$ , превращающие получающееся после исключения третье уравнение в  $0=0$ . Каков ранг матрицы этой системы?

**Упражнение 2.2.8.** При каких значениях  $c$  разрешима система уравнений

$$\begin{aligned} u + v + 2w &= 2, \\ 2u + 3v - w &= 5, \\ 3u + 4v + w &= c. \end{aligned}$$

*Дополнительное замечание.* Часто процесс исключения не заканчивают на матрице  $U$ , а продолжают до получения матрицы еще более простого вида, так называемой ступенчатой матрицы, приведенной по строкам, отличающейся от матрицы  $U$  тем, что все ее ведущие элементы превращены в  $+1$  делением строк на подходящие константы, и нулевые элементы получены не только ниже, но и выше всех ненулевых ведущих элементов. Для рассмотренного выше примера ступенчатая матрица, приведенная по строкам, имеет следующий вид:

$$\begin{bmatrix} 1 & 3 & 0 & 1 \\ 0 & 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Для невырожденной квадратной матрицы соответствующая матрица, приведенная по строкам, будет просто единичной матрицей. Это пример исключения по *методу Гаусса—Жордана* (стр. 45) вместо обычного сведения к разложению  $A=LU$  по методу Гаусса.

Так как исключения Гаусса—Жордана в практических вычислениях с квадратными матрицами требуют больше времени, а ленточная структура матрицы  $A$  не сохраняется в матрице  $A^{-1}$ , то получение такой специальной ступенчатой формы матрицы требует слишком много операций, чтобы ее использовать на компьютере. Тем не менее с теоретической точки зрения она важна как «каноническая форма» матрицы  $A$ : вне зависимости от выбора элементарных операций, включая перестановки строк и деление строк на константу окончательная ступенчатая матрица, приведенная по строкам, всегда одна и та же.

### § 2.3. ЛИНЕЙНАЯ НЕЗАВИСИМОСТЬ, БАЗИС И РАЗМЕРНОСТЬ

Сами по себе числа  $m$  и  $n$  дают неполную картину истинных размеров линейной системы. Матрица в нашем примере имела три строки и четыре столбца, но фактически третья строка была только комбинацией первых двух. После проведения исключения она переходит в нулевую строку и не оказывает никакого влияния на решение однородной системы  $Ax=0$ . Четыре столбца также не могут быть независимыми и пространство столбцов вырождается в двумерную плоскость, потому что второй и четвертый столбцы являются линейными комбинациями первого и третьего.

Важной величиной, которая возникла только в предыдущем параграфе, является **ранг**  $r$ . На стр. 75 ранг был введен чисто вычислительным способом как число ненулевых строк в окончательной матрице  $U$ . Это определение настолько механическое, что может быть предоставлено вычислительной машине. Но было бы неверно остановиться на нем, потому что ранг имеет простой и наглядный смысл: он равен числу линейно независимых строк в матрице  $A$ .

Наша цель — дать этой и подобным ей величинам определения скорее математические, чем вычислительные.

Для начала определим понятие **линейной независимости**. Пусть дано множество векторов  $v_1, v_2, \dots, v_k$ , и рассматриваются всевозможные их линейные комбинации  $c_1v_1 + c_2v_2 + \dots + c_kv_k$ . Очевидно, что тривиальная комбинация со всеми весами, равными нулю ( $c_i=0, i=1, \dots, k$ ), есть нулевой вектор. Вопрос состоит в том, существуют ли другие линейные комбинации, равные нулевому вектору.

**2E.** Если ни одна нетривиальная комбинация векторов  $v_1, \dots, v_k$  не равна нулю, т. е.

$$c_1v_1 + \dots + c_kv_k = 0 \Leftrightarrow c_1 = \dots = c_k = 0, \quad (6)$$

то эти векторы называются **линейно независимыми**. В противном случае эти векторы называются **линейно зависимыми**, и один из них является линейной комбинацией остальных.

**Пример 1.** Если один из заданных векторов, скажем  $v_1$ , является нулевым вектором, то множество векторов  $v_1, \dots, v_k$  линейно зависимо. Действительно, мы можем выбрать  $c_1=3$ , а все остальные  $c_i$  положить равными 0,  $i \neq 1$ . Тогда эта нетривиальная линейная комбинация равна нулевому вектору.

**Пример 2.** Три строки матрицы

$$A = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 5 \\ -1 & -3 & 3 & 0 \end{bmatrix}$$

линейно зависимы. Если эти строки обозначить буквами  $v_1$ ,  $v_2$  и  $v_3$ , то  $5v_1 - 2v_2 + v_3 = 0$ . (Это такая же комбинация, как и полученная ранее для величин  $b_1$ ,  $b_2$  и  $b_3$ , которая должна была равняться нулю, чтобы уравнения системы были совместны; в противном случае третье уравнение системы  $Ux = c$  было бы противоречиво.)

**Пример 3.** Строки единичной матрицы порядка  $n$

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

линейно независимы. Мы обозначим вектор-строки этой матрицы через  $e_1, \dots, e_n$  соответственно. Они представляют собой единичные векторы в направлении осей координат:

$$e_1 = (1, 0, \dots, 0), \dots, e_n = (0, 0, \dots, 1). \quad (7)$$

Обычная процедура доказательства линейной независимости этих векторов такова. *Предположим, что некоторая линейная комбинация  $c_1e_1 + \dots + c_n e_n$  векторов  $e_1, \dots, e_n$  равна нулевому вектору, и требуется доказать, что все веса  $c_i$ ,  $i = 1, \dots, n$ , должны быть равны нулю.* В данном случае мы можем просто вычислить, что

$$c_1e_1 + \dots + c_n e_n = (c_1, c_2, \dots, c_n),$$

и, поскольку эта комбинация является нулевым вектором, что очевидно,  $c_i = 0$ ,  $i = 1, \dots, n$ . Отсюда, так как только тривиальная комбинация дает нулевой вектор, следует, что координатные векторы  $e_1, \dots, e_n$  линейно независимы.

**Пример 4.** Предположим, что матрица  $U$  является верхней треугольной матрицей порядка  $n$  с ненулевыми элементами на диагонали (ведущими элементами). Тогда строки матрицы  $U$  линейно независимы. Для доказательства мы начнем, как и раньше, с предположения о том, что некоторая линейная комбинация строк равна 0,  $c_1v_1 + \dots + c_nv_n = 0$ . Так как  $U$  является верхней треугольной матрицей, ее первый столбец имеет только одну не-

нулевую компоненту  $u_{11}$ . Поэтому первая компонента вектора  $c_1 v_1 + \dots + c_n v_n$  равна  $c_1 u_{11}$ , а так как  $c_1 v_1 + \dots + c_n v_n = 0$ , то  $c_1 u_{11} = 0$  и, следовательно,  $c_1 = 0$ . Рассматривая далее вторые компоненты, видим, что  $c_1 u_{12} + c_2 u_{22} = 0$  и, так как  $c_1 = 0$  и  $u_{22} \neq 0$ , мы получаем  $c_2 = 0$ . Продолжая таким же образом, легко доказать, что все веса будут равны нулю. Таким образом, только одна тривиальная комбинация дает нулевой вектор, и, следовательно, строки матрицы  $U$  линейно независимы.

Те же самые рассуждения применимы и к столбцам матрицы  $U$ , а именно столбцы с ненулевыми ведущими элементами линейно независимы. Аналогичные рассуждения применимы фактически к любой ступенчатой матрице при условии, что мы рассматриваем только строки (или столбцы) с ненулевыми ведущими элементами.

**2F.**  *$r$  ненулевых строк ступенчатой матрицы  $U$  линейно независимы. Линейно независимы также  $r$  ее столбцов, которые содержат ведущие элементы.*

Подчеркнем, что понятие линейной независимости не зависит от выбора осей координат. Если заданы  $k$  точек в  $n$ -мерном пространстве, то векторы, выходящие из начала координат и оканчивающиеся в этих точках, либо линейно независимы, либо линейно зависимы, вне зависимости от того, где мы поместим оси координат. Вращение осей может изменить координаты этих векторов, но не окажет никакого влияния на их линейную зависимость или независимость.

С другой стороны, для заданного произвольного множества  $v_1, \dots, v_k$  векторов проверка их зависимости или независимости, очевидно, требует вычислений. Так как мы должны рассматривать линейные комбинации  $c_1 v_1 + \dots + c_k v_k$ , то естественно будет сформировать матрицу  $A$ , столбцами которой будут данные векторы. Тогда, если мы через  $c$  обозначим вектор весов  $c = (c_1, \dots, c_k)$  и вспомним операцию матричного умножения, то легко убедимся, что

$$Ac = \begin{bmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_k \\ | & | & \dots & | \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} = c_1 v_1 + \dots + c_k v_k.$$

Следовательно, векторы  $v_1, \dots, v_k$  линейно зависимы тогда и только тогда, когда существует нетривиальное решение системы  $Ac = 0$ . Этот факт устанавливается посредством процесса исключения. Если ранг матрицы  $A$  равен  $k$ , то свободных переменных не будет, нуль-пространство будет содержать только нулевой вектор и, следовательно, векторы будут линейно независимы. Если

же ранг матрицы  $A$  меньше, чем  $k$ , то по крайней мере одна переменная будет свободной. Поэтому ее можно выбрать отличной от нуля и, следовательно, столбцы матрицы  $A$ , т. е. векторы  $v_1, \dots, v_k$  будут линейно зависимыми.

Один случай особенно важен. Пусть векторы имеют  $m$  компонент и  $A$  будет прямоугольной матрицей порядка  $m \times k$ . Предположим, что  $k > m$ . Тогда матрица  $A$  не может иметь ранг  $k$ , потому что число ненулевых ведущих элементов не может превысить число строк. Ранг должен быть меньше, чем  $k$ , а однородная система  $Ac=0$ , имеющая больше неизвестных, чем уравнений, всегда имеет ненулевое решение  $c$ .

|| **2G.** Множество  $k$  векторов из  $\mathbf{R}^m$  обязательно будет линейно зависимо, если  $k > m$ .

Читателю нетрудно увидеть здесь иную форму утверждения 2С.

**Пример 5.** Рассмотрим матрицу

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & 2 \end{bmatrix},$$

имеющую три столбца. В двумерном пространстве  $\mathbf{R}^2$  не может быть трех линейно независимых векторов. Чтобы найти комбинацию столбцов, равную нулевому вектору, мы решим систему  $Ac=0$ :

$$A \rightarrow U = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Если присвоить значение 1 свободной переменной  $c_3$ , то обратной подстановкой из  $Uc=0$  получим  $c_2 = -1$ ,  $c_1 = 1$ . Линейная комбинация с этими тремя весами, т. е. первый столбец минус второй плюс третий, равна нулю.

**Упражнение 2.3.1.** Установить, будут ли линейно независимыми следующие векторы:

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad v_4 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

**Упражнение 2.3.2.** Установить, будут ли линейно зависимы или линейно независимы строки матрицы  $A$  из примера 5, столбцы которой линейно зависимы.

**Упражнение 2.3.3.** Доказать, что если хотя бы один диагональный элемент матрицы

$$T = \begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}$$

равен нулю, то строки этой матрицы линейно зависимы.

**Упражнение 2.3.4.** Верно ли утверждение, что если векторы  $v_1, v_2, v_3$  линейно независимы, то векторы  $w_1 = v_1 + v_2, w_2 = v_1 + v_3, w_3 = v_2 + v_3$  также линейно независимы? (Указание: предположить, что  $c_1 w_1 + c_2 w_2 + c_3 w_3 = 0$ , и найти все возможные коэффициенты  $c_1, c_2, c_3$ .)

Следующий шаг в обсуждении векторных пространств — определить, что мы понимаем под *множеством векторов, порождающим данное векторное пространство*. Мы использовали этот термин в начале главы, когда говорили о плоскости, которая являлась линейной оболочкой двух столбцов матрицы, и называли эту плоскость пространством столбцов. Общее определение таково:

**2Н.** Если векторное пространство  $V$  состоит из всех линейных комбинаций векторов  $w_1, \dots, w_l$ , то говорят, что эти векторы *порождают* векторное пространство  $V$ , а само  $V$  называют *линейной оболочкой* этих векторов. Другими словами, каждый вектор  $v$  из  $V$  может быть представлен в виде линейной комбинации векторов  $w_1, \dots, w_l$ ,

$$v = c_1 w_1 + \dots + c_l w_l \quad (8)$$

с некоторыми коэффициентами  $c_i$ .

Такое определение допускает, что для вектора  $v$  может существовать несколько наборов коэффициентов  $c_i$ , т. е. коэффициенты не обязательно единственны, так как порождающее множество векторов может быть слишком большим и даже содержать нулевой вектор.

**Пример 6.** Векторы  $w_1 = (1, 0, 0), w_2 = (0, 1, 0), w_3 = (-2, 0, 0)$  порождают плоскость  $xy$  в трехмерном пространстве. Ту же плоскость порождают первые два вектора, тогда как векторы  $w_1$  и  $w_3$  порождают только прямую.

**Пример 7.** Пространство столбцов некоторой матрицы  $A$  размера  $m \times n$  является линейной оболочкой столбцов этой матрицы. Оно служит подпространством  $m$ -мерного пространства  $\mathbf{R}^m$  (конечно, оно может совпадать с  $\mathbf{R}^m$ ). Пространство строк матрицы  $A$  определяется аналогичным образом. Это подпространство пространства  $\mathbf{R}^n$ , порождаемое строками матрицы  $A$  (мы можем считать строки элементами из  $\mathbf{R}^n$ , несмотря на то что компоненты записаны горизонтально, а не в виде столбца). Если  $m = n$ , то как пространство строк, так и пространство столбцов будут подпространствами пространства  $\mathbf{R}^n$  и могут даже совпадать.

**Пример 8.** Координатные векторы  $e_1 = (1, 0, \dots, 0), \dots, e_n = (0, \dots, 0, 1)$  порождают пространство  $\mathbf{R}^n$ . Для доказательства достаточно показать, как любой вектор  $x = (x_1, x_2, \dots, x_n)$  может быть записан в виде линейной комбинации векторов  $e_i, i = 1, \dots, n$ .

Очевидно, что искомыми весами будут сами компоненты  $x_i$ ,  $i = 1, \dots, n$ , т. е.

$$x = x_1 e_1 + x_2 e_2 + \dots + x_n e_n.$$

В данном примере мы, кроме этого, знаем, что множество векторов  $e_1, \dots, e_n$  является линейно независимым, т. е., грубо говоря, ни один вектор этого множества не является лишним. Матрица в примере (5) имеет лишние столбцы, не вносящие ничего нового в пространство столбцов, а здесь порождающее множество векторов имеет минимальный размер. Иначе говоря, если какой-либо вектор  $e_i$  будет удален, оставшиеся векторы не будут порождать все  $\mathbb{R}^n$ . Такое множество векторов называют базисом:

**21. Базисом** векторного пространства называют множество векторов, обладающее следующими свойствами:

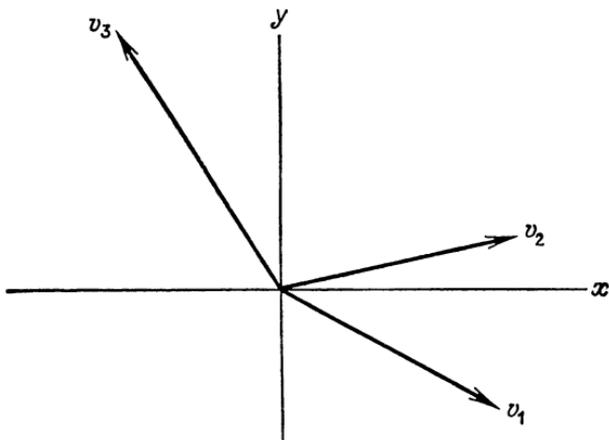
- (1) Оно линейно независимо.
- (2) Оно порождает это векторное пространство.

Сочетание этих двух свойств является основополагающим для теории векторных пространств. Они означают, что каждый вектор  $v$  из данного векторного пространства может быть разложен единственным образом в линейную комбинацию базисных векторов,  $v = a_1 v_1 + \dots + a_k v_k$ . (Такое разложение существует, потому что векторное пространство порождается своим базисом; если же имеется еще одно разложение  $v = b_1 v_1 + \dots + b_k v_k$ , то, вычитая второе разложение из первого, получим  $0 = \sum_{i=1}^k (a_i - b_i) v_i$ , откуда в силу линейной независимости векторов  $v_1, \dots, v_k$  имеем  $a_i - b_i = 0$ ,  $i = 1, \dots, k$ . Следовательно, веса в разложении вектора  $v$  по базисным векторам  $v_i$  однозначно определяются вектором  $v$ .)

**Пример 9.** Рассмотрим обычную плоскость  $xy$  (рис. 2.3). Вектор  $v_1$  сам по себе линейно независим, но не порождает  $\mathbb{R}^2$ . Три вектора  $v_1, v_2, v_3$ , конечно, порождают  $\mathbb{R}^2$ , но не являются линейно независимыми. Любые два из этих векторов, скажем  $v_1$  и  $v_2$ , обладают обоими свойствами — они порождают  $\mathbb{R}^2$  и линейно независимы — и поэтому образуют базис в  $\mathbb{R}^2$ . Отметим, что *базис векторного пространства не единствен*.

**Пример 10.** Рассмотрим ступенчатую матрицу  $U$ , соответствующую матрице  $A$  из примера 2:

$$U = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Рис. 2.3. Порождающее множество и базис в  $\mathbb{R}^2$ .

Четыре ее столбца, как всегда, порождают пространство столбцов, но не являются линейно независимыми. Существует много способов выбора базиса в этом пространстве, но мы предпочитаем выбирать его специальным образом: *столбцы, содержащие ведущие элементы* (в данном случае первый и третий, соответствующие базисным переменным), *составляют базис пространства столбцов*. Мы отмечали в утверждении 2F, что эти столбцы линейно независимы, но легко видеть, что они, кроме того, порождают пространство столбцов. В действительности пространство столбцов матрицы  $U$  является  $xy$ -плоскостью в  $\mathbb{R}^3$ . Заметим также, что первые две строки матрицы  $U$  (и вообще ненулевые строки любой ступенчатой матрицы) образуют базис ее пространства строк.

**Упражнение 2.3.5.** Описать словами или изобразить графически в плоскости  $xy$  пространство столбцов и пространство строк матрицы  $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$ . Найти базис пространства столбцов.

**Упражнение 2.3.6.** Указав ведущие элементы, найти базис пространства столбцов матрицы

$$U = \begin{bmatrix} 0 & 1 & 4 & 3 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Выразить каждый столбец, не вошедший в базис, в виде линейной комбинации базисных столбцов. Найти также базис пространства строк.

**Упражнение 2.3.7.** Предположим, что мы представляем себе все матрицы порядка  $2 \times 2$  как «векторы». Хотя эти матрицы не являются векторами в обычном смысле, но мы определили правила сложения матриц и умножения

матриц на скаляр, и множество матриц замкнуто относительно этих операций. Найти базис в этом векторном пространстве. Какое подпространство порождается множеством всех ступенчатых матриц  $U$ ?

**Упражнение 2.3.8.** Найти два различных базиса в подпространстве всех векторов из  $\mathbf{R}^3$ , имеющих равные первые две компоненты.

**Упражнение 2.3.9.** Найти контрпример к следующему утверждению: если  $v_1, \dots, v_4$  составляют базис векторного пространства  $\mathbf{R}^4$  и если  $W$  — подпространство этого пространства, то некоторое подмножество этих базисных векторов будет образовывать базис подпространства  $W$ .

Несмотря на тот факт, что выбор базиса не единствен, так как его построение может быть осуществлено бесконечно многими способами, существует нечто общее для всех базисов. Это свойство является внутренним свойством самого пространства:

|| **2J.** Любые два базиса векторного пространства  $V$  содержат одинаковое число векторов. Это число, фигурирующее во всех базисах и выражающее число «степеней свободы» пространства, называется **размерностью** пространства  $V$ .

Конечно, мы должны доказать, что все возможные базисы содержат одно и то же число векторов. Прежде всего мы вернемся к некоторым из рассмотренных примеров и установим размерности соответствующих пространств. Каждый базис плоскости  $xy$  на рис. 2.3 содержит два вектора, и размерность этой плоскости равна двум. В общем случае размерность пространства  $\mathbf{R}^n$  равна  $n$ , причем координатные векторы  $e_i, i = 1, \dots, n$ , образуют удобный базис этого пространства. Пространство строк матрицы  $U$  в примере 10 является двумерным подпространством пространства  $\mathbf{R}^4$ , а ее пространство столбцов является двумерным подпространством пространства  $\mathbf{R}^3$ . Нулевая матрица  $A$  — это исключительный случай, поскольку как пространство ее столбцов, так и пространство ее строк состоит только из одного нулевого вектора пространств  $\mathbf{R}^m$  и  $\mathbf{R}^n$  соответственно. Мы будем считать, что базисом такого пространства является пустое множество и что его размерность равна нулю.

Сформулируем и докажем теперь теорему, которая эквивалентна утверждению 2J:

|| **2K.** Пусть  $v_1, \dots, v_m$  и  $w_1, \dots, w_n$  — два базиса одного и того же векторного пространства  $V$ . Тогда  $m = n$ .

**Доказательство.** Предположим, например, что  $m < n$ , и постараемся прийти к противоречию. Так как векторы  $v_1, \dots, v_m$  образуют базис, то они должны порождать все пространство  $V$  и, следовательно, каждый вектор  $w_j, j = 1, \dots, n$ , может быть записан в виде линейной комбинации векторов  $v_i$ ,

$i = 1, \dots, m,$

$$\omega_j = a_{1j}v_1 + \dots + a_{mj}v_m = \sum_{i=1}^m a_{ij}v_i.$$

Правда, мы не можем определить веса  $a_{ij}$ , но тем не менее можем рассмотреть линейные комбинации векторов  $\omega_j$ :

$$c_1\omega_1 + \dots + c_n\omega_n = \sum_{j=1}^n c_j \left[ \sum_{i=1}^m a_{ij}v_i \right] = \sum_{i=1}^m v_i \left[ \sum_{j=1}^n a_{ij}c_j \right]. \quad (9)$$

Это изменение порядка суммирования соответствует другому способу сложения всех элементов матрицы—мы можем сначала суммировать вдоль каждой строки и сложить затем вместе эти подсуммы или же сначала суммировать вдоль каждого столбца.

Рассмотрим теперь уравнения  $\sum a_{ij}c_j = 0$  для  $i = 1, \dots, m$ . Это система  $m$  однородных уравнений с  $n$  неизвестными, причем  $m < n$ . Следовательно, она имеет решение  $c = (c_1, \dots, c_n)$ , отличное от тривиального решения  $c = 0$ . Подставляя  $\sum a_{ij}c_j = 0$  в (9), получаем для этих весов  $c_1, \dots, c_n$  равенство

$$c_1\omega_1 + \dots + c_n\omega_n = 0.$$

Отсюда следует, что векторы  $\omega_j$ ,  $j = 1, \dots, n$ , линейно зависимы. Однако это противоречит предположению о том, что они образуют базис, и мы должны откинуть возможность неравенства  $m < n$ .

Приведенное доказательство аналогично использованному на стр. 80 для доказательства того факта, что каждое множество из  $m+1$  векторов пространства  $\mathbb{R}^m$  должно быть линейно зависимым. Фактически мы пришли к следующему общему результату: *в подпространстве размерности  $k$  ни одно множество векторов, состоящее более чем из  $k$  векторов, не может быть линейно независимым, и ни одно множество, состоящее менее чем из  $k$  векторов, не может порождать все пространство.*

Имеются и другие «двойственные» утверждения, из которых мы приведем только одно. Оно позволяет нам начинать поиск базиса пространства с любого множества векторов, либо слишком маленького, либо слишком большого.

**2Л.** Любое линейно независимое множество векторов векторного пространства  $V$  может быть дополнено до базиса путем добавления новых векторов, если это необходимо.

Любое порождающее множество векторов векторного пространства  $V$  может быть превращено в базис путем отбрасывания некоторых векторов этого множества, если это необходимо.

**Упражнение 2.3.10.** Предположим, что векторы  $x$ ,  $y$  и  $z$  из  $\mathbb{R}^3$  удовлетворяют уравнению  $x + y + z = 0$ . Подпространства каких размерностей они могут образовывать? Проиллюстрировать каждую возможность отдельным примером.

**Упражнение 2.3.11.** Для матрицы  $A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 4 \end{bmatrix}$  расширить множество строк до базиса в  $\mathbb{R}^3$  и выделить в множестве столбцов базис пространства  $\mathbb{R}^3$ .

**Упражнение 2.3.12.** Предположим, что пространство  $V$  имеет размерность  $k$ . Доказать, используя, например, теорему 2L, что

(i) любые  $k$  линейно независимых векторов образуют базис в пространстве  $V$ ;

(ii) любые  $k$  векторов, порождающие пространство  $V$ , образуют базис этого пространства.

Иначе говоря, требуется доказать, что если число векторов, равно размерности пространства, то одно из свойств базиса является следствием другого.

**Упражнение 2.3.13.** Найти размерность пространства симметрических матриц размера  $3 \times 3$  и определить его базис.

**Упражнение 2.3.14.** Доказать, что если  $V$  и  $W$  — трехмерные подпространства пространства  $\mathbb{R}^5$ , то  $V$  и  $W$  должны иметь общий ненулевой вектор. Указание: начать с рассмотрения базисов этих двух подпространств, которые вместе содержат 6 элементов.

Сделаем одно последнее замечание о языке линейной алгебры. Мы нигде не использовали выражений «базис матрицы», «ранг пространства строк» или «размерность базиса». Они не имеют смысла. Есть *размерность пространства строк матрицы*, и она совпадает с *рангом этой матрицы*.

## § 2.4. ЧЕТЫРЕ ОСНОВНЫХ ПОДПРОСТРАНСТВА

В предыдущем параграфе мы имели дело скорее с определениями, чем с построениями. Мы знаем, что базис существует, но не знаем, как его найти. Теперь же, начиная с явного описания подпространства, мы будем стремиться вычислить его конкретный базис.

Обычно подпространства описывают одним из двух способов. Первый способ — это когда задается множество векторов, порождающих данное подпространство; так, например, обстояло дело с пространством строк или пространством столбцов, когда указывались порождающие строки или столбцы. Второй способ заключается в задании перечня ограничений на подпространство, т. е. когда указывают не векторы, порождающие это подпространство, а ограничения, которым должны удовлетворять векторы этого подпространства. Нуль-пространство, например, состоит из всех векторов, удовлетворяющих системе  $Ax = 0$ , и каждое уравнение этой системы представляет собой такое ограничение. При описании подпространства первым способом могут оказаться лишними некоторые строки или столбцы, а при втором — некоторые ограничения. В обоих случаях, для того чтобы можно было выписать базис, необходим некоторый систематический подход.

Читатель может догадаться, что это за подход. А именно, мы покажем, как использовать матрицы  $L$  и  $U$  (и  $P$ ), которые возникают в процессе исключения, чтобы найти базис каждого из подпространств, связанных с матрицей  $A$ . Затем, хотя это и делает настоящий раздел длиннее других, мы рассмотрим два экстремальных случая:

(i) когда ранг очень мал ( $r=1$ ), пространства строк и столбцов особенно просты;

(ii) когда ранг очень велик ( $r=n$ ,  $r=m$  или  $r=m=n$ ), матрица  $A$  имеет либо обратную слева матрицу  $B$ , либо обратную справа матрицу  $C$ , либо обычную обратную  $A^{-1}$ .

Чтобы обсудить вопрос всесторонне, рассмотрим последовательно каждое из четырех основных подпространств. При этом, поскольку легче всего находятся подпространства, связанные с матрицей  $U$ , нашей основной задачей будет установление их связи с подпространствами для исходной матрицы  $A$ .

**1. Пространство строк матрицы  $A$ .** После применения метода исключения к матрице  $A$  мы получаем ступенчатую матрицу  $U$ . Пространство строк матрицы  $U$  очевидно: его размерность равна рангу  $r$ , а базис — это ее  $r$  ненулевых строк. К счастью, его столь же легко связать с матрицей  $A$ .

**2М.** *Пространство строк матрицы  $A$  имеет ту же самую размерность  $r$ , что и пространство строк матрицы  $U$ , и тот же самый базис, поскольку эти два пространства строк совпадают.*

Это так, поскольку каждая элементарная операция оставляет неизменным пространство строк. Каждая строка в новой матрице  $U$  является линейной комбинацией строк исходной матрицы  $A$ , и потому новое пространство строк содержится в старом, а так как каждый шаг исключения может быть обращен с помощью другой элементарной операции, то старое пространство строк также содержится в новом.

Заметим, что мы не начинали с  $m$  строк матрицы  $A$ , которые порождают ее пространство строк, и не исключали  $m-r$  из них для построения базиса, хотя, согласно 2L, мы могли бы поступить так. Но решить, какие строки нужно оставить, а какие исключить, бывает трудно, поэтому оказывается гораздо легче брать ненулевые строки матрицы  $U$ .

Приведенные рассуждения, кроме того, оправдывают данное ранее определение ранга матрицы  $A$ . Тогда мы брали  $r$  равным числу ненулевых ведущих элементов (или ненулевых строк) матрицы  $U$ , не рассматривая все различные ступенчатые матрицы, которые могут быть получены при различных последовательностях

шагов исключения. Теперь же мы знаем, что все базисы имеют одно и то же число членов. Следовательно, ранг  $r$  матрицы  $A$  может быть определен по-новому: *рангом матрицы  $A$  является размерность ее пространства строк.*

**2. Нуль-пространство матрицы  $A$ .** Напомним, что первоначальной целью исключения было упрощение системы линейных уравнений без изменения совокупности решений. Система  $Ax=0$  редуцировалась к системе  $Ux=0$ , и этот процесс обратим. Следовательно, *нуль-пространство матрицы  $A$  совпадает с нуль-пространством матрицы  $U$ .* Среди  $m$  ограничений, налагаемых  $m$  уравнениями  $Ax=0$ , только  $r$  линейно независимы. Они выражаются любыми  $r$  линейно независимыми строками матрицы  $A$ , или  $r$  ненулевыми строками матрицы  $U$ . Если мы воспользуемся последним замечанием, получим вполне определенный способ нахождения базиса нуль-пространства.

**2N.** Нуль-пространство матрицы  $A$  (которое мы обозначим через  $\mathfrak{N}(A)$ ) имеет размерность  $n-r$ . Его базис можно построить путем сведения системы уравнений  $Ax=0$  к эквивалентной системе  $Ux=0$ , которая имеет  $n-r$  свободных переменных, соответствующих столбцам матрицы  $U$ , не содержащим ведущих элементов. Присваивая по очереди каждой свободной переменной значение 1, а остальным свободным переменным значения 0 и разрешая систему  $Ux=0$  с помощью обратной подстановки относительно оставшихся (базисных) переменных, мы получим  $n-r$  векторов, составляющих базис пространства  $\mathfrak{N}(A)$ .

В нашем примере (стр. 71) свободными переменными были  $v$  и  $y$ , а базис образовывали векторы

$$x_1 = \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad v=1, \quad y=0; \quad x_2 = \begin{bmatrix} -1 \\ 0 \\ -\frac{1}{3} \\ 1 \end{bmatrix}, \quad v=0, \quad y=1.$$

Легко видеть (как в примере, так и в общем случае), что векторы  $x_i$  должны быть линейно независимыми. Действительно, любая линейная комбинация  $c_1x_1 + c_2x_2$  имеет значение  $c_1$  на месте  $v$ -компоненты и значение  $c_2$  на месте  $y$ -компоненты. Следовательно, существует только одна возможность получить равенство  $c_1x_1 + c_2x_2 = 0$  — это положить  $c_1 = c_2 = 0$ . Эти два вектора порождают нуль-пространство. Общее решение системы  $Ax=0$  записывается в виде линейной комбинации  $vx_1 + yx_2$ . Таким образом,  $n-r = 4-2=2$  вектора  $x_i$  образуют базис.

Нуль-пространство матрицы  $A$  называется также *ядром* этой матрицы  $A$ , а его размерность называется *дефектом* матрицы.

Обозначая дефект матрицы  $A$  символом  $\nu(A)$ , запишем соотношение, устанавливающее взаимосвязь дефекта и ранга

$$\nu(A) = \dim \mathfrak{N}(A) = n - r.$$

Эту формулу необходимо запомнить.

**3. Пространство столбцов матрицы  $A$ .** Пространство столбцов матрицы  $A$  часто называют *образом* этой матрицы и обозначают символом  $\mathfrak{R}(A)$ . Это согласуется с обычным определением образа функции  $f$  как множества всех возможных значений  $f(x)$ . Если определена некоторая функция  $f(x)$ , то  $x$  принадлежит области определения этой функции, а значение  $f(x)$  принадлежит образу. В нашем случае (с векторами) в качестве функции  $f(x)$  можно взять  $f(x) = Ax$ . Ее область определения состоит из всех векторов  $x \in \mathbb{R}^n$ , а образ состоит из всех векторов  $Ax$  (другими словами, из всех векторов  $b$ , для которых разрешима система  $Ax = b$ ). Мы знаем, что это множество совпадает со всеми линейными комбинациями вектор-столбцов матрицы  $A$ , т. е. образ матрицы — это то же самое, что и ее пространство столбцов. Мы хотим сохранить удобный термин «пространство столбцов», но позаимствовать и краткое обозначение  $\mathfrak{R}(A)$ .

Наша задача — найти базис подпространства  $\mathfrak{R}(A)$  и его размерность. Одна разумная идея заключается в следующем: сделать столбцы матрицы  $A$  строками новой матрицы, и тогда мы вновь будем иметь дело с пространством строк. Обозначим новую матрицу через  $A^T$  и будем называть ее *транспонированной матрицей* к матрице  $A$ . Так как столбцы матрицы  $A$  являются строками матрицы  $A^T$ , то последняя будет прямоугольной матрицей размера  $n \times m$ . Каждый элемент матрицы  $A$  при транспонировании переносится в новую позицию, симметричную первоначальной позиции относительно главной диагонали, т. е.  $(A^T)_{ij} = A_{ji}$ . В то же время строки матрицы  $A$  становятся столбцами матрицы  $A$ , и отсюда вытекает полезное замечание: *пространство строк матрицы  $A$  совпадает с пространством столбцов матрицы  $A^T$ , т. е. с  $\mathfrak{R}(A^T)$* . Это особенно полезно, потому что  $\mathfrak{R}(A^T)$  порождается вектор-столбцами (столбцы матрицы  $A^T$  являются строками матрицы  $A$ , записанными вертикально), и поэтому теперь даже понятие пространства строк отвечает соглашению о том, что векторы должны записываться вертикально.

Конечно, мы можем привести матрицу  $A^T$  к ступенчатому виду и таким образом изучать пространство столбцов матрицы  $A$ . Однако мы хотим не этого. Введение транспонированной матрицы полезно с многих точек зрения, но не с этой. Мы хотим избежать введения новой величины, ранга матрицы  $A^T$ , и приведения

матрицы  $A^T$  к матрице ступенчатого вида. Причем, если это возможно, мы предпочли бы вычислить размерность пространства столбцов через исходные числа  $m$ ,  $n$  и  $r$ .

Необходимо отметить, что пространство столбцов матрицы  $A$  не совпадает с пространством столбцов матрицы  $U$ . Элементарные операции оставляют неизменными пространство строк и нуль-пространство матрицы  $A$ , а пространство столбцов совершенно изменяется; сравните, например, столбцы матриц

$$A = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 5 \\ -1 & -3 & 3 & 0 \end{bmatrix} \quad \text{и} \quad U = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Тем не менее всякий раз, когда некоторые столбцы матрицы  $U$  образуют базис в пространстве столбцов матрицы  $U$ , соответствующие столбцы матрицы  $A$  образуют базис в пространстве столбцов матрицы  $A$ . Объясним, почему это так: дело в том, что  $Ax=0$  тогда и только тогда, когда  $Ux=0$ . Две эти системы эквивалентны и имеют одно и то же множество решений. Вспоминая правило умножения матрицы на вектор, видим, что равенство  $Ax=0$  выражает линейную зависимость между столбцами матрицы  $A$ , с компонентами вектора  $x$  в качестве весов. Поэтому каждая такая зависимость эквивалентна линейной зависимости  $Ux=0$  между столбцами матрицы  $U$  с теми же самыми весами. Следовательно, если некоторое множество столбцов матрицы  $A$  линейно независимо, то соответствующие столбцы матрицы  $U$  также линейно независимы и наоборот<sup>1</sup>). В обеих матрицах  $A$  и  $U$  последний столбец равен сумме первого и умноженного на  $1/3$  третьего столбцов, а второй столбец равен утроенному первому.

Теперь, чтобы найти базис пространства  $\mathfrak{R}(A)$ , мы должны решить более простую задачу нахождения базиса пространства столбцов матрицы  $U$ . Это уже было сделано (стр. 79), и вывод был таков:  $r$  столбцов матрицы  $U$ , содержащие ненулевые ведущие элементы, образуют базис пространства столбцов матрицы  $U$ . Мы преобразуем этот результат применительно к матрице  $A$ :

**20.** *Размерность пространства столбцов  $\mathfrak{R}(A)$  равна рангу  $r$ , который также равен размерности пространства строк, т. е. число линейно независимых столбцов равно числу линейно независимых строк. Базис пространства  $\mathfrak{R}(A)$  образуется теми  $r$  столбцами, которые соответствуют столбцам матрицы  $U$  с ненулевыми ведущими элементами.*

<sup>1</sup>) Я думаю, что это наиболее тонкое из доказательств, появляющихся до сих пор в книге. К счастью, оно не бесполезно: заключение 20, к которому оно приводит, является также наиболее тонким и наиболее важным среди всех, которые были сделаны до сих пор.

Тот факт, что пространство строк и пространство столбцов имеют одинаковую размерность  $r$ , является одной из наиболее важных теорем линейной алгебры. Это часто записывают сокращенно в виде: *ранг по строкам = ранг по столбцам*. Эта фраза выражает результат, который, например, для матрицы размера  $10 \times 12$  совсем не очевиден.

Чтобы убедиться еще раз, что пространство столбцов матрицы  $U$  имеет размерность  $r$ , рассмотрим типичную ситуацию, когда ранг  $r$  равен 3. Ступенчатая матрица  $U$  имеет три линейно независимые строки:

$$U = \begin{bmatrix} d_1 & * & * & * & * & * \\ \hline 0 & 0 & 0 & d_2 & * & * \\ 0 & 0 & 0 & 0 & 0 & d_3 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Мы утверждаем, что она также имеет только три линейно независимых столбца. Столбцы имеют только три ненулевые компоненты. Поэтому, если мы сможем показать, что три базисных столбца — первый, четвертый и шестой — линейно независимы, то они должны быть базисом (для пространства столбцов матрицы  $U$ , но не  $A$ !). Предположим, что некоторая линейная комбинация этих столбцов равна нулю:

$$c_1 \begin{bmatrix} d_1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} * \\ d_2 \\ 0 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} * \\ * \\ d_3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Двигаясь снизу вверх, как обычно, получаем, что  $c_3 = 0$ , потому что  $d_3 \neq 0$ ; далее,  $c_2 = 0$ , потому что  $c_3 = 0$  и  $d_2 \neq 0$ , и, наконец,  $c_1 = 0$ . Это устанавливает линейную независимость базисных столбцов и завершает доказательство. Так как  $Ax = 0$  тогда и только тогда, когда  $Ux = 0$ , мы заключаем, что первый, четвертый и шестой столбцы матрицы  $A$  — какой бы она ни была (мы ее в этом примере совсем не знаем) — образуют базис в пространстве  $\mathfrak{N}(A)$ .

Перейдем к четвертому основному подпространству, которое до сих пор оставалось вне нашего поля зрения. Так как первые три подпространства обозначались через  $\mathfrak{N}(A^T)$ ,  $\mathfrak{N}(A)$  и  $\mathfrak{R}(A)$ , то нетрудно догадаться, что четвертое подпространство следует обозначить через  $\mathfrak{N}(A)$ .

**4. Нуль-пространство матрицы  $A^T$ .** Нуль-пространство матрицы  $A^T$  является подпространством  $m$ -мерного пространства  $\mathbf{R}^m$ , состоящим из тех векторов  $y \in \mathbf{R}^m$ , для которых выполняется равенство  $A^T y = 0$ . Таким образом, линейная комбинация столбцов матрицы  $A^T$  с весами  $y_1, \dots, y_m$  должна равняться нулевому столбцу. Так как столбцы матрицы  $A^T$  являются строками матрицы  $A$ , то мы можем преобразовать систему  $A^T y = 0$  в уравнение для вектор-строк:

$$\|y^T A = [y_1 \ \dots \ y_m] \begin{bmatrix} \\ \\ \\ A \\ \\ \\ \end{bmatrix} = [0 \ \dots \ 0].$$

Такая вектор-строка  $y^T$  иногда называется *левым нуль-вектором* матрицы  $A$ . Линейная комбинация строк матрицы  $A$  с весами  $y_1, \dots, y_m$  равна нулевой строке. Размерность подпространства  $\mathfrak{N}(A^T)$  найти легко. Для любой матрицы число базисных переменных плюс число свободных переменных равно числу столбцов этой матрицы. Другими словами,

$$\begin{aligned} \text{ранг} + \text{дефект} &= \text{размерность пространства столбцов} + \\ &+ \text{размерность нуль-пространства} = \\ &= \text{число столбцов}. \end{aligned} \quad (10)$$

Это правило применимо и к матрице  $A^T$ , имеющей  $m$  столбцов, которая ничем не хуже, чем матрица  $A$ . Но ранг по строкам равен рангу по столбцам, т. е.  $r$  и, следовательно,

$$r + \dim \mathfrak{N}(A^T) = m. \quad (11)$$

**2Р.** *Левое нуль-пространство  $\mathfrak{N}(A^T)$  имеет размерность  $m - r$ .*

Для того чтобы найти базис пространства  $\mathfrak{N}(A^T)$ , начнем с представления  $PA = LU$  или  $L^{-1}PA = U$ . Последние  $m - r$  строк матрицы  $U$  нулевые, и, следовательно, последние  $m - r$  строк матрицы  $L^{-1}P$  образуют базис левого нуль-пространства. Когда они умножаются на  $A$ , то получается нулевой вектор.

Теперь, когда мы знаем размерности всех четырех рассматриваемых подпространств, можно подвести итог. Не будет преувеличением назвать этот результат так:

**Основная теорема линейной алгебры, часть 1.**

1.  $\mathfrak{N}(A^T)$  — пространство строк матрицы  $A$ ; размерность  $r$ ,
2.  $\mathfrak{N}(A)$  — нуль-пространство матрицы  $A$ ; размерность  $n - r$ ,
3.  $\mathfrak{N}(A)$  — пространство столбцов матрицы  $A$ ; размерность  $r$ ,
4.  $\mathfrak{N}(A^T)$  — левое нуль-пространство матрицы  $A$ ; размерность  $m - r$ .

**Упражнение 2.4.1.** Верно ли такое утверждение: если  $m=n$ , то нуль-пространство матрицы  $A$  равно ее левому нуль-пространству?

**Упражнение 2.4.2.** Найти размерность и построить базисы четырех подпространств, связанных с матрицей из примера 2.2.4:

$$A = \begin{bmatrix} 0 & 1 & 4 & 0 \\ 0 & 2 & 8 & 0 \end{bmatrix}.$$

**Упражнение 2.4.3.** Найти размерность и базисы четырех основных подпространств для матрицы

$$A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix}.$$

**Упражнение 2.4.4.** Описать четыре подпространства, связанные с матрицей

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

**Упражнение 2.4.5.** Показать, что если произведение двух матриц  $A$  и  $B$  равно нулевой матрице ( $AB=0$ ), то пространство столбцов матрицы  $B$  содержится в нуль-пространстве матрицы  $A$ .

**Упражнение 2.4.6.** Объяснить, почему  $Ax=b$  тогда и только тогда, когда  $(\text{ранг } A) = (\text{ранг } A')$ , где матрица  $A'$  получена из матрицы  $A$  добавлением вектора  $b$  в качестве дополнительного столбца. Указание: ранг матрицы  $A$  равен размерности ее пространства столбцов, а система  $Ax=b$  разрешима тогда и только тогда, когда вектор  $b$  лежит в подпространстве  $\mathfrak{R}(A)$ .

### МАТРИЦЫ РАНГА ОДИН

Одна из основных задач в математике — показать, как что-нибудь сложное может быть составлено из более простых частей. Мы уже рассмотрели один пример такого синтеза, когда нижняя треугольная матрица была разложена в произведение элементарных матриц. Теперь мы вернемся к рангу  $r$  как к новому критерию простоты и введем класс матриц *ранга*  $r=1$ . Следующая матрица является типичной матрицей этого класса:

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 2 & 2 \\ 8 & 4 & 4 \\ -2 & -1 & -1 \end{bmatrix}.$$

Каждая строка этой матрицы кратна первой строке, и потому пространство ее строк одномерно. На самом деле эту матрицу мы можем записать в виде *произведения вектор-столбца и век-*

тор-строки:

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 2 & 2 \\ 8 & 4 & 4 \\ -2 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \\ -1 \end{bmatrix} [2 \ 1 \ 1].$$

Произведение матрицы размера  $4 \times 1$  на матрицу размера  $1 \times 3$  будет матрицей размера  $4 \times 3$ . И эта матрица имеет ранг, равный единице. Заметим, что в то же самое время все столбцы этой матрицы кратны друг другу, т. е. пространство столбцов также имеет размерность, равную единице, и сводится к прямой.

То же самое происходит с любой другой матрицей ранга один: она может быть представлена в простейшей форме  $A = uv^T$ . Все строки этой матрицы кратны одному и тому же вектору  $v^T$ , а все столбцы кратны вектору  $u$ .

В последнем параграфе мы покажем, как разложить любую матрицу ранга  $r$  в сумму  $r$  матриц ранга один.

**Упражнение 2.4.7.** Пусть заданы числа  $a$ ,  $b$  и  $c$  и  $a \neq 0$ . Как надо выбрать  $d$ , чтобы матрица

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

имела ранг, равный единице? Выбрав  $d$ , разложить матрицу  $A$  в произведение  $uv^T$ .

**Упражнение 2.4.8.** Вычислить произведение  $AB$  матриц

$$A = \begin{bmatrix} 2 & -2 \\ 4 & -4 \\ 0 & 0 \end{bmatrix} \quad \text{и} \quad B = \begin{bmatrix} 1 & 1 & 2 \\ 3 & 3 & 6 \end{bmatrix}$$

ранга один. Записывая  $A$  и  $B$  в виде  $uv^T$  и  $wz^T$ , проверить, что их произведение кратно матрице  $uz^T$  и что соответствующий множитель равен скалярному произведению  $v^T w$ .

**Упражнение 2.4.9.** Для предыдущей матрицы  $A$  изобразить пространство строк и нуль-пространство в плоскости  $xy$ .

## СУЩЕСТВОВАНИЕ ОБРАТНЫХ МАТРИЦ

Мы уже знаем из § 1.5, что если матрица  $A$  имеет левую обратную матрицу  $B$  ( $BA = I$ ) и правую обратную матрицу  $C$  ( $AC = I$ ), то  $B = B(AC) = (BA)C = C$ . Теперь с помощью понятия ранга матрицы можно легко установить, какие матрицы имеют обратные. Строго говоря, обратная матрица существует только тогда, когда ранг исходной матрицы имеет максимальное из возможных значений.

Ранг матрицы всегда удовлетворяет неравенствам  $r \leq m$  и  $r \leq n$ , так как матрица размера  $m \times n$  не может иметь более чем  $m$  линейно независимых столбцов. Мы хотим доказать, что если  $r = m$ , то существует правая обратная матрица, а если  $r = n$ , то существует левая обратная матрица. В первом случае система  $Ax = b$  всегда имеет решение, а во втором случае решение (если оно существует) единственно. Только для квадратной матрицы могут одновременно выполняться оба равенства  $r = m$  и  $r = n$ . Поэтому только для систем с квадратными матрицами могут одновременно выполняться как условие существования решения, так и условие единственности этого решения.

**2Q. Существование:** система  $Ax = b$  имеет **по крайней мере** одно решение  $x$  для каждого вектора  $b$  тогда и только тогда, когда столбцы матрицы  $A$  порождают все пространство  $\mathbb{R}^m$ , т. е.  $r = m$ . В этом случае существует правая обратная матрица  $C$  размера  $n \times m$ , такая, что  $AC = I_m$  — единичная матрица порядка  $m$ , причем это возможно только тогда, когда  $m \leq n$ .

**Единственность:** система  $Ax = b$  имеет **не более** одного решения  $x$  для каждого вектора  $b$  тогда и только тогда, когда все столбцы матрицы  $A$  линейно независимы, т. е.  $r = n$ . В этом случае существует левая обратная матрица  $B$  размера  $n \times m$ , такая, что  $BA = I_n$  — единичная матрица порядка  $n$ , причем это возможно только тогда, когда  $m \geq n$ .

В первом случае  $x = Cb$  — это одно из возможных решений, так как  $Ax = ACb = b$ . Но могут быть и другие решения, если существуют другие правые обратные матрицы.

Во втором случае, если существует решение системы  $Ax = b$ , обязательно  $x = BAx = Bb$ , и других решений быть не может<sup>1)</sup>.

**Пример.** Рассмотрим простую матрицу размера  $2 \times 3$  и ранга  $r = 2$

$$A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \end{bmatrix}.$$

Так как  $r = m = 2$ , то предыдущая теорема гарантирует существование правой обратной матрицы  $C$ :

$$AC = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

<sup>1)</sup> Число решений для случая «единственности» равно либо нулю, либо единице, в то время как для случая «существования» — либо единице, либо бесконечности.

В действительности здесь существует много правых обратных матриц, потому что последняя строка матрицы  $C$  совершенно произвольна. Это случай существования, но не единственности:

$$C = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{5} \\ c_{31} & c_{32} \end{bmatrix}.$$

В качестве примера противоположной ситуации рассмотрим матрицу  $A^T$ , т. е. транспонированную матрицу к матрице  $A$ . Матрица  $A^T$  будет иметь размер  $3 \times 2$ , а ее ранг по-прежнему будет равен двум, поэтому она будет иметь бесконечно много левых обратных

$$BA^T = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 5 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Теперь последний столбец матрицы  $B$  совершенно произволен. На самом деле это матрицы, транспонированные к найденным выше матрицам  $C$ .

Этот пример подсказывает общий способ построения матриц  $C$  и  $B$  и доказательства утверждения 2Q. Мы выделим два случая:

(1) **Существование.** Пусть столбцы матрицы  $A$  порождают все пространство  $\mathbb{R}^m$ . Тогда ранг матрицы  $A$  (размерность ее пространства столбцов) равен  $m$ . Значит, каждый вектор  $b$ , включая координатные векторы  $e_1, \dots, e_m$ , может быть представлен в виде линейной комбинации столбцов матрицы  $A$ . Другими словами, мы можем найти по крайней мере одно решение  $x_i$  системы  $Ax_i = e_i$  для каждого  $i = 1, \dots, m$ . Если  $C$  — матрица размера  $n \times m$ , столбцами которой являются эти решения  $x_1, \dots, \dots, x_m$ , то в «постолбцовой» записи

$$AC = A[x_1 \ \dots \ x_m] = [e_1 \ \dots \ e_m] = I_m.$$

Отсюда следует, что матрица  $C$  является искомой правой обратной для матрицы  $A$ .

(2) **Единственность.** Пусть  $n$  столбцов матрицы  $A$  линейно независимы. Тогда ранг пространства столбцов равен  $n$  и нет свободных переменных, т. е. нуль-пространство содержит только нулевой вектор. Поэтому если существует некоторое решение системы  $Ax = b$ , то оно единственно. Простейший способ построения левой обратной матрицы заключается в рассмотрении матрицы  $A^T$ . Это будет матрица размера  $n \times m$ , имеющая, согласно основной теореме, тот же самый ранг, что и матрица  $A$ , т. е. ранг  $r = n$ , где  $n$  теперь равно числу строк матрицы  $A^T$ .

Следовательно, к матрице  $A^T$  применима первая часть доказываемой теоремы, т. е. для  $A^T$  существует правая матрица, скажем матрица  $Q$ , такая, что  $A^T Q = I_n$ . Транспонируя это равенство, получаем  $Q^T A = I_n$ . Итак,  $Q^T$  — требуемая левая обратная матрица для матрицы  $B$ .

Есть более простой способ построения этих обратных матриц. Пусть

$$B = (A^T A)^{-1} A^T \quad \text{и} \quad C = A^T (A A^T)^{-1}.$$

Очевидно, что  $BA = I$  и  $AC = I$ , но не столь очевидно существование матриц  $(A^T A)^{-1}$  и  $(A A^T)^{-1}$ . Здесь должны сыграть свою роль условия на пространство столбцов. Позже в §3 мы покажем, что матрица  $A^T A$  обратима, когда ранг матрицы  $A$  равен  $n$ , а матрица  $A A^T$  обратима, когда ранг матрицы  $A$  равен  $m$ .

Естественно представлять себе матрицу  $A$  как отображение из  $\mathbf{R}^n$  в  $\mathbf{R}^m$ : любой вектор  $x$  из  $\mathbf{R}^n$  преобразуется в некоторый вектор  $Ax$  из  $\mathbf{R}^m$ . Такое отображение *линейно*, поскольку из свойств умножения матриц следует, что  $A(cx + dy) = cAx + dAy$ . Это очень важное свойство, и мы посвятим приложение А связи между линейными отображениями и матрицами. Когда выполнено условие существования, т. е. когда  $r = m$ , это отображение называется отображением «на», потому что для каждого вектора  $b \in \mathbf{R}^m$  существует по крайней мере один вектор  $x \in \mathbf{R}^n$ , такой, что  $Ax = b$ . Образом (пространством столбцов) является все  $\mathbf{R}^m$ . В случае выполнения условия единственности, т. е. когда  $r = n$ , преобразование называется *взаимно однозначным*, потому что для каждого  $b$  существует не более одного такого вектора  $x \in \mathbf{R}^n$ , что  $Ax = b$ .

Нелинейный пример: функция  $y = x^2$  из  $\mathbf{R}^1$  в  $\mathbf{R}^1$  не является отображением «на», потому что для числа  $y = -4$  не существует такого  $x \in \mathbf{R}^1$ , что  $x^2 = -4$ , она не является также и взаимно однозначным отображением, потому что для числа  $y = 4$  существуют два числа  $x = -2$  и  $x = 2$ , таких, что  $x^2 = 4$ . Напротив, функция  $y = x^3$  является как отображением «на», так и взаимно однозначным отображением. В этом случае мы получаем взаимно однозначное соответствие между вещественными числами  $x$  и их кубами  $x^3$ , или, в другую сторону, взаимно однозначное соответствие между вещественными числами  $y$  и кубическими корнями  $y^{1/3}$  из них. Второе отображение является (двусторонним!) обратным к первому. Обратимые отображения — это именно взаимно однозначные соответствия; они являются одновременно взаимно однозначными отображениями и отображениями «на».

Прямоугольная матрица может обладать одним из указанных свойств и не обладать другим. Так же обстоит дело и с нелинейной функцией  $y(x)$ ; приведите соответствующие примеры в качестве упражнения. Для квадратных матриц ситуация со-

вершенно иная: если  $m = n$ , то матрица  $A$  обладает левой обратной тогда и только тогда, когда она обладает правой обратной, т. е. *существование влечет за собой единственность, а единственность влечет за собой существование*. Так как любая правая обратная матрица равна любой левой обратной, то существует лишь одна правая, она же левая, обратная, т. е.  $B = C = A^{-1}$ . Условие обратимости таково: ранг матрицы  $A$  должен быть максимальным, т. е.  $r = m = n$ . Мы можем выразить это и другим способом: для того чтобы квадратная матрица  $A$  порядка  $n$  была невырожденной, необходимо и достаточно выполнение любого из следующих условий:

(1) Столбцы матрицы  $A$  порождают все  $\mathbf{R}^n$ , и, следовательно, система  $Ax = b$  имеет по крайней мере одно решение для каждого вектора  $b$ .

(2) Столбцы матрицы  $A$  линейно независимы, и, следовательно, система  $Ax = 0$  имеет только одно решение  $x = 0$ .

Этот список можно значительно удлинить, особенно если взглянуть в последующие главы. Каждое условие в этом списке эквивалентно любому другому и влечет за собой невырожденность матрицы  $A$ .

(3) Строки матрицы  $A$  порождают все  $\mathbf{R}^n$ .

(4) Строки матрицы  $A$  линейно независимы.

(5) Метод исключения Гаусса с перестановкой строк, если это необходимо, обходится без нулевых ведущих элементов:  $PA = LDU$ , где все  $d_i \neq 0$ .

(6) Существует матрица  $A^{-1}$ , такая, что  $AA^{-1} = A^{-1}A = I^1$ .

(7) Определитель матрицы  $A$  отличен от нуля.

(8) Нуль не является собственным значением матрицы  $A$ .

(9) Матрица  $A^T A$  положительно определена.

(10) Матрица  $AA^T$  положительно определена.

Приведем типичный пример применения изложенного выше. Рассмотрим все многочлены  $P(t)$  степени  $n-1$ . Существует только один такой многочлен, принимающий нулевые значения в  $n$  данных точках  $t_1, \dots, t_n$ , — это  $P(t) \equiv 0$ . Никакой другой многочлен степени  $n-1$  не может иметь  $n$  корней. Это утверждение о единственности, и оно влечет за собой утверждение о существовании: для любых данных значений  $b_1, \dots, b_n$  существует многочлен степени  $n-1$ , принимающий эти значения:  $P(t_i) = b_i, i = 1, \dots, n$ . Заметим, что здесь мы имеем дело с квадратной матрицей, так как число коэффициентов в  $P(t)$  (которое равно  $n$ ) равно числу уравнений. Действительно, если  $P(t) = x_1 + x_2 t + \dots + x_n t^{n-1}$ , то уравнения  $P(t_i) = b_i$  могут быть

<sup>1)</sup> Это означает, что обратимость есть синоним невырожденности. Эти два термина применяются только к квадратным матрицам.

записаны в матричном виде:

$$\begin{bmatrix} 1 & t_1 & t_1^2 & \dots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \dots & t_2^{n-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & t_n & t_n^2 & \dots & t_n^{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix}.$$

Матрица коэффициентов  $A$  имеет порядок  $n$  и известна как *матрица Вандермонда*. Напомним доказательство сформулированного выше утверждения. Так как система  $Ax=0$  имеет только нулевое решение  $x=0$  (другими словами,  $P(t_i)=0$ ,  $i=1, \dots, n$ , только тогда, когда  $P \equiv 0$ ), то отсюда следует, что матрица  $A$  невырождена. Поэтому система  $Ax=b$  всегда имеет решение — многочлен степени  $n-1$  может быть проведен через любые  $n$  значений  $b_i$  в различных точках  $t_i$ . Позже мы вычислим определитель этой матрицы, и он действительно будет отличен от нуля.

Имеется и еще один пример. Так как матрица  $A$  невырождена, то невырожденной является и матрица  $A^T$ , потому что она имеет тот же самый ранг. Поэтому мы можем решить не только систему  $Ax=b$ , но также и систему  $A^T y=c$  для всех векторов  $c$ . Последняя система возникает в совершенно иной ситуации. Предположим, что мы ищем формулу для приближенного численного интегрирования вида

$$\int_0^1 f(t) dt \approx \sum_1^n y_i f(t_i).$$

Интеграл от функции  $f$  должен быть вычислен приближенно через свои значения  $f(t_i)$  в  $n$  различных точках  $t_i$ . Имея в распоряжении только  $n$  коэффициентов  $y_i$ , нельзя сделать эту формулу точной для всех функций  $f$ , но можно (и это делается практически) сделать ее точной для всех многочленов  $f=P$  степени не выше  $n-1$ . Требуя точность формулы для  $f=1$ ,  $f=t$ ,  $\dots$ ,  $f=t^{n-1}$ , мы получаем систему

$$\begin{aligned} \int_0^1 1 dt &= y_1 + \dots + y_n, \\ \int_0^1 t dt &= y_1 t_1 + \dots + y_n t_n, \\ &\vdots \\ \int_0^1 t^{n-1} dt &= y_1 t_1^{n-1} + \dots + y_n t_n^{n-1}. \end{aligned}$$

Это в точности система  $s = A^T y$ , которая, как было доказано, разрешима, и поэтому коэффициенты  $y_i$ ,  $i = 1, \dots, n$ , можно найти. Такая формула интегрирования существует, потому что многочлены степени  $n - 1$  имеют только  $n - 1$  корней.

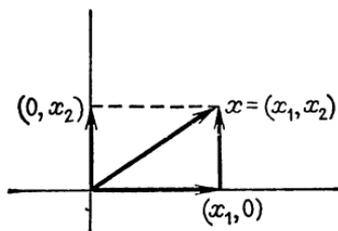
**Упражнение 2.4.10.** Построить нелинейную функцию  $y(x)$ , которая является взаимно однозначным отображением, но не является отображением «на» и нелинейную функцию  $z(x)$ , которая является отображением «на», но не является взаимно однозначной.

**Упражнение 2.4.11.** Объяснить, почему для системы с матрицей  $A$  условие существования выполнено тогда и только тогда, когда для системы с матрицей  $A^T$  выполнено условие единственности, и наоборот.

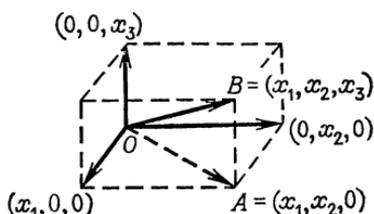
**Упражнение 2.4.12.** Построить все возможные левые и правые обратные матрицы для матрицы  $A = [1 \ 1 \ 0]$  размера  $1 \times 3$  и для матрицы  $A^T$ .

## § 2.5. ОРТОГОНАЛЬНОСТЬ ВЕКТОРОВ И ПОДПРОСТРАНСТВ

Мы начнем этот параграф с определения длины вектора. В двумерном случае эта длина  $\|x\|$  является длиной гипотенузы прямоугольного треугольника (рис. 2.4а), которая была вычис-



(a)



(b)

Рис. 2.4. Длина двумерных и трехмерных векторов.

лена много лет назад Пифагором<sup>1)</sup>:  $\|x\|^2 = x_1^2 + x_2^2$ . В трехмерном пространстве вектор  $x = (x_1, x_2, x_3)$  является диагональю параллелепипеда (рис. 2.4б), и его длина находится с помощью *двух* применений формулы Пифагора. Сначала рассматривается диагональ  $OA = (x_1, x_2, 0)$  в основании параллелепипеда, где по формуле Пифагора  $OA^2 = x_1^2 + x_2^2$ . Эта диагональ образует прямой угол с вертикальной стороной  $(0, 0, x_3)$ . Поэтому мы можем вновь воспользоваться формулой Пифагора (в плоскости  $OA$  и  $AB$ ).

<sup>1)</sup> Или, возможно, египтянами, но настоящее доказательство принадлежит пифагорейцам.

Гипотенуза прямоугольного треугольника  $OAB$  имеет длину  $\|x\|$ , равную  $\sqrt{OA^2 + AB^2}$ , т. е.

$$\|x\|^2 = OA^2 + AB^2 = x_1^2 + x_2^2 + x_3^2.$$

Обобщение понятия длины для  $n$ -мерного вектора  $x = (x_1, \dots, x_n)$  очевидно. **Длиной  $\|x\|$  вектора  $x$  из  $R^n$  называется положительный квадратный корень из суммы квадратов его компонент, т. е.**

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}. \quad (12)$$

Геометрическое доказательство этого факта состоит в применении формулы Пифагора  $n-1$  раз, добавляя каждый раз по одной компоненте. В случае размерности единица ( $n=1$ ) длина вектора равна абсолютной величине единственной компоненты  $x_1$ .

Предположим, что заданы два вектора  $x$  и  $y$  (рис. 2.5). Каким образом можно определить, будут ли эти векторы перпендику-

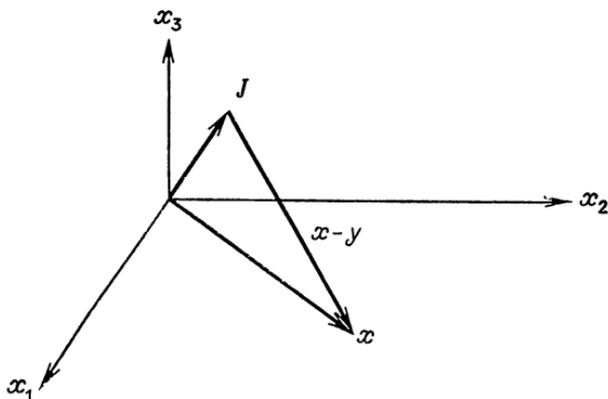


Рис. 2.5. Плоскость треугольника со сторонами  $x$ ,  $y$  и  $x-y$ .

лярными? Иначе говоря, что является критерием ортогональности? В случае двумерной плоскости ответ на этот вопрос может быть дан с помощью тригонометрии. Нам же нужно обобщение на  $n$ -мерный случай, но и тогда мы можем действовать в плоскости, порождаемой векторами  $x$  и  $y$ . На этой плоскости вектор  $x$  ортогонален к вектору  $y$ , если они образуют прямоугольный треугольник, и, следовательно, мы можем использовать в качестве критерия формулу Пифагора

$$\|x\|^2 + \|y\|^2 = \|x-y\|^2. \quad (13)$$

Используя это условие совместно с формулой (12), получаем, что

$$(x_1^2 + \dots + x_n^2) + (y_1^2 + \dots + y_n^2) = (x_1 - y_1)^2 + \dots + (x_n - y_n)^2.$$

Правая часть последнего равенства в точности равна выражению

$$(x_1^2 + \dots + x_n^2) - 2(x_1y_1 + \dots + x_ny_n) + (y_1^2 + \dots + y_n^2).$$

Поэтому условие (13) выполняется, а значит, векторы  $x$  и  $y$  ортогональны, когда сумма «перекрестных членов» равна нулю, т. е.

$$x_1y_1 + \dots + x_ny_n = 0. \quad (14)$$

Заметим, что эта величина в точности совпадает с произведением вектор-строки  $x^T$  на вектор-столбец  $y$  (как матриц):

$$x^T y = [x_1 \dots x_n] \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = x_1y_1 + \dots + x_ny_n. \quad (15)$$

Используя сокращенное обозначение для суммы, правую часть (15) можно записать как  $\sum x_i y_i$ . Это выражение появляется при обсуждении любого вопроса, связанного с геометрией  $n$ -мерного пространства. Оно называется скалярным произведением двух векторов и обозначается символом  $(x, y)$  или  $x \cdot y$ , но мы предпочитаем сохранить обозначение  $x^T y$ :

**2R.** Величина  $x^T y$  является скалярным произведением векторов (столбцов)  $x$  и  $y$  из  $\mathbf{R}^n$ . Она равна нулю тогда и только тогда, когда векторы  $x$  и  $y$  ортогональны.

Отметим, что понятия длины и скалярного произведения связаны между собой следующим образом:  $x^T x = x_1^2 + \dots + x_n^2 = \|x\|^2$ . Существует только один вектор длины нуль, другими словами, только один вектор ортогонален сам себе — это нулевой вектор. Нулевой вектор ортогонален также каждому вектору  $y \in \mathbf{R}^n$ .

**Упражнение 2.5.1.** Найти длины и скалярное произведение векторов  $x = (1, 4, 0, 2)^T$  и  $y = (2, -2, 1, 3)^T$ .

Мы показали, что векторы  $x$  и  $y$  ортогональны тогда и только тогда, когда их скалярное произведение равно нулю. В следующей главе скалярное произведение будет обсуждаться в большем объеме<sup>1)</sup>. Там мы будем интересоваться и неортогональными векторами, так как скалярное произведение дает естественное опре-

<sup>1)</sup> Точнее, мы будем рассматривать его под другим углом зрения.

деление косинуса в  $n$ -мерном пространстве и определяет угол между любыми двумя векторами. В этом параграфе, однако, основной целью по-прежнему остается изучение четырех основных подпространств, причем свойство, которое нас интересует — ортогональность.

Прежде всего, между линейной независимостью и ортогональностью существует простая связь: *если ненулевые векторы  $v_1, \dots, v_k$  взаимно ортогональны (каждый вектор ортогонален любому другому), то они линейно независимы.*

**Доказательство.** Предположим, что  $c_1 v_1 + \dots + c_k v_k = 0$ . Чтобы показать, что любой коэффициент  $c_i$  должен равняться нулю, возьмем скалярное произведение обеих частей на вектор  $v_i$ :

$$v_i^T (c_1 v_1 + \dots + c_k v_k) = v_i^T 0 = 0. \quad (16)$$

Из взаимной ортогональности векторов  $v_i$  следует, что в левой части (16) остается только один член:  $c_i v_i^T v_i$ . Так как векторы  $v_i$  по предположению ненулевые, то  $v_i^T v_i \neq 0$  и, следовательно,  $c_i = 0$ . Таким образом, существует только одна линейная комбинация векторов  $v_i$ ,  $i = 1, \dots, k$ , равная нулю, — это тривиальная комбинация со всеми  $c_i = 0$ . Поэтому векторы  $v_1, \dots, v_k$  линейно независимы.

Наиболее важным примером взаимно ортогональных векторов является множество координатных векторов  $e_1, \dots, e_n$  в  $\mathbf{R}^n$ . Они представляют собой столбцы единичной матрицы и составляют простейший базис пространства  $\mathbf{R}^n$ . Кроме того, они являются *единичными* векторами, т. е. длина каждого вектора  $e_i$ ,  $i = 1, \dots, n$ , равна единице,  $\|e_i\| = 1$ . Геометрически эти векторы указывают направления координатных осей. Если систему векторов  $e_1, \dots, e_n$  повернуть, то мы получим новое ортонормированное множество векторов, т. е. новую систему взаимно ортогональных единичных векторов. В плоскости такое вращение дает ортонормированную пару векторов

$$v_1 = (\cos \theta, \sin \theta), \quad v_2 = (-\sin \theta, \cos \theta).$$

**Упражнение 2.5.2.** Привести пример линейно независимых векторов из  $\mathbf{R}^2$ , не являющихся взаимно ортогональными, доказав тем самым, что утверждение, обратное предыдущей теореме, не верно. Привести также пример взаимно ортогональных векторов, которые не являются линейно независимыми, потому что нарушены условия теоремы.

**Упражнение 2.5.3.** Известно, что две прямые на плоскости перпендикулярны, когда произведение тангенсов углов их наклона равно  $-1$ . Применить это свойство к векторам  $x = (x_1, x_2)$  и  $y = (y_1, y_2)$ , у которых тангенсы углов наклона соответственно равны  $x_2/x_1$  и  $y_2/y_1$ , для того чтобы вновь вывести условие ортогональности  $x^T y = 0$ .

**Упражнение 2.5.4.** Почему  $i$ -я строка матрицы  $B$  должна быть ортогональна  $j$ -му столбцу матрицы  $B^{-1}$ , если  $i \neq j$ ?

**Упражнение 2.5.5.** Какие пары из следующих трех векторов

$$v_1 = \begin{bmatrix} 1 \\ 2 \\ -2 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 4 \\ 0 \\ 4 \\ 0 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

ортогональны?

**Упражнение 2.5.6.** В  $\mathbb{R}^3$  найти все векторы, ортогональные как к вектору  $(1, 1, 1)$ , так и к вектору  $(1, -1, 0)$ . Построить, используя эти векторы, систему взаимно ортогональных единичных векторов (ортонормированную систему) в  $\mathbb{R}^3$ .

### ОРТОГОНАЛЬНЫЕ ПОДПРОСТРАНСТВА

Перейдем далее к понятию ортогональности двух подпространств. В обычном трехмерном пространстве все подпространства исчерпываются прямыми и плоскостями, проходящими через начало координат, и двумя «экстремальными» подпространствами — подпространством, состоящим только из одного нулевого вектора, и всем пространствам  $\mathbb{R}^3$ . Подпространство  $\{0\}$  ортогонально ко всем подпространствам. Прямая может быть ортогональна как к другой прямой, так и к плоскости, но плоскость не может быть ортогональна к плоскости<sup>1)</sup>, а все пространство  $\mathbb{R}^3$  ортогонально только к нулевому подпространству  $\{0\}$ .

В  $n$ -мерном случае основное определение таково:

**2S.** Два подпространства  $V$  и  $W$  одного и того же пространства  $\mathbb{R}^n$  называются **ортогональными**, если любой вектор  $v \in V$  ортогонален любому вектору  $w \in W$ , т. е.  $v^T w = 0$  для всех  $v \in V$  и  $w \in W$ .

**Пример.** Предположим, что  $V$  является плоскостью, порожденной векторами  $v_1 = (1, 0, 0, 0)$  и  $v_2 = (1, 1, 0, 0)$ , а  $W$  — прямая, порожденная вектором  $w_1 = (0, 0, 4, 5)$ . Тогда, так как вектор  $w_1$  ортогонален как к вектору  $v_1$ , так и к вектору  $v_2$ , прямая  $W$  будет ортогональна к плоскости  $V$ .

**Упражнение 2.5.7.** В предыдущем примере найти вектор  $w_2$ , такой, что плоскость  $W$ , порожденная векторами  $w_1$  и  $w_2$ , ортогональна плоскости  $V$ . Найти также  $v_3$ , такой, что трехмерное подпространство, порожденное векторами  $v_1, v_2, v_3$ , ортогонально к исходной прямой  $W$ .

**Упражнение 2.5.8.** Показать, что если подпространства  $V$  и  $W$  ортогональны, то они имеют только один общий вектор, а именно нулевой вектор, т. е.  $V \cap W = \{0\}$ .

<sup>1)</sup> Я должен признаться, что фронтальная и боковая стены очень похожи на перпендикулярные плоскости в  $\mathbb{R}^3$ . Но по нашему определению это не так! В этих плоскостях существуют прямые  $v$  и  $w$ , пересекающиеся не под прямым углом.

Объясним теперь, зачем нам понадобилась ортогональность подпространств. Напомним, что два из четырех основных подпространств, связанных с некоторой матрицей  $A$ , являются подпространствами пространства  $\mathbb{R}^n$  — нуль-пространство  $\mathfrak{N}(A)$  и пространство строк  $\mathfrak{R}(A^T)$ , а два других лежат в  $\mathbb{R}^m$ . Наиболее важный факт об этих пространствах, помимо их размерности,— это то, что они ортогональны:

2Т. Для любой матрицы размера  $m \times n$  нуль-пространство  $\mathfrak{N}(A)$  и пространство строк  $\mathfrak{R}(A^T)$  являются ортогональными подпространствами в  $\mathbb{R}^n$ . Аналогично левое нуль-пространство  $\mathfrak{N}(A^T)$  и пространство столбцов  $\mathfrak{R}(A)$  являются ортогональными подпространствами в  $\mathbb{R}^m$ .

Первое доказательство. Предположим, что  $w$  — произвольный вектор из нуль-пространства  $\mathfrak{N}(A)$ . Тогда  $Aw = 0$ , причем эта система  $m$  уравнений может быть записана более подробно:

$$Aw = \begin{bmatrix} \text{строка 1} & \dots \\ \text{строка 2} & \dots \\ & \cdot \\ & \cdot \\ & \cdot \\ \text{строка } m & \dots \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ \cdot \\ w_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}. \quad (17)$$

Первое уравнение означает, что вектор  $w$  ортогонален к первой строке матрицы  $A$  или, более правильно (сохраняя соглашение понимать под вектором вектор-столбец), к первому столбцу матрицы  $A^T$ . Второе уравнение утверждает ортогональность вектора  $w$  ко второму столбцу матрицы  $A^T$ . Продолжая аналогичные рассуждения, установим, что вектор  $w$  ортогонален к каждому столбцу матрицы  $A^T$ . Следовательно, он ортогонален ко всему пространству, порожденному этими столбцами, или, иначе говоря, к каждому вектору  $v$  из пространства столбцов матрицы  $A^T$ . Это верно для любого вектора  $w$  из нуль-пространства, и, следовательно,  $\mathfrak{N}(A) \perp \mathfrak{R}(A^T)$ . Второе утверждение в теореме об ортогональности  $\mathfrak{N}(A^T) \perp \mathfrak{R}(A)$  совпадает с первым, если его применить к матрице  $A^T$ . (Первое утверждение было доказано для любых матриц, включая и транспонированную к любой заданной матрице  $A$ .) С другой стороны, мы можем начать с любого вектора  $u$  из левого нуль-пространства, и, используя

соотношение

$$y^T A = [y_1 \ \dots \ y_m] \begin{bmatrix} с & с \\ т & т \\ о & о \\ л & л \\ б & б \\ е & е \\ ц & ц \\ 1 & n \end{bmatrix} = [0 \ \dots \ 0],$$

легко убедиться, что вектор  $y$  ортогонален ко всем столбцам матрицы  $A$ . Следовательно, он ортогонален к каждой линейной комбинации столбцов, т. е. любой вектор  $y \in \mathfrak{N}(A^T)$  ортогонален к каждому вектору  $w \in \mathfrak{N}(A)$ .

Второе доказательство. Мы хотим установить тот же самый результат с помощью рассуждений, не столь тесно связанных с координатами. Сравнение этих двух доказательств будет полезно читателю как пример «абстрактного» метода исследования в противовес «конкретному». Я надеюсь, что это доказательство яснее и лучше запоминается.

Предположим, что  $w \in \mathfrak{N}(A)$ , а  $v \in \mathfrak{N}(A^T)$ . Тогда  $Aw = 0$ , а вектор  $v$  имеет представление  $v = A^T x$  для некоторого вектора  $x$ . (Иначе говоря, вектор  $v$  является линейной комбинацией столбцов матрицы  $A^T$ , а  $x_1, \dots, x_m$  — веса этой комбинации.) Следовательно,

$$w^T v = w^T (A^T x) = (w^T A^T) x = (Aw)^T x = 0^T x = 0. \quad (18)$$

**Пример.** Пусть  $A$  — матрица из примера 2.4.2:

$$A = \begin{bmatrix} 0 & 1 & 4 & 0 \\ 0 & 2 & 8 & 0 \end{bmatrix} \rightarrow U = \begin{bmatrix} 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Второй столбец является базисным, и потому вторая компонента является базисной, а три другие компоненты — свободными. Тогда если мы последовательно присвоим каждой свободной переменной значение, равное единице, и решим систему  $Ux = 0$ , то мы найдем три вектора, образующие базис нуль-пространства матрицы  $A$ :

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ -4 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Все три перпендикулярны к строкам матрицы  $A$ , как это и гарантирует утверждение 2Т.

Пространство столбцов матрицы  $A$  одномерно (ранг по строкам равен рангу по столбцам) и порождается одним базисным столбцом  $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ . С другой стороны, левое нуль-пространство находится с помощью такой комбинации строк матрицы  $A$ , при которой получается нулевая строка в матрице  $U$ :  $(-2)$  (строка 1)  $+$   $(1)$  (строка 2)  $= 0$ . Следовательно, вектор  $y^T = (-2, 1)$  лежит в левом нуль-пространстве и, как и утверждалось, ортогонален к пространству столбцов:

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

Теперь я должен попросить вас потерпеть еще немного. Нуль-пространство, конечно, перпендикулярно к пространству строк. Но это еще не все:  $\mathfrak{N}(A)$  содержит все ортогональные к пространству строк векторы. Оно состоит из всех решений системы  $Ax = 0$ .

**2U. Определение.** Пусть  $V$  — подпространство пространства  $\mathbb{R}^n$ . Тогда пространство всех  $n$ -мерных векторов, ортогональных к подпространству  $V$ , называется **ортогональным дополнением к  $V$**  и обозначается символом  $V^\perp$ .

Используя эту терминологию, нуль-пространство  $\mathfrak{N}(A)$  можно назвать ортогональным дополнением к пространству  $\mathfrak{R}(A^T)$ , т. е.  $\mathfrak{N}(A) = (\mathfrak{R}(A^T))^\perp$ . В то же время верно и противоположное утверждение: пространство строк содержит все ортогональные к нуль-пространству векторы. Это не столь очевидно из построения, поскольку, решая систему  $Ax = 0$ , мы начинали с пространства строк и находили все векторы  $x$ , ортогональные к нему. Теперь мы идем в противоположном направлении. Предположим, однако, что некоторый вектор  $z \in \mathbb{R}^n$  ортогонален к нуль-пространству, но не содержится в пространстве строк. Тогда, добавляя вектор  $z$  в качестве строки к матрице  $A$ , мы тем самым расширим пространство строк, не изменяя нуль-пространство. Но мы знаем, что существует неизменная формула:  $r + (n - r) = n$  или

$\dim(\text{пространство строк}) + \dim(\text{нуль-пространство}) = \text{число столбцов}$ .

Так как последние два числа не изменяются при добавлении строки  $z$ , то и первое число также не может измениться. Поэтому каждый вектор, ортогональный к нуль-пространству, содержится в пространстве строк, т. е.  $\mathfrak{N}(A^T) = (\mathfrak{N}(A))^\perp$ .

Те же рассуждения, примененные к матрице  $A^T$ , дают двойственный результат: левое нуль-пространство  $\mathfrak{N}(A^T)$  и пространство столбцов  $\mathfrak{N}(A)$  являются ортогональными дополнениями друг к другу в  $\mathbf{R}^m$ . Это завершает вторую половину основной теоремы линейной алгебры. Первая половина дает размерности четырех подпространств, включая тот факт, что ранг по строкам равен рангу по столбцам, а теперь мы знаем, что они не только перпендикулярны, но и являются ортогональными дополнениями друг друга.

## 2V. Основная теорема линейной алгебры, часть 2.

$$\begin{aligned}\mathfrak{N}(A) &= (\mathfrak{N}(A^T))^\perp, & \mathfrak{N}(A^T) &= (\mathfrak{N}(A))^\perp, \\ \mathfrak{N}(A^T) &= (\mathfrak{N}(A))^\perp, & \mathfrak{N}(A) &= (\mathfrak{N}(A^T))^\perp.\end{aligned}$$

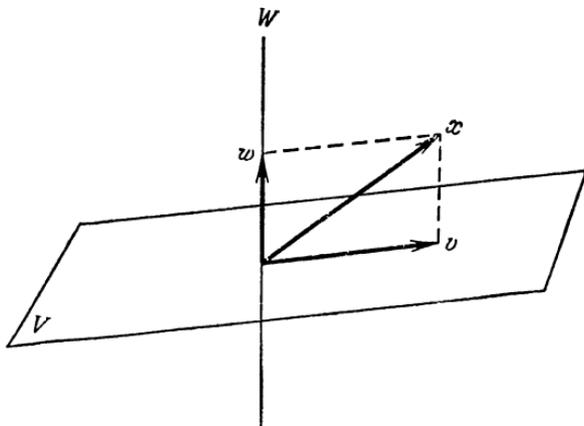
Последнее равенство означает, что система  $Ax=b$  имеет решение тогда и только тогда, когда вектор  $b$  ортогонален к  $\mathfrak{N}(A^T)$ , т. е. вектор  $b$  принадлежит пространству столбцов тогда и только тогда, когда он ортогонален к каждому решению системы  $A^T y=0$ .

Мы должны подчеркнуть, что два подпространства  $V$  и  $W$  могут быть ортогональными и не быть ортогональными дополнениями друг к другу. В трехмерном пространстве прямая  $V$ , порождаемая вектором  $(1, 0, 0)$ , ортогональна к прямой  $W$ , порождаемой вектором  $(0, 0, 1)$ , но  $V$  не совпадает с  $W^\perp$ . Ортогональным дополнением к  $W$  является двумерное подпространство, содержащее все векторы вида  $(x_1, x_2, 0)$ . Прямая  $V$  может быть только частью подпространства  $W^\perp$ , потому что ее размерность меньше размерности  $W^\perp$ . При соответствующих размерностях два ортогональных подпространства обязательно являются ортогональными дополнениями друг к другу, как было в случае пространства строк и нуль-пространства. С другой стороны, если  $W=V^\perp$ , то это указывает, что соотношение между размерностями  $V$  и  $W$  правильное и автоматически означает, что  $V=W^\perp$ . Пространство будет просто разложено на две взаимно перпендикулярные части  $V$  и  $W$ , как на рис. 2.6.

Теорема, которая соответствует рис. 2.6, имеет следующую формулировку.

Если  $V$  и  $W$  — подпространства пространства  $\mathbf{R}^n$  и если выполнено хотя бы одно из перечисленных ниже условий, то  $V$  и  $W$  являются ортогональными дополнениями друг к другу:

- (1)  $W=V^\perp$  ( $W$  состоит из всех векторов, ортогональных к  $V$ );
- (2)  $V=W^\perp$  ( $V$  состоит из всех векторов, ортогональных к  $W$ );
- (3)  $V$  и  $W$  ортогональны и  $\dim V + \dim W = n$ .

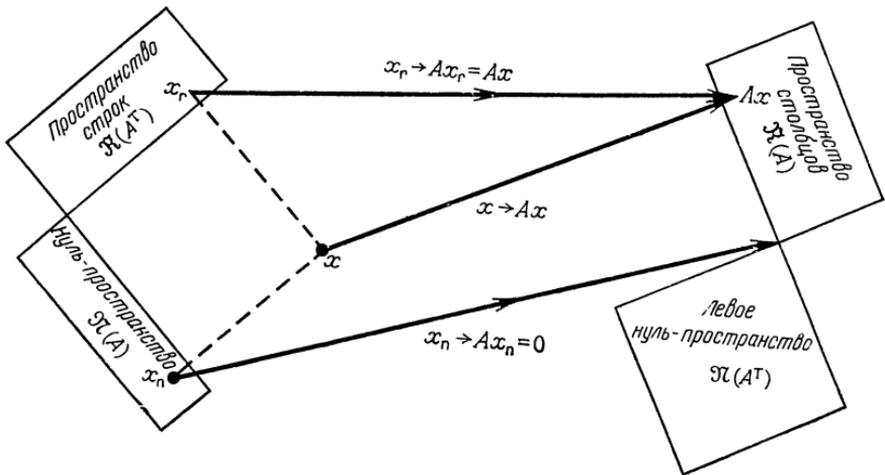
Рис. 2.6. Ортогональное разложение пространства  $\mathbf{R}^3$ .

Если выполняется хотя бы одно из этих трех эквивалентных условий, то любой вектор  $x$  может быть разложен единственным образом в сумму  $x = v + w$ , где  $w \in W$ , а  $v \in V$ . Составляющие этого разложения являются проекциями вектора  $x$  на подпространства  $V$  и  $W$  и взаимно ортогональны, т. е.  $v^T w = 0$ .

Разрешите подвести итог предыдущему и данному параграфам. Взяты вместе, они дают очень полную картину действия матрицы  $A$ . В предыдущем параграфе определяется *размерность* четырех основных подпространств (в частности, пространство строк  $\mathfrak{R}(A^T)$  и пространство столбцов  $\mathfrak{R}(A)$  имеют одинаковую размерность  $r$ ). В данном параграфе определяется *взаимное расположение* этих четырех пространств: два из них являются ортогональными дополнениями друг к другу в  $\mathbf{R}^n$ , а два других — ортогональными дополнениями друг к другу в  $\mathbf{R}^m$ . Действие произвольной матрицы  $A$  иллюстрируется (очень схематично) на рис. 2.7. Произвольный вектор  $x \in \mathbf{R}^n$  разлагается в сумму  $x = x_r + x_n$ . Матрица  $A$  преобразует компоненту  $x_r \in \mathfrak{R}(A^T)$  в вектор  $Ax_r = Ax$ , принадлежащий пространству столбцов матрицы  $A$ , в то время как компонента  $x_n \in \mathfrak{N}(A)$  преобразуется в нулевой вектор<sup>1)</sup>.

**2W.** *Отображение пространства строк в пространство столбцов невырожденно, т. е. обратимо; каждый вектор  $b$  из пространства столбцов является образом одного и только одного вектора  $x$ , принадлежащего пространству строк.*

<sup>1)</sup> Мы не знаем, как нарисовать два ортогональных подпространства размерности  $r$  и  $n - r$ . Если вы уже поняли смысл этих размерностей и понятия ортогональности, то рис. 2.7 не смутит вас.

Рис. 2.7. Действие матрицы  $A$ .

**Доказательство.** Если вектор  $b$  принадлежит пространству столбцов, то он является некоторой комбинацией  $Ax$  столбцов матрицы  $A$ . Раскладывая вектор  $x$  в сумму  $x_r + x_n$ , где  $x_r \in \mathfrak{R}(A^T)$ , а  $x_n \in \mathfrak{N}(A)$ , получаем, что  $Ax_r = Ax_r + Ax_n = Ax = b$ . Таким образом, мы нашли в пространстве строк подходящий вектор  $x_r$ . Если бы там был другой вектор  $x'_r$ , такой, что  $Ax'_r = b$ , то  $A(x_r - x'_r) = b - b = 0$ . Поэтому вектор  $x_r - x'_r$  принадлежит как пространству строк, так и нуль-пространству. Следовательно, он ортогонален сам себе и должен равняться нулевому вектору, т. е.  $x_r = x'_r$ .

Каждая матрица  $A$  обратима, если ее соответствующим образом воспринимать как отображение некоторого  $r$ -мерного подпространства на другое  $r$ -мерное подпространство, а именно  $\mathfrak{R}(A^T)$  на  $\mathfrak{R}(A)$ . На ортогональном дополнении  $\mathfrak{N}(A)$  матрица  $A$  действует как нулевая матрица. Точно так же матрица  $A^T$  является обратимым отображением в противоположном направлении, из  $\mathfrak{N}(A)$  на  $\mathfrak{R}(A^T)$ . Это не значит, что матрица  $A^T$  является обратной к  $A$ , просто для каждого  $y \in \mathfrak{R}(A^T)$  существует один и только один элемент  $x \in \mathfrak{N}(A)$ , такой, что  $y = A^T x$ . Фактически отображение  $A^T$  переводит вектор  $Ax$  в  $A^T Ax$ , в то время как обратное к  $A$  отображение или, в вырожденном случае, обсуждаемом в § 3.4, псевдообратное к  $A$  отображение переводит  $Ax$  обратно в  $x$  при  $x \in \mathfrak{R}(A)$ .

**Упражнение 2.5.9.** Найти ортогональное дополнение к плоскости, порожденной векторами  $(1, 1, 2)$  и  $(1, 2, 3)$ , выбирая их в качестве строк матрицы  $A$  и решая систему  $Ax = 0$ . Напомним, что это дополнение является прямой.

**Упражнение 2.5.10.** Построить однородное уравнение с тремя неизвестными, решениями которого являются линейные комбинации векторов  $(1, 1, 2)$

и (1, 2, 3). Эта задача обратна к задаче из предыдущего упражнения, но, разумеется, на самом деле они совпадают.

**Упражнение 2.5.11.** Основная теорема линейной алгебры часто формулируется в виде *альтернативы Фредгольма*: для любых  $A$  и  $b$  одна и только одна из следующих задач:

$$(1) Ax = b, \quad (2) A^T y = 0, \quad y^T b \neq 0,$$

имеет решение. Иначе говоря, либо вектор  $b$  лежит в пространстве  $\mathfrak{R}(A)$ , либо существует такой вектор  $y \in \mathfrak{R}(A^T)$ , что  $y^T b \neq 0$ . Получить подходящий вектор  $y = b_n$  можно, раскладывая вектор  $b$  в сумму  $b = b_r + b_n$ , где  $b_r \in \mathfrak{R}(A)$ ,  $b_n \in \mathfrak{R}(A^T)$ .

**Упражнение 2.5.12.** Найти базис нуль-пространства матрицы

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 4 \end{bmatrix}$$

и проверить, что он ортогонален к пространству строк. Разложить вектор  $x = (3, 3, 3)$  в сумму  $x = x_r + x_n$  двух векторов  $x_r \in \mathfrak{R}(A^T)$  и  $x_n \in \mathfrak{R}(A)$ .

**Упражнение 2.5.13.** Показать, что ортогональное дополнение к  $V + W$  равно пересечению ортогональных дополнений к  $V$  и  $W$ , т. е.  $(V + W)^\perp = V^\perp \cap W^\perp$ .

**Упражнение 2.5.14.** Проиллюстрировать действие матрицы  $A^T$  рисунком, аналогичным рис. 2.7, направляя  $\mathfrak{R}(A)$  обратно, в пространство строк  $\mathfrak{R}(A^T)$ , а левое нуль-пространство — в нулевой вектор.

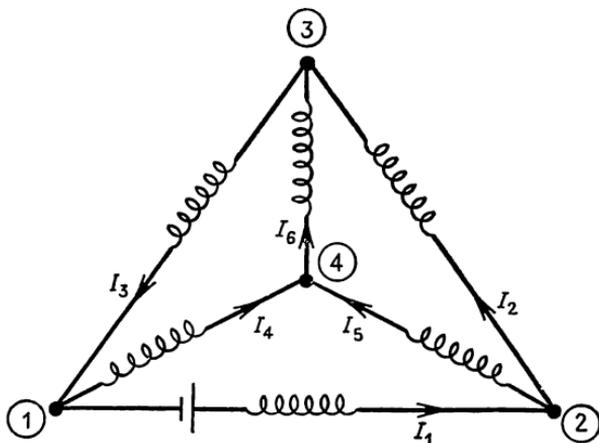


Рис. 2.8. Электрическая цепь с четырьмя вершинами и шестью ребрами.

### МАТРИЦЫ ИНЦИДЕНТНОСТИ И ЗАКОНЫ КИРХГОФА

Для контура, изображенного на рис. 2.8, по которому течет постоянный ток, справедливы законы Кирхгофа:

(1) В каждой вершине контура сумма входящих токов равна сумме выходящих токов. Например, в первой вершине  $I_3 = I_1 + I_4$ .

(2) Сумма падений напряжений вдоль любого замкнутого контура равна нулю. Если через  $E_k$  обозначить падение напряжения на ребре  $I_k$  в направлении, указанном стрелкой на рис. 2.8, то для замкнутого треугольника с вершинами 1—4—3 второй закон требует  $E_4 + E_6 + E_3 = 0$ .

Мы хотим выразить эти законы на языке матричной алгебры. Отметим, что они основываются только на «теории графов», т. е. зависят лишь от способа соединения вершин ребрами и от направлений стрелок, но не зависят от величин сопротивлений (или мощностей батарей) в контуре<sup>1</sup>). Связи между вершинами полностью описываются матрицей инцидентности графа этого контура, имеющей строку для каждой вершины и столбец для каждого ребра. В каждом столбце этой матрицы два ненулевых элемента  $+1$  и  $-1$  отмечают начальную и конечную вершины ребра. Матрица инцидентности, соответствующая рис. 2.8, имеет следующий вид:

$$M = \begin{bmatrix} 1 & 0 & -1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{bmatrix} \begin{array}{l} \text{вершина 1} \\ \text{вершина 2} \\ \text{вершина 3} \\ \text{вершина 4.} \end{array}$$

Рассмотрим первую строку матрицы  $M$ , соответствующую первой вершине. В ней имеется один элемент, равный  $-1$ , соответствующий входящему ребру  $I_3$ , и два элемента, равные  $+1$ , соответствующие выходящим ребрам  $I_1$  и  $I_4$ . Вспоминая, что первый закон Кирхгофа дает  $I_3 = I_1 + I_4$ , запишем его сразу в матричном виде: если через  $I$  обозначен вектор-столбец, составленный из шести значений тока  $I_1, \dots, I_6$ , то  $MI = 0$ . Это система четырех уравнений с шестью неизвестными.

Чтобы записать в матричном виде второй закон, зафиксируем в первой вершине потенциал  $p_1$ , равный 0, и определим потенциалы  $p_i$  во всех остальных вершинах с помощью условия, что падение напряжения между вершинами  $i$  и  $j$  (если они соединены ребром в этом направлении) равняется разности потенциалов, т. е.  $p_i - p_j$ . Второй закон утверждает, что суммарное падение напряжения вдоль замкнутого контура равно нулю, и, следовательно, мы вернемся опять к тому же самому значению потенциала, с которого начали. В терминах матрицы  $M$  вычисление падения напряжения  $p_i - p_j$  на некотором ребре соответствует

<sup>1</sup>) Другими словами, законы Кирхгофа являются «условиями равновесия», которые дополняются законом Ома  $E = IR$ .

умножению вектор-потенциала  $p$  на столбец матрицы  $M$ , соответствующий этому ребру, так как по построению этот столбец имеет компоненту  $+1$ , соответствующую вершине  $i$ , и  $-1$ , соответствующую вершине  $j$ . Другими словами, *вектор падения напряжения  $E$  равен вектору  $M^T p$* .

Теперь оба закона могут быть сформулированы очень коротко: *вектор  $I$  принадлежит нуль-пространству матрицы  $M$ , а вектор  $E$  принадлежит ее пространству строк*. Так как эти два пространства ортогональны для любой матрицы  $M$ , то векторы  $E$  и  $I$  должны быть перпендикулярны, т. е.  $E^T I = 0$ .

**Пример.** Предположим, что 20-вольтовая батарея, подключенная к первому ребру, дает ток  $I_1$ . Из соображений симметрии половина этого тока будет возвращаться через третью вершину, а половина — через четвертую, не проходя через шестое ребро, т. е.  $I_6 = 0$ . Если величина каждого из сопротивлений равна пяти ( $R = 5$ ), то можно проверить, что основной ток  $I_1 = 2$ , а возвратные токи  $I_2 = I_3 = -I_4 = I_5 = 1$ . (Замечание: стрелки были поставлены случайным образом только для того, чтобы зафиксировать направления вдоль каждого ребра, поэтому равенство  $I_4 = -1$  означает, что ток течет по этому ребру в направлении, противоположном стрелке.) Эти токи дают падение напряжения  $E = IR = 10V$  в нижнем сопротивлении и по  $5V$  в каждом из сопротивлений 2, 3, 4, 5, которые уравниваются 20-вольтовой батареей. Таким образом, законы Кирхгофа соблюдаются и, следовательно, величины токов указаны правильно.

**Упражнение 2.5.15.** Найти падение напряжения на каждом ребре и потенциалы  $p_2, p_3, p_4$ , если  $p_1 = 0$ . Проверить, что  $E^T I = 0$ .

**Упражнение 2.5.16.** Нарисовать контур, соответствующий матрице инцидентности

$$M = \begin{bmatrix} 1 & 0 & 0 & -1 & 1 & 0 \\ -1 & 1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 1 & 0 & -1 \end{bmatrix},$$

с 6-вольтовой батареей в ребре 1—3 и единичными сопротивлениями во всех ребрах. Каковы будут значения токов  $I$  и падений напряжения  $E$ ?

## § 2.6. ПАРЫ ПОДПРОСТРАНСТВ И ПРОИЗВЕДЕНИЯ МАТРИЦ

Эта глава преследует единственную цель — изучение системы  $Ax = b$ . Каждое новое понятие и каждое определение, включая векторные пространства и линейную независимость, базис и раз-

мерность, ранг и дефект, скалярное произведение и ортогональность, были введены потому, что это требовалось для одной этой цели. Сейчас мы хотим вернуться к этим понятиям, чтобы установить между ними некоторые другие связи, которых мы вначале не отметили<sup>1)</sup>.

Многие из них просты. Они возникают из рассмотрения не отдельного подпространства или отдельной матрицы, а взаимосвязей между двумя подпространствами или двумя различными матрицами. Вот первая из них:

**2X.** Если  $V$  и  $W$  — два подпространства некоторого векторного пространства, то их пересечение  $V \cap W$  также будет подпространством этого пространства.

Доказательство очевидно. Предположим, что  $x$  и  $x'$  принадлежат  $V \cap W$ , т. е. они принадлежат как  $V$ , так и  $W$ . Тогда, так как  $V$  и  $W$  — векторные пространства,  $x + x'$  и  $sx$  также принадлежат как  $V$ , так и  $W$ . Следовательно, операции сложения и умножения на скаляр не выводят за пределы пересечения. Геометрически это выглядит следующим образом: пересечение двух плоскостей, проходящих через начало координат (или «гиперплоскостей» в  $\mathbb{R}^n$ ), также является подпространством. То же самое будет верно относительно пересечения нескольких подпространств и даже бесконечного множества подпространств.

**Пример 1.** Пересечение двух ортогональных подпространств  $V$  и  $W$  является нулевым подпространством  $\{0\}$ .

**Пример 2.** Рассматривая множество матриц порядка  $n$  как векторное пространство, а подмножества верхних и нижних треугольных матриц как подпространства, легко убедиться, что их пересечение состоит из множества диагональных матриц порядка  $n$ . Это множество, несомненно, является подпространством, потому что сложение двух диагональных матриц или умножение диагональной матрицы на скаляр не выводит нас за пределы этого множества.

**Пример 3.** Предположим, что  $V$  — нуль-пространство матрицы  $A$  размера  $k \times n$  и  $W$  — нуль-пространство матрицы  $B$  размера  $l \times n$ . Тогда  $V \cap W$  — нуль-пространство матрицы

$$C = \begin{bmatrix} A \\ B \end{bmatrix},$$

сформированной из  $k$  строк матрицы  $A$  и  $l$  строк матрицы  $B$ .

<sup>1)</sup> Наиболее важные определения даются в 2X и 2Y. Остаток этого параграфа развивает понятия, которые скорее относятся к теории, чем к приложениям.

**Доказательство.** Вектор  $x$  принадлежит нуль-пространству матрицы  $C$ , т. е.  $Cx=0$ , тогда и только тогда, когда  $Ax=0$  и  $Bx=0$ .

**Пример 4.** Пусть  $v_1, \dots, v_k$ —множество векторов из  $\mathbf{R}^n$ . Рассмотрим все подпространства, содержащие это множество;  $\mathbf{R}^n$ —одно из таких подпространств (каждое векторное пространство является подпространством самого себя), но, очевидно, имеются и другие. Пересечение всех этих подпространств является наименьшим подпространством, содержащим данные векторы. Это подпространство, порожденное векторами  $v_1, \dots, v_k$ .

Обычно после того, как обсуждено и проиллюстрировано на примерах понятие пересечения двух множеств, естественно рассмотреть их объединение. Для векторных пространств это, однако, не естественно. *Объединение  $V \cup W$  двух подпространств, вообще говоря, не будет подпространством.* Рассмотрим, например, оси  $x$  и  $y$  на плоскости. Каждая ось сама по себе является подпространством, но их объединение не является подпространством, потому что сумма двух векторов  $(1, 0)$  и  $(0, 1)$  не принадлежит этим осям. Нетрудно видеть, что аналогичная ситуация будет всегда, исключая, разве что, случай, когда одно из подпространств содержится в другом. Только тогда их объединение (которое совпадает с бóльшим подпространством) также будет подпространством.

Тем не менее мы хотим «скомбинировать» два подпространства и потому вместо их объединения рассмотрим их сумму.

**2У.** Если  $V$  и  $W$  — два подпространства данного пространства, то их сумма  $V + W$ , определяемая как множество всевозможных комбинаций  $x = v + w$ , где  $v$  — произвольный вектор из  $V$ , а  $w$  — произвольный вектор из  $W$ , также будет подпространством этого пространства.

Это подпространство — не что иное, как подпространство, порожденное множеством  $V \cup W$ , т. е. наименьшее векторное подпространство, содержащее как  $V$ , так и  $W$ . Сумма осей  $x$  и  $y$  является всей плоскостью  $xy$ ; сумма двумерной плоскости и прямой, не лежащей в этой плоскости, является трехмерным подпространством.

**Пример 5.** Предположим, что  $V$  и  $W$  — ортогональные дополнения друг к другу в  $\mathbf{R}^n$ . Тогда их сумма равна всему  $\mathbf{R}^n$ , т. е.  $V + W = \mathbf{R}^n$ ; каждый вектор  $x \in \mathbf{R}^n$  может быть разложен в сумму  $x = x_V + x_W$ , где  $x_V \in V$ ,  $x_W \in W$ .

**Пример 6.** Если  $V$  — пространство верхних треугольных матриц, а  $W$  — пространство нижних треугольных матриц, то  $V + W$  —

пространство всех матриц. Каждая матрица может быть записана как сумма верхней треугольной и нижней треугольной матриц многими способами (можно разными способами разлагать диагональ).

**Пример 7.** Если  $V$  — пространство столбцов матрицы  $A$ , а  $W$  — пространство столбцов матрицы  $B$ , то  $V + W$  — пространство столбцов расширенной матрицы  $Q = [A \ B]$ . Размерность пространства  $V + W$  может быть меньше, чем сумма размерностей пространств  $V$  и  $W$  (потому что два эти пространства могут перекрываться), но ее легко найти:

$$\dim(V + W) = \text{ранг } Q. \quad (19)$$

Как ни странно, *вычисление пространства  $V \cap W$  более сложно*. Предположим, что нам заданы два базиса  $v_1, \dots, v_k$  и  $w_1, \dots, w_l$ , и мы хотим получить базис пересечения двух подпространств. Конечно, для этого не достаточно проверить, будут ли какие-нибудь из векторов  $v_i$  совпадать с какими-либо из  $w_j$ . Пространства  $V$  и  $W$  могут совпадать, а базисы при этом быть совершенно разными.

Наиболее эффективный метод таков: составим ту же самую матрицу  $Q$ , столбцами которой являются векторы  $v_1, \dots, v_k, w_1, \dots, w_l$ , и вычислим нуль-пространство  $\mathfrak{N}(Q)$ . Мы покажем, что базис этого подпространства позволяет построить базис подпространства  $V \cap W$  и что эти два пространства имеют одинаковую размерность:

$$\dim(V \cap W) = \dim \mathfrak{N}(Q). \quad (20)$$

Это приводит к формуле, которая важна сама по себе. Складывая равенства (19) и (20), получаем

$$\dim(V + W) + \dim(V \cap W) = r(Q) + v(Q),$$

где  $r(Q)$  — ранг, а  $v(Q)$  — дефект матрицы  $Q$ . Из предыдущих разделов мы знаем, что сумма ранга и дефекта любой матрицы равна числу столбцов этой матрицы. В данном случае матрица  $Q$  имеет  $k + l$  столбцов, а так как  $k = \dim V$  и  $l = \dim W$ , то мы приходим к следующему заключению:

$$\| \quad \dim(V + W) + \dim(V \cap W) = \dim V + \dim W. \quad (21)$$

Неплохая формула.

**Пример 8.** Каждое из пространств  $V$  и  $W$  верхних и нижних треугольных матриц имеет размерность, равную  $n(n + 1)/2$ . Пространство  $V + W$  — пространство всех матриц — имеет размерность  $n^2$ , а пространство  $V \cap W$  — пространство диагональных матриц —

имеет размерность  $n$ . Таким образом, как и предсказывает формула (21), имеет место равенство

$$n^2 + n = n(n+1)/2 + n(n+1)/2.$$

Вернемся теперь к доказательству формулы (20). Только один раз в этой книге нас будут интересовать не столько действительные вычисления, сколько техника доказательства, и только один раз мы воспользуемся при изучении одного пространства сравнением его с другим, и это приведет нас к полезной формуле (21). Отметим, прежде всего, что нуль-пространство матрицы  $Q$  является подпространством пространства  $\mathbf{R}^{k+l}$ , в то время как пространство  $V \cap W$  является подпространством в  $\mathbf{R}^m$ . При этом мы хотим доказать, что эти два подпространства имеют одинаковую размерность. Идея доказательства заключается в установлении между этими подпространствами взаимно однозначного соответствия.

Пусть задан произвольный вектор  $x$  из нуль-пространства матрицы  $Q$ . Перепишем уравнение  $Qx=0$  в следующем виде:

$$Qx = x_1v_1 + \dots + x_kv_k + x_{k+1}w_1 + \dots + x_{k+l}w_l = 0$$

или, эквивалентно,

$$x_1v_1 + \dots + x_kv_k = -x_{k+1}w_1 - \dots - x_{k+l}w_l.$$

Левая часть этого уравнения представляет собой вектор, принадлежащий пространству  $V$ , а правая часть — вектор, принадлежащий пространству  $W$ . Так как эти векторы равны, то они представляют собой некоторый вектор  $y$ , принадлежащий подпространству  $V \cap W$ . Таким образом, мы установили соответствие между вектором  $x \in \mathfrak{N}(Q)$  и вектором  $y \in V \cap W$ . Легко проверить, что это соответствие сохраняет операции сложения и умножения на скаляр, т. е. если вектор  $x \in \mathfrak{N}(Q)$  соответствует вектору  $y \in V \cap W$ , а вектор  $x' \in \mathfrak{N}(Q)$  — вектору  $y' \in V \cap W$ , то вектор  $x + x' \in \mathfrak{N}(Q)$  соответствует вектору  $y + y' \in V \cap W$ , а вектор  $sx \in \mathfrak{N}(Q)$  соответствует вектору  $sy \in V \cap W$ . Кроме того, это соответствие является взаимно однозначным, т. е. каждый вектор  $y \in V \cap W$  соответствует одному и только одному вектору  $x \in \mathfrak{N}(Q)$  (упр. 2.6.5).

Это точная иллюстрация понятия *изоморфизма* между двумя векторными пространствами. Пространства  $V \cap W$  и  $\mathfrak{N}(Q)$  различны, но с алгебраической точки зрения они совершенно одинаковые. Они совпадают во всем: линейно независимые множества соответствуют линейно независимым множествам, а базис одного пространства соответствует базису другого. Поэтому их размерности равны, что и завершает доказательство формул (20) и (21). Такого типа результаты позволяют алгебраистам идентифицировать два различных математических объекта, совпадающих по сущности своей. Известно, что любые два пространства с одина-

ковым множеством скаляров, имеющие одинаковую (конечную) размерность, всегда изоморфны, но это слишком общее утверждение, чтобы оно могло взволновать нас. Интерес возникает, когда оказываются идентичными два совершенно внешне не похожих пространства, например  $\mathfrak{N}(Q)$  и  $V \cap W$ .

**Пример 9.** Предположим, что пространство  $V$  порождается первыми двумя столбцами, а пространство  $W$  — последними двумя столбцами матрицы

$$Q = \begin{bmatrix} 4 & 2 & 0 & 0 \\ 0 & 3 & -6 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Эта матрица уже имеет ступенчатый вид, причем третья переменная является свободной (третий столбец не имеет ведущего элемента). Поэтому первый, второй и четвертый столбцы порождают пространство столбцов  $V + W$ , которое, имея размерность, равную трем, должно совпадать со всем  $\mathbf{R}^3$ . Чтобы найти  $V \cap W$ , вычислим нуль-пространство матрицы  $Q$ . Присвоим свободной переменной  $x_3$  значение, равное единице ( $x_3 = 1$ ), и разрешим полученную систему относительно базисных переменных. Тогда получим  $x_4 = 0$ ,  $x_2 = 2$ ,  $x_1 = -1$ . Соответствующее решение системы  $Qx = 0$  имеет вид  $x = (-1, 2, 1, 0)^T$ , и наше взаимно однозначное соответствие между  $\mathfrak{N}(Q)$  и  $V \cap W$  отображает его в вектор

$$y = (-1) (\text{столбец } 1) + (2) (\text{столбец } 2) = (-1) \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} + (2) \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 0 \end{bmatrix},$$

который и является базисом для (одномерного) пересечения  $V \cap W$ .

**Упражнение 2.6.1.** Пусть задано пространство квадратных матриц четвертого порядка, и пусть  $V$  будет подпространством трехдиагональных матриц, а  $W$  — подпространством верхних треугольных матриц. Описать подпространство  $V + W$ , элементами которого являются верхние матрицы Хессенберга, и подпространство  $V \cap W$ . Проверить формулу (21).

**Упражнение 2.6.2.** Предположим, что  $V \cap W = \{0\}$ . Тогда из формулы (21) следует равенство  $\dim(V + W) = \dim V + \dim W$ . Показать, что каждый элемент  $x \in V + W$  может быть единственным образом представлен в виде суммы  $x = v + w$ , где  $v \in V$ , а  $w \in W$ . Иначе говоря, нужно предположить, что  $x = v' + w'$ , и доказать, что  $v = v'$  и  $w = w'$ . Если  $V \cap W = \{0\}$ , то пространство  $V + W$  называется *прямой суммой* пространств  $V$  и  $W$  и часто записывается как  $V \oplus W$ . Ортогональные подпространства всегда составляют прямую сумму так же, как любые два подпространства, пересечение которых состоит только из нулевого вектора

**Упражнение 2.6.3.** Предположим, что подпространство  $V$  порождается вектор-столбцами  $(1, 1, 0, 1)^T$  и  $(1, 2, 0, 0)^T$ . Найти подпространство  $W$ , такое, что в обозначениях предыдущего упражнения  $V \oplus W = \mathbf{R}^4$ .

**Упражнение 2.6.4.** Найти базис подпространства  $V + W$ , где подпространство  $V$  порождается  $v_1 = (1, 1, 0, 0)$  и  $v_2 = (1, 0, 1, 0)$ , а подпространство  $W$  порождается  $w_1 = (0, 1, 0, 1)$  и  $w_2 = (0, 0, 1, 1)$ . Найти также размерность подпространства  $V \cap W$  и его базис.

**Упражнение 2.6.5.** Проверить утверждение, что каждому вектору  $y \in V \cap W$  соответствует один и только один вектор  $x \in \mathfrak{N}(Q)$ , описав, как для данного вектора  $y \in V \cap W$  найти единственный подходящий вектор  $x \in \mathfrak{N}(Q)$ .

### ОСНОВНЫЕ ПРОСТРАНСТВА ДЛЯ ПРОИЗВЕДЕНИЙ МАТРИЦ $AB$

Перейдем теперь от пар подпространств к произведениям матриц. Как всегда, объектом нашего изучения будут не отдельные элементы матрицы  $AB$ , которые, возможно, не имеют ничего общего с элементами матриц  $A$  и  $B$ , а векторы (строки или столбцы), которые скорее, чем отдельные элементы, могут унаследовать некоторые свойства матриц  $A$  и  $B$ ; точнее, даже не столько отдельные строки или столбцы, сколько подпространства, которые они порождают. Эти подпространства отражают свойства всей матрицы целиком.

Наш основной вопрос: какие связи можно установить между четырьмя основными подпространствами, связанными с матрицей  $A$ , четырьмя подпространствами, связанными с матрицей  $B$ , и четырьмя подпространствами, связанными их произведением  $AB$ ? Для этих матриц (все они могут быть прямоугольными) имеют место четыре основных соотношения.

(i) Нуль-пространство матрицы  $AB$  содержит нуль-пространство матрицы  $B$ , т. е.  $\mathfrak{N}(AB) \supseteq \mathfrak{N}(B)$ .

(ii) Пространство столбцов матрицы  $AB$  содержится в пространстве столбцов матрицы  $A$ , т. е.  $\mathfrak{R}(AB) \subseteq \mathfrak{R}(A)$ .

(iii) Левое нуль-пространство матрицы  $AB$  содержит левое нуль-пространство матрицы  $A$ , т. е.  $\mathfrak{N}((AB)^T) \supseteq \mathfrak{N}(A^T)$ .

(iv) Пространство строк матрицы  $AB$  содержится в пространстве строк матрицы  $B$ , т. е.  $\mathfrak{R}((AB)^T) \subseteq \mathfrak{R}(B^T)$ .

Доказательство крайне просто.

(i) Если  $Bx = 0$ , то  $ABx = 0$ . Следовательно,  $\mathfrak{N}(B) \subseteq \mathfrak{N}(AB)$ .

(ii) Предположим, что  $b \in \mathfrak{R}(AB)$ , или, иначе говоря,  $ABx = b$  для некоторого вектора  $x$ . Тогда  $b \in \mathfrak{R}(A)$ , поскольку  $Ay = b$  для  $y = Bx$ . Следовательно,  $\mathfrak{R}(AB) \subseteq \mathfrak{R}(A)$ .

Другое доказательство следует прямо из свойств умножения матриц, потому что каждый столбец матрицы  $AB$  является линейной комбинацией столбцов матрицы  $A$ .

(iii) Так как  $(AB)^T = B^T A^T$ , то третье утверждение совпадает с первым после того, как матрицу  $A$  в (i) заменить на матрицу  $B^T$ , а матрицу  $B$  — на  $A^T$ .

(iv) Четвертое утверждение совпадает со вторым после той же самой замены. Кроме того, оно прямо следует из свойств умножения матриц:  $i$ -я строка матрицы  $AB$  является линейной комбинацией строк матрицы  $B$  с весами, являющимися элементами  $i$ -й строки матрицы  $A$ . Следовательно, пространство строк матрицы  $AB$  содержится в пространстве строк матрицы  $B$ .

**Упражнение 2.6.6.** Показать на примере, что нуль-пространство матрицы  $AB$  может не содержать нуль-пространство матрицы  $A$ , а пространство столбцов матрицы  $AB$  может не содержаться в пространстве столбцов матрицы  $B$ .

**Следствие.** Ранг  $r$  и дефект  $\nu$  удовлетворяют неравенствам

$$r(AB) \leq r(A) \quad \text{и} \quad r(AB) \leq r(B), \quad (22)$$

$$\nu(AB) \geq \nu(B). \quad (23)$$

Так как пространство столбцов матрицы  $AB$  содержится в пространстве столбцов матрицы  $A$ , а ранг матрицы равен размерности ее пространства столбцов, то отсюда непосредственно следует, что  $r(AB) \leq r(A)$ . Аналогично из (iv) вытекает, что  $r(AB) \leq r(B)$ , а неравенство (23) для дефекта следует из (i). Отметим, что мы не пытаемся доказать неравенство  $\nu(AB) \geq \nu(A)$ , которое для произвольных матриц не выполняется.

**Упражнение 2.6.7.** Показать на примере матриц, составленных из одних нулей, что  $\nu(AB)$  может быть меньше, чем  $\nu(A)$ .

Существует еще одно характерное приложение, которое связано с полученными соотношениями (i) — (iv). Рассмотрим произвольную матрицу  $A$  размера  $m \times n$  и ее разложение  $PA = LU$  (или  $A = P^{-1}LU$ ). Вспоминая, что последние  $m - r$  строк матрицы  $U$  равны нулю, предположим, что мы отбросили их, чтобы получить матрицу  $\bar{U}$  порядка  $r \times n$ . (В случае  $r = m$  отбрасывать нечего и  $\bar{U} = U$ .) Теперь внимательно посмотрим, как перемножаются матрицы:

$$A = (P^{-1}L)U = (P^{-1}L) \begin{bmatrix} \bar{U} \\ 0 \end{bmatrix}. \quad (24)$$

Последние  $m - r$  столбцов матрицы  $P^{-1}L$  умножаются только на нижние нулевые строки матрицы  $U$ , в силу чего мы можем отбросить также и эти столбцы. Обозначим получившуюся матрицу через  $\bar{L}$ , хотя в действительности ее надо было обозначить через  $P^{-1}\bar{L}$ , так как она состоит из первых  $r$  столбцов матрицы  $P^{-1}L$ . До сих пор наши рассуждения не дали ничего, кроме следующего утверждения: **отбрасывание нулей в равенстве (24) приводит к новому разложению  $A = \bar{L}\bar{U}$ .**

К произведению матриц  $\bar{L}$  и  $\bar{U}$  мы применим утверждение (ii) для пространства столбцов произведений: пространство столбцов матрицы  $A = \bar{L}\bar{U}$  содержится в пространстве столбцов матрицы  $\bar{L}$ . Мы знаем, что пространство столбцов матрицы  $A$  имеет размерность  $r$ . Так как матрица  $\bar{L}$  имеет только  $r$  столбцов, то ее пространство столбцов не может быть больше пространства столбцов матрицы  $A$  и, следовательно, *эти два пространства столбцов совпадают*. Матрица  $\bar{L}$  имеет то же пространство столбцов, что и матрица  $A$ , а матрица  $\bar{U}$  имеет то же пространство строк, что и матрица  $A$ .

**Упражнение 2.6.8.** Разложить матрицу

$$A = \begin{bmatrix} 0 & 1 & 4 & 0 \\ 0 & 2 & 8 & 0 \end{bmatrix}$$

в произведение  $A = \bar{L}\bar{U}$  и проверить, что столбцы матрицы  $\bar{L}$  являются базисом пространства столбцов матрицы  $A$ .

**Упражнение 2.6.9.** Повторить предыдущее задание для матрицы  $A$ , требующей применения матрицы перестановки  $P$ :

$$A = \begin{bmatrix} 0 & 0 \\ 1 & 2 \\ 4 & 8 \\ 0 & 0 \end{bmatrix}.$$

**Упражнение 2.6.10.** Умножение каждого столбца матрицы  $\bar{L}$  на строку с тем же номером матрицы  $\bar{U}$  позволяет разложить матрицу  $A = \bar{L}\bar{U}$  в сумму  $r$  матриц единичного ранга. Построить матрицы  $\bar{L}$  и  $\bar{U}$  и разложить в сумму  $r$  матриц единичного ранга следующую матрицу ранга 2:

$$A = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 5 \\ -1 & -3 & 3 & 0 \end{bmatrix}.$$

Наконец, напомним, что *подматрица*  $C$  образуется путем вычеркивания некоторых (или никаких) строк матрицы  $A$  и некоторых (или никаких) столбцов этой матрицы. Нетрудно угадать ограничение на ранг матрицы  $C$ .

**2Z.** Предположим, что  $A$  — матрица размера  $m \times n$  и ее ранг  $r(A)$  равен  $r$ . Тогда

- (i) для каждой ее подматрицы  $C$  имеем  $r(C) \leq r(A) = r$ ;
- (ii) существует по крайней мере одна подматрица  $C$  размера  $r \times r$ , ранг которой равен  $r$ , т. е.  $r(C) = r(A) = r$ .

**Доказательство.** Сведем матрицу  $A$  к матрице  $C$  в два этапа. На первом этапе оставим число столбцов неизменным и вы-

черкнем только нежелательные строки. Пространство строк этой промежуточной матрицы  $B$ , очевидно, содержится в пространстве строк матрицы  $A$ , в силу чего  $r(B) \leq r(A) = r$ . На втором этапе матрица  $B$  сводится к матрице  $C$  вычеркиванием нежелательных столбцов. Поэтому пространство столбцов матрицы  $C$  содержится в пространстве столбцов матрицы  $B$  и, следовательно,  $r(C) \leq r(B) \leq r$ . Это устанавливает (i).

Чтобы доказать (ii), предположим, что матрица  $B$  образована из  $r$  линейно независимых строк матрицы  $A$ . Тогда пространство строк матрицы  $B$  имеет размерность  $r$ , т. е.  $r(B) = r(A) = r$ . Пространство столбцов матрицы  $B$  также должно иметь размерность  $r$ . Предположим далее, что  $C$  образована из  $r$  линейно независимых столбцов матрицы  $B$ . Тогда пространство столбцов матрицы  $C$  имеет размерность  $r$  и, следовательно,  $r(C) = r$ . Это завершает доказательство (ii): каждая матрица ранга  $r$  содержит невырожденную подматрицу размера  $r \times r$ .

**Пример 10.** Рассмотрим еще раз матрицу

$$A = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 5 \\ -1 & -3 & 3 & 0 \end{bmatrix}$$

размера  $3 \times 4$ , имеющую ранг  $r = 2$ , с подматрицей

$$C = \begin{bmatrix} 1 & 3 \\ 2 & 9 \end{bmatrix}.$$

Каждая подматрица  $A$ , имеющая размер  $3 \times 3$ , вырождена, а подматрица  $C$  невырождена.

Последняя теорема не заслуживает особого выделения. Она похожа на *действительно важную* теорему из конца предыдущего параграфа (первую после рис. 2.7) и имеющую номер 2W. Там доказано, что каждая матрица  $A$  задает обратимое отображение ее  $r$ -мерного пространства строк в ее  $r$ -мерное пространство столбцов. Эти пространства и это отображение дают полную информацию о матрице  $A$ . При этом матрица может быть восстановлена по такому отображению. Здесь речь идет лишь об отыскании обратимой подматрицы  $C$  и ничего не говорится о выборе этой подматрицы; у матрицы  $A$  может существовать много различных обратимых подматриц порядка  $r$ , как показывают примеры. Мы дадим новое, эквивалентное прежнему определение ранга: рангом матрицы  $A$  называется *порядок максимальной ее невырожденной подматрицы*.

**Упражнение 2.6.11.** Для матрицы  $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 6 & 9 \end{bmatrix}$  найти максимальную невырожденную подматрицу и ранг.

Ниже мы предлагаем ряд более абстрактных упражнений, чтобы получить некоторую практику доказательств.

**Упражнение 2.6.12.** Предположим, что  $A$  — матрица размера  $m \times n$ , а  $B$  — матрица размера  $n \times t$  и  $n < t$ . Доказать, что их произведение  $AB$  вырождено.

**Упражнение 2.6.13.** Доказать, что  $r(A+B) \leq r(A) + r(B)$ .

**Упражнение 2.6.14.** Пусть  $\mathfrak{N}(AB) = \mathfrak{N}(B)$  и некоторый вектор  $y$  лежит в  $\mathfrak{N}(B) \cap \mathfrak{N}(A)$ . Доказать, что  $y = 0$ .

**Упражнение 2.6.15.** Пусть  $A$  — квадратная и обратимая матрица. Показать, что матрица  $AB$  имеет те же самые нуль-пространство, пространство строк и ранг, что и матрица  $B$ . Указание: применить утверждение (i) из 2Z к произведению  $A^{-1}(AB)$ .

**Упражнение 2.6.16.** Рассмотрим четырехмерное векторное пространство  $V$  всех квадратных матриц второго порядка, и пусть задана некоторая матрица  $A$  из  $V$ . Предположим, что пространство  $V$  преобразуется в себя по следующему правилу: каждая матрица  $B \in V$  отображается в матрицу  $AB \in V$ . Если матрица  $A$  обратима, то это отображение не имеет нуль-пространства, т. е. существует только одна матрица, отображающаяся в нулевую, — это нулевая матрица.

(i) Пусть  $A = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}$ ; описать нуль-пространство соответствующего ей отображения, т. е. множество всех квадратных матриц  $B$ , таких, что  $AB = 0$ , и вычислить его размерность (которая равна дефекту отображения, а не дефекту матрицы  $A$ ).

(ii) Описать также образ или «пространство столбцов» этого отображения, т. е. пространство всех матриц  $AB$ , где  $A = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}$ , а  $B$  — произвольная матрица из  $V$ . Вычислить размерность этого пространства (которая равна рангу отображения, а не рангу матрицы). Проверить, что  $\text{ранг} + \text{дефект} = \dim V$ .

## ОБЗОРНЫЕ УПРАЖНЕНИЯ

2.1. Найти базис подпространства пространства  $\mathbf{R}^4$ , компоненты векторов которого удовлетворяют равенствам  $x_1 = x_2 = x_3$ .

2.2. Найти ступенчатый вид  $U$  матрицы

$$A = \begin{bmatrix} 1 & 2 & 0 & 2 & 1 \\ -1 & -2 & 1 & 1 & 0 \\ 1 & 2 & -3 & -7 & -2 \end{bmatrix}.$$

Каковы размерности четырех основных подпространств, связанных с этой матрицей?

2.3. Определить базис двумерного подпространства из  $\mathbf{R}^3$ , не содержащего ни один из координатных векторов  $(1, 0, 0)$ ,  $(0, 1, 0)$ ,  $(0, 0, 1)$ .

2.4. Найти ранг и нуль-пространство матрицы

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

2.5. Построить матрицу, нуль-пространство которой порождается вектором  $[1, 0, 1]^T$

2.6. Определить, верны или ложны следующие утверждения. Если ложно, то привести контрпример

- (i) Если векторы  $x_1, \dots, x_m$  порождают подпространство  $S$ , то  $\dim S = m$ .
- (ii) Пересечение двух подпространств векторного пространства не может быть пусто.
- (iii) Если  $Ax = Ay$ , то  $x = y$ .
- (iv) Пространство строк матрицы  $A$  имеет единственный базис, который может быть вычислен путем приведения матрицы  $A$  к ступенчатому виду
- (v) Если квадратная матрица  $A$  имеет линейно независимые столбцы, то этим свойством обладает и матрица  $A^2$

2.7. Найти базисы для четырех основных подпространств, связанных с матрицами

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

2.8. Найти общее решение системы  $u + v + w = 1$ ,  $u - w = 2$ .

2.9. Существует ли матрица, пространство строк которой содержит вектор  $[1 \ 1 \ 1]^T$ , а пространство столбцов — вектор  $[1 \ 0 \ 0]^T$ ?

2.10. Найти все векторы, перпендикулярные к векторам  $(1, 4, 4, 1)$  и  $(2, 9, 8, 2)$ .

2.11. Могут ли векторы  $(1, 2, 1)$ ,  $(0, 1, 3)$  и  $(3, 7, 6)$  лежать в одной и той же плоскости пространства  $\mathbb{R}^3$ ?

2.12. Которому из четырех основных подпространств, связанных с матрицей  $A$ , должен быть ортогонален вектор  $b$  для того, чтобы система  $Ax = b$  имела решение?

2.13. Найти все решения систем

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 1 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \quad \text{и} \quad \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 1 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}.$$

2.14. Могут ли векторы  $(1, 1, 3)$ ,  $(2, 3, 6)$  и  $(1, 4, 3)$  образовывать базис в  $\mathbb{R}^3$ ?

2.15. Найти матрицы  $A$ , для которых число решений системы  $Ax = b$  равняется:

- (i) 0 или 1 в зависимости от вектора  $b$ ;
- (ii) 1 или  $\infty$  в зависимости от вектора  $b$ ;
- (iii) 0 или  $\infty$  в зависимости от вектора  $b$ ;
- (iv) 1 независимо от вектора  $b$ .

2.16. Как должно быть связано число  $r$  с  $n$  и  $m$  в каждом случае предыдущего упражнения?

# ОРТОГОНАЛЬНЫЕ ПРОЕКЦИИ И МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

## § 3.1. СКАЛЯРНЫЕ ПРОИЗВЕДЕНИЯ И ТРАНСПОНИРОВАНИЕ

Мы уже знаем, что скалярное произведение двух векторов  $x$  и  $y$  есть число  $x^T y$ . До сих пор нас интересовал лишь вопрос, равно ли нулю это число, другими словами, ортогональны ли эти два вектора. Теперь же мы будем допускать возможность, что скалярные произведения не равны нулю, т. е. что углы не являются прямыми, и интересоваться соотношением между скалярным произведением и углом. В этом параграфе мы также коснемся вопроса о связи между скалярным произведением и транспонированием. В предыдущей главе транспонирование осуществлялось путем переворачивания матрицы, как оладьи; теперь эта процедура будет заменена более строгой.

Суммируя материал оставшейся части главы, следует отметить, что *ортогональный случай все-таки является наиболее важным*. Предположим, что задана точка  $b$  в  $n$ -мерном пространстве и мы хотим найти расстояние от этой точки до прямой, порожденной вектором  $a$ , т. е. мы ищем на этой прямой точку  $p$ , ближайшую к  $b$ . Тогда прямая, соединяющая точки  $b$  и  $p$  (изображенная на рис. 3.1 пунктиром), перпендикулярна к исходному вектору  $a$ .

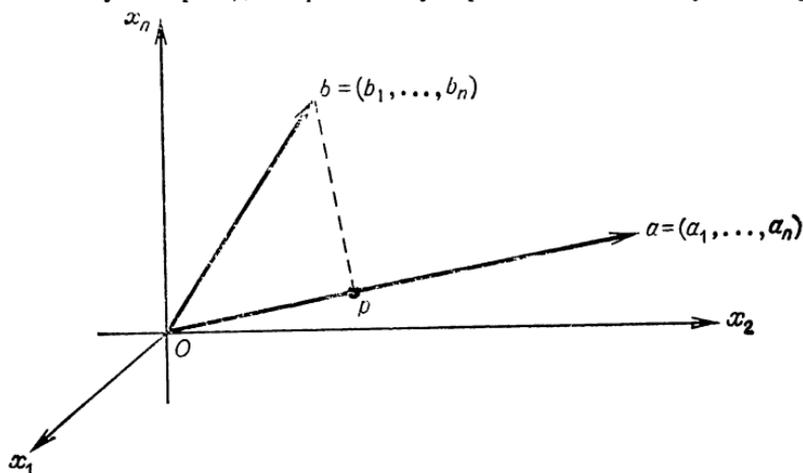


Рис. 3.1 Проекция в  $n$ -мерном пространстве.

Этот факт позволяет нам найти ближайшую точку  $p$  и вычислить расстояние от нее до  $b$ . Несмотря на то что исходные векторы  $a$  и  $b$  не были ортогональны, для решения задачи автоматически привлекается ортогональность.

Ситуация будет аналогичной, если вместо прямой, определенной вектором  $a$ , задана плоскость или вообще любое подпространство  $S$  пространства  $\mathbf{R}^n$ . Задача вновь сводится к отысканию точки  $p$  в этом подпространстве, которая является ближайшей к  $b$ , и эта точка вновь оказывается проекцией точки  $b$  на подпространство  $S$ . Если мы опустим перпендикуляр из точки  $b$  на  $S$ , то  $p$  будет точкой пересечения этого перпендикуляра с подпространством  $S$ . Геометрически это соответствует решению задачи о расстояниях между точками и подпространствами. Однако остаются и некоторые вопросы, а именно:

- 1) Имеет ли эта задача практическое значение?
- 2) Существует ли аналитическая формула для определения точки  $p$ , если подпространство  $S$  задается определенным базисом (или просто набором векторов, порождающих его)?
- 3) Существует ли устойчивый (с вычислительной точки зрения) способ вычисления точки  $p$  при помощи этой формулы?

Ответ на первые два вопроса, безусловно, является положительным. Наша задача, до сих пор описанная в геометрических терминах, в точности совпадает с задачей о *решении переопределенной системы уравнений методом наименьших квадратов*. Вектор  $b$  представляет собой информацию, полученную в результате проведения серии экспериментов или опросов, и в ней содержится слишком много ошибок, чтобы он мог попасть в данное подпространство  $S$ . Другими словами, если мы попытаемся представить вектор  $b$  как вектор из подпространства  $S$ , то нам не удастся этого сделать, поскольку соответствующая система уравнений окажется несовместной и, значит, не имеет решения. А метод наименьших квадратов даст нам точку  $p$ , наилучшую из возможных. Сомневаться в важности этого приложения не приходится<sup>1)</sup>.

Второй вопрос, об отыскании формулы для точки  $p$ , решается очень просто, когда подпространство представляет собой прямую. В этом и следующем параграфах мы будем проектировать один вектор на другой различными способами и соотносить это проектирование с вопросом о скалярных произведениях и об углах. К счастью, формула для  $p$  остается достаточно простой и при проектировании на подпространства большей размерности (при условии, что в них задан базис). Это наиболее важный случай; он соответствует задаче о наименьших квадратах с несколькими

<sup>1)</sup> В экономике и статистике это называется *регрессионным анализом*.

параметрами и решается в § 3.2. Остаются еще две возможности, заслуживающие отдельного рассмотрения:

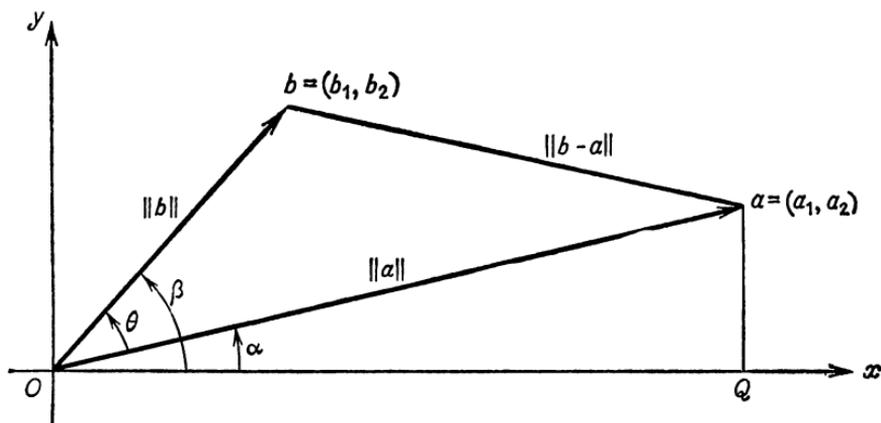
а) Если подпространство  $S$  описывается набором порождающих его векторов, причем линейно зависимых, то эта вырожденность приводит к нарушению формулы для  $p$ . Точка  $p$  остается единственной точкой, ближайшей к  $b$ , но она может быть выражена различными способами в виде комбинации линейно зависимых наборов векторов. Поэтому требуется дополнительное правило, позволяющее выделить одну комбинацию, и это приводит нас в § 3.4 к понятию псевдообратной матрицы.

б) Предположим гораздо более благоприятную ситуацию, когда описывающие  $S$  векторы не только линейно независимы, но и взаимно ортогональны. В этом случае формула для  $p$  становится особенно простой и все сводится к случаю проектирования на прямую. Столь же простыми будут и вычисления. Это обстоятельство наводит на мысль о необходимости предварительной подготовки базиса с тем, чтобы составляющие его векторы стали взаимно ортогональными. Любой базис может быть преобразован в ортогональный, и в § 3.3 описывается простой процесс Грама — Шмидта, служащий для этой цели.

Остается ответить на третий вопрос о том, может ли быть реализована устойчивая схема для метода наименьших квадратов. Сделать это труднее по следующей причине: подпространство  $S$  определяется набором векторов, и если эти векторы будут близки к линейной зависимости, то оказывается практически невозможным установить их линейную зависимость или независимость. Другими словами, размерность подпространства  $S$  (число линейно независимых векторов) является исключительно неустойчивой, что делает неустойчивыми как точку  $p$ , так и псевдообратную матрицу. В более устойчивом случае «нормальные уравнения», описанные в § 3.2, дают простейший способ для вычисления  $p$  (то же самое относится и к процессу ортогонализации Грама — Шмидта или к сингулярному разложению из § 3.4).

### СКАЛЯРНЫЕ ПРОИЗВЕДЕНИЯ И НЕРАВЕНСТВО ШВАРЦА

Вернемся к обсуждению вопроса о скалярных произведениях и углах. Ниже будет показано, что скалярные произведения связаны не с самими углами, а с их *косинусами*. Поэтому мы обратимся к тригонометрии, т. е. к двумерному случаю, чтобы найти это соотношение (рис. 3.2). Пусть  $\alpha$  есть угол между вектором  $a$  и осью  $x$ . Вспомогательная, что  $\|a\|$  есть длина вектора  $a$ , равная гипотенузе треугольника  $OaQ$ , получаем выражения для синуса и ко-

Рис. 3.2. Косинус угла  $\theta = \beta - \alpha$ .

синуса угла  $\alpha$  в виде

$$\sin \alpha = \frac{a_2}{\|a\|}, \quad \cos \alpha = \frac{a_1}{\|a\|}.$$

То же самое справедливо и для вектора  $b$  с соответствующим углом  $\beta$ : его синус равен  $b_2/\|b\|$ , а косинус равен  $b_1/\|b\|$ . Теперь, поскольку угол  $\theta$  в точности равен  $\beta - \alpha$ , его косинус дается хорошо известным тригонометрическим тождеством

$$\cos \theta = \cos \beta \cos \alpha + \sin \beta \sin \alpha = \frac{a_1 b_1 + a_2 b_2}{\|a\| \|b\|}. \quad (1)$$

Числитель этой формулы совпадает со скалярным произведением векторов  $b$  и  $a$ , откуда мы и получаем требуемое соотношение:

|| 3А. Косинус угла между двумя векторами равен

$$\cos \theta = \frac{a^T b}{\|a\| \|b\|}. \quad (2)$$

Отметим, что эта формула правильна и в метрическом отношении: если мы вдвое увеличим вектор  $b$ , то как числитель, так и знаменатель увеличатся вдвое и косинус останется неизменным. Изменение знака вектора  $b$  на противоположный одновременно изменит знак у  $\cos \theta$ , что означает изменение угла на  $180^\circ$ .

*Замечание.* Тот же самый результат может быть получен и из теоремы косинусов, которая связывает длины сторон в произвольном треугольнике и имеет вид

$$\|b - a\|^2 = \|b\|^2 + \|a\|^2 - 2\|b\|\|a\|\cos \theta. \quad (3)$$

Если  $\theta$  является прямым углом, то мы приходим к теореме Пифагора, а при произвольном  $\theta$  выражение  $\|b - a\|^2$  может быть за-

писано в виде  $(b-a)^T(b-a)$  и вместо (3) получаем:

$$b^T b - 2a^T b + a^T a = b^T b + a^T a - 2\|b\|\|a\|\cos\theta.$$

После сокращений мы приходим к формуле (2) для косинуса. Фактически это доказывает и нашу формулу для  $n$ -мерного случая, поскольку все происходит лишь в плоском треугольнике  $Oab$ .

Теперь мы хотим найти проекцию  $p$ . Эта точка должна быть кратна вектору  $a$ , т. е.  $p = \bar{x}a$ , поскольку любая точка этой прямой кратна  $a$ , и задача состоит в отыскании коэффициента  $\bar{x}$ . Для этого нам нужен лишь простой геометрический факт, состоящий в том, что прямая, соединяющая конец вектора  $b$  с точкой  $p$ , перпендикулярна к вектору  $a$ :

$$(b - \bar{x}a) \perp a, \quad \text{или} \quad a^T(b - \bar{x}a) = 0, \quad \text{или} \quad \bar{x} = \frac{a^T b}{a^T a}.$$

**3В.** Проекция  $p$  точки  $b$  на прямую, определенную вектором  $a$ , задается формулой

$$p = \frac{a^T b}{a^T a} a. \quad (4)$$

Расстояние (в квадрате) от этой точки до прямой равняется

$$\begin{aligned} \left\| b - \frac{a^T b}{a^T a} a \right\|^2 &= b^T b - 2 \frac{(a^T b)^2}{a^T a} + \left( \frac{a^T b}{a^T a} \right)^2 a^T a = \\ &= \frac{(b^T b)(a^T a) - (a^T b)^2}{(a^T a)}. \end{aligned} \quad (5)$$

Это позволяет нам повторить рис. 3.1 уже с указанием формулы для определения точки  $p$  (рис. 3.3).

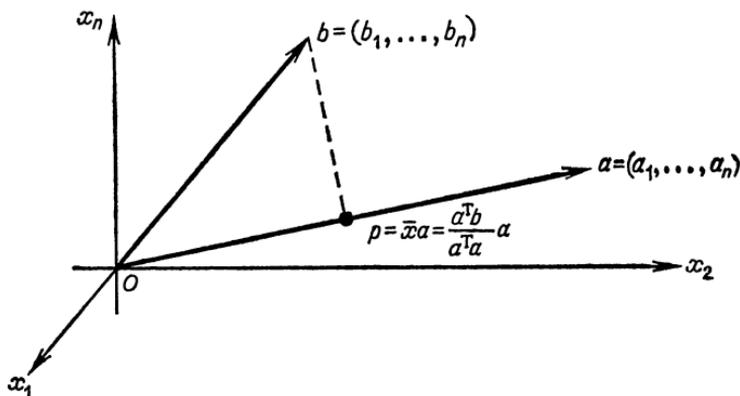


Рис. 3.3. Проекция вектора  $b$  на вектор  $a$ .

Полученная формула имеет замечательное следствие, представляющее собой, вероятно, наиболее важное неравенство в математике. (Как частный случай оно включает в себе и неравенство о среднем геометрическом и среднем арифметическом и эквивалентно — см. упр. 3.1.1 — неравенству треугольника.) Сам по себе этот результат кажется случайным следствием из формулы (5) для расстояния от точки до прямой, т. е. между  $b$  и  $xa$ . Это расстояние, а также его квадрат обязательно должны быть неотрицательными. Поэтому числитель в формуле (5) должен быть неотрицателен:  $(b^T b)(a^T a) - (a^T b)^2 \geq 0$ . Если мы добавим к обеим частям неравенства величину  $(a^T b)^2$  и извлечем корень, то формула (5) переписется в таком виде:

$$\| \quad 3С. \text{ Любые два вектора удовлетворяют неравенству Шварца} \quad \| \\ |a^T b| \leq \|a\| \|b\|. \quad (6)$$

*Замечание.* Согласно формуле (2), отношение двух частей неравенства (6) в точности равно  $|\cos \theta|$ . Поскольку все значения косинуса лежат в пределах от  $-1$  до  $1$ , мы получаем другое доказательство неравенства (6), причем доказательство, в некотором смысле более доступное для понимания благодаря очевидности свойств косинуса. Но мы хотим подчеркнуть простоту нашего доказательства, которое сводится к механическим вычислениям в формуле (5). Выражение в левой части (5) является неотрицательным, и оно будет оставаться неотрицательным даже после введения ниже дополнительных возможностей для скалярных произведений и длин векторов. Поэтому выражение в правой части (5) также будет неотрицательным, и неравенство Шварца доказано, причем без привлечения тригонометрии<sup>1)</sup>.

И последнее замечание: равенство в (6) достигается тогда и только тогда, когда вектор  $b$  является кратным  $a$ . В этом случае конец вектора  $b$  совпадает с точкой  $p$  и расстояние (5) между этой точкой и прямой равно нулю.

### ТРАНСПОНИРОВАНИЕ МАТРИЦЫ

Обратимся теперь к транспонированию матриц. До сих пор матрица  $A^T$  определялась нами лишь как зеркальное отражение матрицы  $A$  относительно ее главной диагонали: строки матрицы  $A$  превращались в столбцы матрицы  $A^T$ , и наоборот. Другими сло-

<sup>1)</sup> Имя Коши также связывается с этим неравенством  $|a^T b| \leq \|a\| \|b\|$ , а в русской литературе даже называют его неравенством Коши — Шварца — Буняковского! История математики показывает, что это, по-видимому, правильно,

вами, элемент  $i$ -й строки и  $j$ -го столбца матрицы  $A^T$  совпадает с элементом  $(j, i)$  матрицы  $A$ :

$$(A^T)_{ij} = (A)_{ji}. \quad (7)$$

Но в транспонировании матрицы заложен и более глубокий смысл, поскольку оно тесно связано со скалярным произведением. На самом деле эта связь может быть использована для того, чтобы дать новое и значительно более «абстрактное» определение транспонирования:

**3Д.** Транспонированная матрица  $A^T$  может быть определена с помощью следующего свойства: скалярное произведение вектора  $Ax$  на вектор  $y$  равняется скалярному произведению вектора  $x$  на  $A^T y$ . Формально это просто означает, что

$$(Ax)^T y = x^T A^T y = x^T (A^T y). \quad (8)$$

Новое определение преследует две цели:

(i) Оно показывает нам, как следует модифицировать транспонирование, когда мы пользуемся другими определениями скалярного произведения. Это будет важно в случае комплексных чисел; новое скалярное произведение появится в § 5.5.

(ii) Оно позволяет нам транспонировать произведение  $AB$  без перемножения этих матриц, причем соответствующее правило совершенно аналогично известному правилу обращения произведения матриц  $(AB)^{-1} = B^{-1}A^{-1}$ .

**3Е.** Транспонированная к  $AB$  матрица равняется произведению транспонированных матриц, взятых в обратном порядке,

$$(AB)^T = B^T A^T. \quad (9)$$

**Доказательство.** Скалярное произведение  $(AB)x$  на  $y$  должно равняться, согласно 3Д, скалярному произведению вектора  $x$  на  $(AB)^T y$ . С другой стороны, оно также равняется скалярному произведению  $A(Bx)$  на  $y$ , что совпадает с произведением  $Bx$  на  $A^T y$  по 3Д. Отсюда по 3Д получаем, что исходное скалярное произведение равно произведению вектора  $x$  на  $B^T(A^T y)$ , откуда и получается равенство  $(AB)^T = B^T A^T$ . Это легко показать на конкретном примере:

$$AB = \begin{bmatrix} 1 & 4 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 1 & 5 \\ 6 & 0 & 3 \end{bmatrix},$$

$$B^T A^T = \begin{bmatrix} 0 & 2 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 8 & 6 \\ 1 & 0 \\ 5 & 3 \end{bmatrix}.$$

Нам нужна еще одна матрица. При проектировании на прямую в знаменателе стояло число  $a^T a$ . При проектировании на подпространство это число превращается в матрицу  $A^T A$ , ранг которой может быть вычислен заранее.

|| 3F. Для произвольной матрицы  $A$  размера  $m \times n$ , имеющей ранг  $r$ , произведение  $A^T A$  является симметрической матрицей ранга  $r$ .

Вспомним, что симметрической называется матрица, совпадающая со своей транспонированной. Поэтому вычислим транспонированную к  $A^T A$  матрицу по правилу (9):

$$(A^T A)^T = A^T (A^T)^T. \quad (10)$$

Но при повторном транспонировании мы вновь вернемся к исходной матрице  $A$ . Поэтому правая часть в (10) совпадает с  $A^T A$  и (10) показывает, что матрица совпадает со своей транспонированной. Другими словами, матрица  $A^T A$  является симметрической.

Для отыскания ее ранга мы покажем, что матрицы  $A$  и  $A^T A$  имеют одинаковые нуль-пространства. Тогда, поскольку ранг плюс размерность нуль-пространства равняется числу столбцов матрицы,  $r + (n - r) = n$ , и матрицы  $A$  и  $A^T A$  имеют по  $n$  столбцов, отсюда немедленно вытекает равенство рангов матриц. Если вектор  $x$  принадлежит нуль-пространству матрицы  $A$ , то  $Ax = 0$  и  $A^T Ax = A^T 0 = 0$ , так что  $x$  принадлежит и нуль-пространству матрицы  $A^T A$ . Обратно, пусть  $A^T Ax = 0$ . Умножая скалярно на вектор  $x$ , получаем

$$x^T A^T Ax = 0, \quad \text{или} \quad \|Ax\|^2 = 0, \quad \text{или} \quad Ax = 0.$$

Следовательно,  $x$  принадлежит нуль-пространству матрицы  $A$  и оба нуль-пространства совпадают.

Отметим один случай, который является как наиболее важным, так и наиболее распространенным. Это случай, когда столбцы матрицы  $A$  линейно независимы, так что ее ранг  $r$  равен  $n$ . Тогда, согласно 3F, матрица  $A^T A$  размера  $n \times n$  также будет иметь ранг, равный  $n$ . Следовательно, эта матрица обратима.

|| 3G. Если столбцы матрицы  $A$  линейно независимы, так что  $r = n$ , то  $A^T A$  является квадратной симметрической обратной матрицей.

Легко привести пример такой матрицы:

$$A = \begin{bmatrix} 1 & 4 \\ 2 & 0 \\ 3 & 1 \end{bmatrix}, \quad A^T A = \begin{bmatrix} 14 & 7 \\ 7 & 17 \end{bmatrix}.$$

Обе матрицы имеют линейно независимые столбцы, ранг равен 2, и нуль-пространство отсутствует.

**Упражнение 3.1.1.** (а) Пусть заданы два произвольных положительных числа  $x$  и  $y$ . Рассмотрим векторы  $b = (\sqrt{x}, \sqrt{y})$  и  $a = (\sqrt{y}, \sqrt{x})$ . Используя неравенство Шварца, сравнить среднее арифметическое чисел  $x$  и  $y$  с их средним геометрическим.

(б) Пусть задан вектор с началом в нуле и концом в точке  $x$  и к нему добавлен вектор длины  $\|y\|$ , соединяющий точку  $x$  с  $x+y$ . Третья сторона треугольника образуется отрезком, соединяющим начало координат с этой точкой  $x+y$ . Тогда неравенство треугольника утверждает, что это расстояние не может превосходить сумму первых двух:

$$\|x+y\| \leq \|x\| + \|y\|.$$

Возводя обе части в квадрат, привести это к неравенству Шварца.

**Упражнение 3.1.2.** Проверить справедливость теоремы Пифагора для треугольника  $Obp$ , изображенного на рис. 3.3, используя формулу (5) для длины стороны  $bp$ .

**Упражнение 3.1.3.** Найти на луче, проходящем через начало координат и через точку  $a = (1, 1, 1)$ , точку  $p$ , ближайшую к точке  $b = (2, 4, 4)$ . Найти также точку, ближайшую к  $a$ , на луче, проходящем через начало координат и точку  $b$ .

**Упражнение 3.1.4.** Объяснить, почему неравенство Шварца превращается в равенство, когда точки  $a$  и  $b$  лежат на одной прямой, проходящей через начало координат, и только в этом случае. Что произойдет, если они будут расположены по разные стороны от начала координат?

**Упражнение 3.1.5.** Какой угол образует вектор  $(1, 1, \dots, 1)$  с осями координат в  $n$ -мерном пространстве?

**Упражнение 3.1.6.** Неравенство Шварца доказывается в одну строчку, если векторы  $a$  и  $b$  заранее нормированы к единичной длине:

$$|a^T b| = \left| \sum a_j b_j \right| \leq \sum |a_j| |b_j| \leq \sum \frac{|a_j|^2 + |b_j|^2}{2} = \frac{1}{2} + \frac{1}{2} = \|a\| \|b\|.$$

Провести доказательство справедливости второго неравенства.

**Упражнение 3.1.7.** Транспонирование матрицы  $AA^{-1} = I$  показывает, что операции транспонирования и обращения перестановочны:  $(A^{-1})^T = (A^T)^{-1}$ . Показать, что если матрица  $A$  является симметрической, то симметрической будет и  $A^{-1}$ .

**Упражнение 3.1.8.** Построить симметрические матрицы  $A$  и  $B$  размера  $2 \times 2$ , произведение  $AB$  которых не является симметрической матрицей. Заметим, что если матрицы  $A$  и  $B$  коммутируют, то их произведение будет симметрической матрицей:  $(AB)^T = B^T A^T = BA = AB$ .

**Упражнение 3.1.9.** Если матрица  $A$  имеет ранг  $r$ , то тот же ранг имеют и матрицы  $A^T$  (ранг по строкам равен рангу по столбцам),  $A^T A$  (согласно ЗФ) и  $AA^T$  (согласно ЗФ, примененному к матрице  $A^T$ ). Показать, что матрица  $AA^T$  может не быть обратимой, даже если  $A^T A$  является таковой.

**Упражнение 3.1.10.** Молекула метана  $\text{CH}_4$  устроена таким образом, что атом углерода находится в центре правильного тетраэдра с четырьмя атомами водорода в вершинах. Если вершины находятся в точках  $(0, 0, 0)$ ,  $(1, 1, 0)$ ,  $(1, 0, 1)$  и  $(0, 1, 1)$  — при этом шесть ребер имеют длину, равную  $\sqrt{2}$ , так что тетраэдр является правильным, — то чему равен косинус угла между лучами, идущими из центра  $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$  в вершины? (Сам этот угол, хорошо известный химикам, равен примерно  $109,5^\circ$ .)

### § 3.2. ПРОЕКЦИИ НА ПОДПРОСТРАНСТВА И АППРОКСИМАЦИИ ПО МЕТОДУ НАИМЕНЬШИХ КВАДРАТОВ

До сих пор система  $Ax = b$  либо имела решение, либо не имела его. Если вектор  $b$  не принадлежит пространству столбцов  $\mathfrak{R}(A)$ , то система является несовместной и метод исключения Гаусса не работает. Это почти всегда происходит в системах  $m$  уравнений ( $m > 1$ ) с одним неизвестным. Например, система

$$\begin{aligned} 2x &= b_1, \\ 3x &= b_2, \\ 4x &= b_3 \end{aligned} \quad (11)$$

будет иметь решение только в случае, когда отношение правых частей равно 2:3:4. Решение  $x$ , если только оно существует, безусловно является единственным, но оно будет существовать лишь когда вектор  $b$  лежит на той же прямой, что и вектор

$$a = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}.$$

Хотя несовместные системы и являются неразрешимыми, они часто возникают на практике, и их нужно уметь решать. Одна возможность состоит в отыскании вектора  $x$  из части системы и игнорировании оставшихся уравнений. Но такой прием трудно обосновать, если все  $m$  уравнений имеют один и тот же «источник». Вместо точного решения одних уравнений и допущения больших ошибок в других целесообразно выбрать вектор  $x$  таким образом, чтобы минимизировать *среднюю ошибку для всех  $m$  уравнений*. Существует много способов для выбора такого осреднения, но наиболее удобный заключается в использовании суммы квадратов

$$E^2 = (2x - b_1)^2 + (3x - b_2)^2 + (4x - b_3)^2. \quad (12)$$

Если уравнение  $ax = b$  имеет точное решение, то минимальная ошибка равна  $E = 0$ . В более вероятном случае, когда вектор  $b$

не пропорционален  $a$ , функция  $E^2$  будет параболой с минимумом в точке, где

$$\frac{dE^2}{dx} = 2[(2x - b_1)2 + (3x - b_2)3 + (4x - b_3)4] = 0. \quad (13)$$

Разрешая относительно  $x$ , получаем приближенное решение системы  $ax = b$  по методу наименьших квадратов в виде

$$\bar{x} = \frac{2b_1 + 3b_2 + 4b_3}{2^2 + 3^2 + 4^2} = \frac{2b_1 + 3b_2 + 4b_3}{29}. \quad (14)$$

Нетрудно дать общую формулу для произвольных  $a \neq 0$  и правой части  $b$ . Во-первых, ошибка  $E$  представляет собой длину вектора  $ax - b$ :

$$E = \|ax - b\| = [(a_1x - b_1)^2 + \dots + (a_mx - b_m)^2]^{1/2}. \quad (15)$$

Возводя в квадрат, получаем параболу

$$E^2 = (ax - b)^T(ax - b) = a^T a x^2 - 2a^T b x + b^T b. \quad (16)$$

Она имеет минимум в точке, где

$$\frac{dE^2}{dx} = 2a^T a x - 2a^T b = 0.$$

**ЗН.** Решение по методу наименьших квадратов одномерной задачи  $ax = b$  имеет вид

$$\bar{x} = \frac{a^T b}{a^T a}. \quad (17)$$

Геометрически это решение совпадает с проекцией:  $p = \bar{x}a$  является точкой на прямой, определенной вектором  $a$ , ближайшей к  $b$ .

Таким образом, мы возвращаемся к геометрической интерпретации задачи о наименьших квадратах — к минимизации расстояния. Дифференцируя параболу  $E^2$  в форме (16) и приравнявая производную нулю, мы фактически использовали аппарат анализа для подтверждения геометрических рассуждений предыдущего параграфа. Прямая, соединяющая точки  $b$  и  $p$ , должна быть перпендикулярна направлению вектора  $a$ , откуда находится значение  $\bar{x}$ :

$$a^T(b - \bar{x}a) = a^T b - \frac{a^T b}{a^T a} a^T a = 0. \quad (18)$$

В заключение коснемся вырожденного случая  $a = 0$ . В этой ситуации все числа, кратные  $a$ , равны нулю и прямая вырождается в точку. Поэтому проекцией  $b$  на вектор  $a$  является точка  $p = 0$ . Но формула (17) для  $\bar{x}$  превращается в неопределенность вида

0/0, что отражает полную произвольность в выборе множителя  $\bar{x}$ . Фактически все значения  $x$  будут давать одну и ту же ошибку  $E = \|Ox - b\|$ , так что  $E^2$  превращается из параболы в горизонтальную прямую и не существует единственной минимизирующей точки  $\bar{x}$ . Одна из целей псевдообращения, которое будет описано в § 3.4, заключается в придании числу  $\bar{x}$  определенного значения. В данной ситуации этим значением будет точка  $\bar{x} = 0$ , которая по крайней мере выглядит более «симметричной», чем все остальные.

**Упражнение 3.2.1.** Предположим, что мы измеряем вес пациента в четырех различных случаях, и он оказывается равным  $b_1 = 150$ ,  $b_2 = 153$ ,  $b_3 = 150$ ,  $b_4 = 151$ . Какое значение является наилучшим в смысле наименьших квадратов для указания веса пациента?

**Упражнение 3.2.2.** Найти наилучшее решение  $\bar{x}$  в смысле метода наименьших квадратов для задачи  $3x = 10$ ,  $4x = 5$ .

### МНОГОМЕРНЫЕ ЗАДАЧИ О НАИМЕНЬШИХ КВАДРАТАХ

Теперь мы готовы к следующему шагу, а именно к проектированию вектора  $b$  на подпространство, а не на прямую. Эта геометрическая задача возникает следующим образом. Пусть мы вновь начинаем с системы  $Ax = b$ , но теперь  $A$  является матрицей размера  $m \times n$ , т. е. вместо одного неизвестного и одного вектор-столбца  $a$  матрица содержит  $n$  столбцов. Мы должны по-прежнему считать, что число  $m$  наблюдений больше числа неизвестных  $n$ , так что следует ожидать, что система  $Ax = b$  окажется несовместной. Вероятно, не существует вектора  $x$ , который бы точно соответствовал  $b$ . Другими словами, вектор  $b$ , вероятно, не является комбинацией столбцов матрицы  $A$ .

Задача вновь состоит в выборе  $\bar{x}$ , минимизирующего ошибку, и эта минимизация снова будет производиться по методу наименьших квадратов. Ошибка равняется  $E = \|Ax - b\|$ , и она в точности представляет собой расстояние от  $b$  до точки  $Ax$ , лежащей в пространстве столбцов матрицы  $A$ . (Вспомним, что  $Ax$  есть линейная комбинация столбцов с коэффициентами  $x_1, \dots, x_n$ .) Поэтому отыскание решения  $\bar{x}$  по методу наименьших квадратов, означающее минимизацию ошибки  $E$ , эквивалентно отысканию точки  $p = A\bar{x}$ , ближайшей к точке  $b$  по сравнению с остальными точками из пространства столбцов.

Для определения вектора  $\bar{x}$  можно использовать либо геометрию, либо анализ. Мы воспользуемся первой из возможностей: точка  $p$  должна быть «проекцией  $b$  на пространство столбцов» и вектор невязки  $A\bar{x} - b$  должен быть ортогонален к этому

**пространству** (рис. 3.4). Ортогональность к пространству состоит в следующем. Каждый вектор в пространстве столбцов матрицы  $A$  является линейной комбинацией столбцов с некоторыми

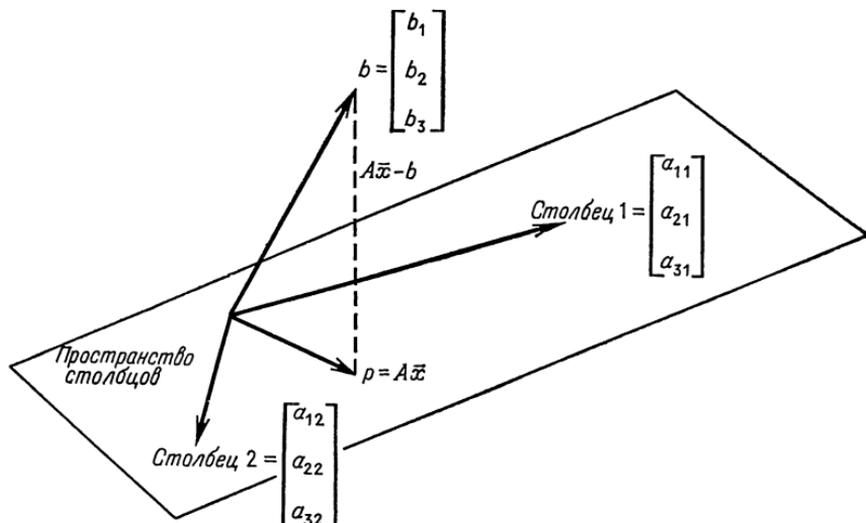


Рис. 3.4. Проекция на пространство столбцов матрицы размера  $3 \times 2$ .

коэффициентами  $y_1, \dots, y_n$ . Другими словами, это вектор вида  $Ay$ . Для всех  $y$  эти векторы в плоскости должны быть перпендикулярны вектору невязки  $A\bar{x} - b$ :

$$(Ay)^T (A\bar{x} - b) = 0, \text{ или } y^T [A^T A\bar{x} - A^T b] = 0. \quad (19)$$

Это справедливо для произвольного вектора  $y$ , откуда следует, что вектор в квадратных скобках должен быть нулевым,  $A^T A\bar{x} - A^T b = 0$ . Таким образом, геометрические соображения приводят нас к фундаментальным уравнениям метода наименьших квадратов.

**31.** Решение по методу наименьших квадратов для несовместной системы  $Ax = b$ , состоящей из  $t$  уравнений с  $n$  неизвестными, удовлетворяет соотношению

$$A^T A\bar{x} = A^T b. \quad (20)$$

Оно известно под названием «**нормальных уравнений**». Если столбцы матрицы  $A$  являются линейно независимыми, то по ЗГ матрица  $A^T A$  обратима и единственное решение дается формулой

$$\bar{x} = (A^T A)^{-1} A^T b. \quad (21)$$

Проекция точки  $b$  на пространство столбцов, таким образом, имеет вид

$$p = A\bar{x} = A(A^T A)^{-1} A^T b. \quad (22)$$

Матрица  $B = (A^T A)^{-1} A^T$ , возникающая в формуле (21),  $\bar{x} = Bb$ , является одной из левых обратных к матрице  $A$ :  $BA = = (A^T A)^{-1} A^T A = I$ . Существование этой левой обратной гарантируется условием единственности из теоремы 2Q, поскольку столбцы матрицы  $A$  являются линейно независимыми.

Для иллюстрации рассмотрим задачу, которую можно решить либо интуитивно, либо с использованием формулы (22). Пусть матрица  $A$  и вектор  $b$  задаются формулами

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \\ 3 \\ 9 \end{bmatrix}.$$

Пространство столбцов матрицы  $A$  легко определяется, поскольку оба столбца матрицы оканчиваются нулями: это пространство совпадает с плоскостью  $xy$  в трехмерном пространстве. Проекция точки  $b$  на эту плоскость будет иметь те же самые компоненты по осям  $x$  и  $y$ , равные 4 и 3 соответственно, но компонента  $z$  исчезнет, так что  $p = (4, 3, 0)^T$ . То же самое следует и из формулы (22):

$$A^T A = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 5 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 7 \\ 7 & 29 \end{bmatrix}, \quad (A^T A)^{-1} = \frac{1}{9} \begin{bmatrix} 20 & -7 \\ -7 & 2 \end{bmatrix},$$

$$p = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 0 & 0 \end{bmatrix} \frac{1}{9} \begin{bmatrix} 29 & -7 \\ -7 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 2 & 5 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \\ 9 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 0 \end{bmatrix}.$$

Можно сразу обратиться и к несовместной системе

$$\begin{aligned} u + 2v &= 4, \\ u + 5v &= 3, \\ 0u + 0v &= 9. \end{aligned}$$

В этом случае лучшее из того, что мы можем сделать, это решить систему из первых двух уравнений (откуда находятся компоненты  $\bar{u}$  и  $\bar{v}$  вектора  $\bar{x}$ ) и игнорировать третье уравнение; ошибка в этом уравнении всегда будет равняться 9.

Отметим, что если вектор  $b$  принадлежит пространству столбцов матрицы, т. е. выражается в виде некоторой комбинации  $b = Ax$  столбцов матрицы, то проекция принимает вид

$$p = A (A^T A)^{-1} A^T A x = Ax = b. \quad (23)$$

Таким образом, точка  $p$  совпадает с  $b$ , как и должно быть.

Противоположный случай возникает, когда вектор  $b$  ортогонален всем столбцам матрицы  $A$ . В этом случае  $b$  не только не лежит в подпространстве столбцов, но даже ортогонален ему. Тогда геометрические соображения показывают, что ближайшей точкой из подпространства является начало координат: вектор  $b$  имеет нулевую компоненту в разложении по подпространству и его проекция есть  $p=0$ . Это подтверждается и вычислением  $\bar{x}$ : столбцы матрицы  $A$  совпадают со строками  $A^T$ , и если вектор  $b$  ортогонален всем им, то  $A^T b = 0$ . Поэтому

$$\bar{x} = (A^T A)^{-1} A^T b = 0 \quad \text{и} \quad p = A \bar{x} = 0.$$

Заметим, что при проектировании на прямую мы приходим к формуле (17). Матрица  $A$  превращается в столбец  $a$ , матрица  $A^T A$  есть число  $a^T a$  и  $\bar{x}$  в формуле (21) равен  $a^T b / a^T a$ , как и должно быть.

**Упражнение 3.2.3.** Используя нормальные уравнения, найти наилучшее решение в смысле метода наименьших квадратов для несовместной системы

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Проделать то же самое для системы  $x=1$ ,  $x=3$ ,  $x=5$ , т. е.

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} [x] = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}.$$

**Упражнение 3.2.4.** (а) Пусть

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, \quad x = \begin{bmatrix} u \\ v \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}.$$

Записать  $E^2 = \|Ax - b\|^2$  и приравнять нулю производные по  $u$  и  $v$ . Сравнить получившиеся уравнения с  $A^T A \bar{x} = A^T b$ . Это сравнение должно показать, что нормальные уравнения могут быть получены как из геометрии, так и из анализа, и они связаны непосредственно с минимизацией величины  $E^2$ .

- (б) Найти решение  $\bar{x}$  и проекцию  $p = A \bar{x}$  точки  $b$  на пространство столбцов.  
 (с) Проверить, что вектор  $b - p$  ортогонален к столбцам матрицы  $A$ .

### МАТРИЦЫ ПРОЕКТИРОВАНИЯ

Выше было установлено, что ближайшая к  $b$  точка дается формулой  $p = A (A^T A)^{-1} A^T b$ . Эта формула является матричной записью для перпендикуляра, опущенного из точки  $b$  на пространство столбцов матрицы  $A$ . Матрица, описывающая эту конструкцию, называется **матрицей проектирования** и обозначается через  $P$ :

$$P = A (A^T A)^{-1} A^T. \quad (24)$$

Эта матрица проектирует любой вектор  $b$  на пространство столбцов матрицы  $A^1$ ). Другими словами, вектор  $p = Pb$  является компонентой вектора  $b$  в пространстве столбцов матрицы, а ошибка  $b - Pb$  является компонентой в его ортогональном дополнении. (Иначе говоря,  $I - P$  также является матрицей проектирования: она проектирует произвольный вектор  $b$  на это ортогональное дополнение, и соответствующая проекция равняется  $(I - P)b = b - Pb$ .) Таким образом, мы имеем матричную формулу для разбиения вектора на две взаимно перпендикулярные составляющие:  $Pb$  принадлежит пространству столбцов  $\mathfrak{N}(A)$ , а другая составляющая  $(I - P)b$  — левому нуль-пространству  $\mathfrak{N}(A^T)$ , которое является ортогональным дополнением к пространству столбцов.

Введенные матрицы проектирования можно рассматривать как с геометрической, так и с алгебраической точек зрения. Они представляют собой семейство матриц с весьма специфическими свойствами и будут использоваться ниже как основные блоки при построении всех симметрических матриц. Поэтому прервем на время изложение метода наименьших квадратов и обратимся к этим свойствам.

**3J.** Матрица проектирования  $P = A(A^T A)^{-1} A^T$  обладает двумя основными свойствами:

- (i) она является идемпотентной, т. е.  $P^2 = P$ ;
- (ii) она является симметрической, т. е.  $P = P^T$ .

И обратно, любая матрица с этими двумя свойствами представляет собой матрицу проектирования на свое пространство столбцов.

Доказательство. Справедливость равенства  $P^2 = P$  легко устанавливается с помощью геометрических соображений. Если мы начинаем с произвольного вектора  $b$ , то вектор  $Pb$  будет лежать в подпространстве, на которое осуществляется проектирование. Поэтому при повторном проектировании, в результате которого мы получаем вектор  $P(Pb)$ , или  $P^2b$ , ничего не изменяется: вектор уже принадлежит подпространству и  $Pb = P^2b$  для всех  $b$ . Алгебраически это можно показать следующим образом:

$$P^2 = A(A^T A)^{-1} A^T A(A^T A)^{-1} A^T = A(A^T A)^{-1} A^T = P.$$

Для доказательства симметричности матрицы  $P$  перемножим транспонированные матрицы в обратном порядке и воспользуемся тождеством  $(B^{-1})^T = (B^T)^{-1}$ , которое приведено в упр. 3.1.7, для матрицы  $B = A^T A$ :

$$P^T = (A^T)^T ((A^T A)^{-1})^T A^T = A((A^T A)^T)^{-1} A^T = A(A^T A)^{-1} A^T = P.$$

<sup>1)</sup> Есть опасность спутать эти матрицы с матрицами перестановки, также обозначаемыми через  $P$ . Однако опасность эта весьма мала, поскольку мы будем избегать одновременного использования обеих матриц на одной странице.

Чтобы показать это геометрически, представим себе два произвольных вектора  $b$  и  $c$ . Если  $b$  проецируется обычным способом, давая в результате вектор  $Pb$ , а  $c$  проецируется на ортогональное дополнение, давая вектор  $(I - P)c$ , то эти две проекции являются взаимно ортогональными:

$$(Pb)^T (I - P)c = b^T P^T (I - P)c = 0.$$

Поскольку это справедливо для любых векторов  $b$  и  $c$ , получаем, что

$$P^T (I - P) = 0, \text{ или } P^T = P^T P, \text{ или } P = (P^T P)^T = P^T P.$$

Следовательно,  $P^T = P$  и матрица  $P$  является симметрической.

Для доказательства обратного мы должны, используя свойства (i) и (ii), показать, что  $P$  является матрицей проектирования на свое пространство столбцов. Это пространство состоит из всех комбинаций  $Pc$  столбцов матрицы  $P$ . Для любого вектора  $b$  вектор  $Pb$  обязательно лежит в этом пространстве, поскольку из правила умножения матриц следует, что он равен линейной комбинации столбцов матрицы  $P$ . Кроме того, можно показать, что *вектор ошибки*  $b - Pb$  ортогонален этому пространству: для любого вектора  $Pc$  из пространства столбцов свойства (i) и (ii) дают

$$(b - Pb)^T Pc = b^T (I - P)^T Pc = b^T (P - P^2)c = 0. \quad (25)$$

Поскольку вектор  $b - Pb$  ортогонален пространству столбцов, он и является искомым перпендикуляром, так что  $P$  оказывается матрицей проектирования.

**Упражнение 3.2.5.** Пусть задан базис  $u_1, u_2$  в подпространстве  $S$  трехмерного пространства и вектор  $b$ , не принадлежащий  $S$ :

$$u_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}.$$

С помощью матрицы  $A$  со столбцами  $u_1$  и  $u_2$  найти матрицу  $P$ , осуществляющую проектирование на подпространство  $S$ . Вычислить проекции вектора  $b$  на  $S$  и на ортогональное дополнение  $S^\perp$ .

**Упражнение 3.2.6.** (а) Показать, что если  $P$  является матрицей проектирования, т. е. обладает свойствами (i) и (ii), то матрица  $I - P$  также будет обладать этими свойствами.

(б) Показать, что если  $P_1$  и  $P_2$  являются матрицами проектирования на подпространства  $S_1$  и  $S_2$  и если  $P_1 P_2 = P_2 P_1 = 0$ ,  $P = P_1 + P_2$  также будет матрицей проектирования. Привести пример для матриц порядка 2.

**Упражнение 3.2.7.** Пусть  $P$  — матрица проектирования на прямую в плоскости  $xy$ . Изобразить на рисунке, как будет действовать на произвольный вектор «матрица отражения»  $H = I - 2P$ . Объяснить геометрически и алгебраически, почему  $H^2 = I$ .

**Упражнение 3.2.8.** Показать, что если вектор  $u$  имеет единичную длину, то матрица  $P = uu^T$  ранга один будет матрицей проектирования, т. е. будет обладать свойствами (i) и (ii). Если  $u = a/\|a\|$ , то  $P$  — матрица проектирования на прямую, определенную вектором  $a$ , и вектор  $Pb$  оканчивается в точке  $p = \frac{a}{\|a\|^2} a$ : матрицы проектирования единичного ранга в точности соответствуют одномерным задачам о наименьших квадратах.

**Упражнение 3.2.9.** Какая матрица размера  $2 \times 2$  проектирует плоскость  $xy$  на ось  $y$ ?

### ПОДГОНКА ДАННЫХ МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ

Пусть мы выполняем серию экспериментов и предполагаем, что выходные данные  $y$  почти линейно зависят от входных данных  $t$ , т. е.  $y = C + Dt$ , например:

(1) Через определенные промежутки времени мы измеряем расстояние до космической станции, летящей к Марсу. В этом случае  $t$  есть время,  $y$  есть расстояние и, если двигательная установка выключена, а гравитационные силы невелики, корабль будет двигаться почти с постоянной скоростью  $v$ :  $y = y_0 + vt$ .

(2) Мы можем изменять нагрузку, приложенную к конструкции, и контролировать возникающие в результате напряжения. В этом случае  $t$  является нагрузкой, а  $y$  — показаниями измерителя напряжений. Тогда если нагрузка не так велика, чтобы материал переходил в пластическое состояние, то  $y$  и  $t$  будут связаны обычным линейным соотношением  $y = C + Dt$  теории упругости.

(3) В вопросах экономики и бизнеса рассматриваются сложные взаимные связи между стоимостью продукции, объемом производства, ценами и прибылью. Несмотря на их сложность, в определенных пределах эти связи могут быть близки к линейным. Пусть, например, выпуск  $t_1$  экземпляров журнала обходится в сумму  $y_1$ , на следующей неделе выпуск  $t_2$  экземпляров обходится в  $y_2$  и т. д. Тогда издатель может оценить стоимость на следующую неделю, предположив, что  $y = C + Dt$  и оценив коэффициенты  $C$  и  $D$  по уже имеющимся данным. Коэффициент  $D$ , равный стоимости каждого дополнительного экземпляра и называемый предельной стоимостью производства, обычно важнее для издателя, чем накладные расходы  $C$ .

Вопрос состоит в том, как вычислить коэффициенты  $C$  и  $D$  по результатам этих экспериментов. Если соотношение действительно является линейным и ошибки в экспериментах отсутствуют, то вопрос решается просто: два измерения величины  $y$  при различных значениях  $t$  определяют прямую  $y = C + Dt$  и результаты всех последующих измерений будут находиться на этой прямой. Но при наличии ошибки эти точки уже не будут попадать на прямую, и мы должны «осреднять» результаты всех эксперимен-

тов для отыскания оптимальной прямой. Причем ее не следует путать с прямой, на которую мы раньше проектировали вектор  $b$ ! Фактически, поскольку нужно определить две величины  $C$  и  $D$ , мы будем заниматься проектированием на двумерное подпространство. Задача о наименьших квадратах определяется результатами экспериментов

$$\begin{aligned} C + Dt_1 &= y_1, \\ C + Dt_2 &= y_2, \\ &\vdots \\ C + Dt_m &= y_m. \end{aligned} \tag{26}$$

Эта система является переопределенной, и при наличии ошибок она не будет иметь решения. Отметим, что вектор неизвестных  $x$  имеет две компоненты  $C$  и  $D$ :

$$\begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \quad \text{или} \quad Ax = b. \tag{27}$$

Наилучшее в смысле метода наименьших квадратов решение минимизирует сумму квадратов ошибок; числа  $\bar{C}$  и  $\bar{D}$  выбираются из условия минимизации величины

$$E^2 = \|b - Ax\|^2 = (y_1 - C - Dt_1)^2 + \dots + (y_m - C - Dt_m)^2.$$

Пользуясь матричной терминологией, мы выбираем такой вектор  $\bar{x}$ , что точка  $p = A\bar{x}$  является ближайшей к  $b$ . Из всех прямых вида  $y = C + Dt$  мы выбираем ту, которая наилучшим образом приближает опытные данные (рис. 3.5). На этом рисунке ошибками являются расстояния по вертикали до прямой, равные  $y - C - Dt$ , и именно эти расстояния мы возводим в квадрат, суммируем и минимизируем.

**Пример.** Пусть заданы результаты четырех измерений (рис. 3.5):

$$\begin{aligned} y = 0 & \text{ при } t = 0, & y = 1 & \text{ при } t = 1, \\ y = 2 & \text{ при } t = 3, & y = 5 & \text{ при } t = 4. \end{aligned}$$

Отметим, что значения  $t$  не обязательно берутся через равные интервалы. Экспериментатор может выбирать любые подходящие значения (в том числе и отрицательные, если это допускается

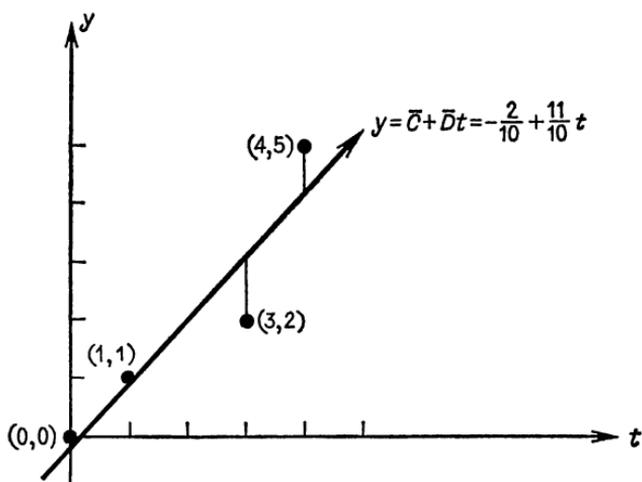


Рис. 3.5. Аппроксимация прямой линией.

экспериментом), не изменяя при этом математических формулировок. В данном случае переопределенная система  $Ax = b$  имеет вид

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 5 \end{bmatrix}.$$

Нам также понадобятся матрица  $A^T A$  и обратная к ней:

$$A^T A = \begin{bmatrix} 4 & 8 \\ 8 & 26 \end{bmatrix}, \quad (A^T A)^{-1} = \frac{1}{20} \begin{bmatrix} 13 & -4 \\ -4 & 2 \end{bmatrix}.$$

Тогда решение  $\bar{x} = (A^T A)^{-1} A^T b$  по методу наименьших квадратов принимает вид

$$\bar{x} = \begin{bmatrix} \bar{C} \\ \bar{D} \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 13 & -4 \\ -4 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} -\frac{2}{10} \\ \frac{11}{10} \end{bmatrix}.$$

Таким образом, оптимальная прямая задается уравнением  $y = -\frac{2}{10} + \frac{11}{10}t$ .

*Замечание.* Метод наименьших квадратов не предполагает, что мы должны приближать экспериментальные данные лишь с помощью прямых линий. Во многих экспериментах связи могут

быть нелинейными, и было бы глупо искать для этих задач линейные соотношения. Пусть, например, мы работаем с радиоактивным материалом. Тогда выходными данными  $y$  являются показания счетчика Гейгера в различные моменты времени  $t$ . Пусть наш материал представляет собой смесь двух радиоактивных веществ, и мы знаем период полураспада каждого из них, но не знаем, в какой пропорции эти вещества смешаны. Если обозначить их количества через  $C$  и  $D$ , то показания счетчика Гейгера будут вести себя подобно сумме двух экспонент<sup>1)</sup>, а не как прямая:

$$y = Ce^{-\lambda t} + De^{-\mu t}. \quad (28)$$

На практике, поскольку радиоактивность измеряется дискретно и через различные промежутки времени, показания счетчика не будут точно соответствовать формуле (28). Вместо этого мы имеем серию показаний  $y_1, \dots, y_m$  в различные моменты времени  $t_1, \dots, t_m$ , и соотношение (28) выполняется лишь приближенно:

$$\begin{aligned} Ce^{-\lambda t_1} + De^{-\mu t_1} &\approx y_1, \\ &\vdots \\ &\vdots \\ Ce^{-\lambda t_m} + De^{-\mu t_m} &\approx y_m. \end{aligned} \quad (29)$$

Если мы имеем более двух показаний,  $m > 2$ , то разрешить эту систему относительно  $C$  и  $D$  практически невозможно. Но при использовании метода наименьших квадратов мы получим оптимальные значения  $\bar{C}$  и  $\bar{D}$ .

Ситуация будет совершенно иной, если нам известны количества веществ  $C$  и  $D$  и нужно отыскать коэффициенты  $\lambda$  и  $\mu$ . Это *нелинейная задача о наименьших квадратах*, и решить ее значительно труднее. Мы по-прежнему будем находить величину  $E^2$ , равную сумме квадратов ошибок, и минимизировать ее. Но  $E^2$  уже не будет многочленом второй степени относительно  $\lambda$  и  $\mu$ , так что приравнивание нулю производных не будет давать линейных уравнений для отыскания оптимальных  $\bar{\lambda}$  и  $\bar{\mu}$ . В упражнениях мы ограничимся рассмотрением только линейного случая метода наименьших квадратов.

**Упражнение 3.2.10.** Показать, что оптимальное приближение по методу наименьших квадратов набора измерений  $y_1, \dots, y_m$  горизонтальной прямой, т. е. постоянной функцией  $y=C$ , представляет собой среднее арифметическое этих измерений:

$$C = \frac{y_1 + \dots + y_m}{m}.$$

<sup>1)</sup> В статистике и экономике это соответствует двум продуктам, которые выпускаются или уничтожаются с вероятностями, определяемыми законом Пуассона. В теории популяций числа  $\lambda$  и  $\mu$  будут отрицательными, если рождаемость превышает смертность.

(Ср. с упр. 3.2.1.) Говоря языком статистики, выбор  $\bar{y}$ , минимизирующий функцию  $E^2 = (y_1 - y)^2 + \dots + (y_m - y)^2$ , представляет собой *среднее* значение выборки, а величина  $E^2$  называется *дисперсией* и обозначается через  $\sigma^2$ .

**Упражнение 3.2.11.** Для следующего набора измерений найти оптимальное линейное приближение и изобразить его:

$$\begin{array}{ll} y=2 & \text{при } t=-1, \\ y=-3 & \text{при } t=1, \end{array} \quad \begin{array}{ll} y=0 & \text{при } t=0, \\ y=-5 & \text{при } t=2. \end{array}$$

**Упражнение 3.2.12.** Пусть мы приближаем указанные выше результаты измерений не с помощью прямой, а с использованием параболы  $y=C+Dt+Et^2$ . Для несовместной системы  $Ax=b$ , образованной в результате четырех замеров, выписать матрицу коэффициентов  $A$ , вектор неизвестных  $x$  и вектор исходных данных  $b$ . Вычислять оптимальный вектор  $\bar{x}$  не обязательно.

### § 3.3. ОРТОГОНАЛЬНЫЕ БАЗИСЫ, ОРТОГОНАЛЬНЫЕ МАТРИЦЫ И ОРТОГОНАЛИЗАЦИЯ ГРАМА — ШМИДТА

Мы уже пытались показать важность свойства ортогональности при решении задач методом наименьших квадратов и продолжаем делать это. Но на самом деле свойство ортогональности является значительно более важным и содержательно более глубоким, чем это может показаться из его применения к методу наименьших квадратов. Каждый раз, когда я представляю себе плоскость  $xu$  или трехмерное пространство, воображение обязательно добавляет к картине набор координатных осей. Они дают нам отправную точку, называемую началом координат, причём *эти воображаемые координатные оси всегда являются взаимно ортогональными*. Выбирая базис в плоскости  $xu$  (что эквивалентно выбору координатных осей), мы стремимся выбрать ортогональный базис.

Если понятие базиса является одним из основных моментов при соединении геометрии векторных пространств с алгеброй, то выделение ортогонального базиса также легко объяснить. Базис необходим для перехода от геометрической конструкции к алгебраическим вычислениям, а ортогональный базис нужен нам для того, чтобы сделать эти вычисления более простыми. Существует еще одно ограничение, введение которого делает базис почти оптимальным: мы начинаем с набора взаимно ортогональных векторов и нормализуем их к векторам единичной длины. Это означает, что каждый вектор  $v$  из рассматриваемого набора делится на свою собственную длину и заменяется на вектор  $v/\|v\|$ . В результате *ортогональный базис превращается в ортонормированный*.

3К. Базис  $v_1, \dots, v_k$  называется **ортонормированным**, если

$$v_i^T v_j = \begin{cases} 0 & \text{при } i \neq j \text{ (свойство ортогональности),} \\ 1 & \text{при } i = j \text{ (нормированность).} \end{cases} \quad (30)$$

Наиболее важным примером этого является так называемый **стандартный базис**. В плоскости  $xy$  или в пространстве  $\mathbf{R}^n$  мы не только задаем взаимно перпендикулярные оси координат, но также отмечаем на каждой оси вектор единичной длины:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \quad \dots, \quad e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}.$$

Этот ортонормированный базис никоим образом не является единственным, поскольку мы можем поворачивать весь набор осей, оставляя неизменными прямые углы между ними. Ниже будут введены соответствующие этим поворотам матрицы вращений, называемые также ортогональными матрицами. С другой стороны, если мы рассматриваем не все пространство  $\mathbf{R}^n$ , а лишь его подпространство, то «стандартные векторы»  $e_i$  могут лежать вне этого подпространства. В этом случае существование даже одного ортонормированного базиса не столь очевидно. Но мы покажем, что такой базис всегда существует и что он легко может быть построен по произвольному базису. Это построение, переводящее неортогональный набор осей в ортогональный, известно под названием *ортогонализации Грама — Шмидта*.

Итак, три основных пункта настоящего параграфа состоят в следующем:

- (1) Решение методом наименьших квадратов системы  $Ax = b$ , когда столбцы матрицы  $A$  являются ортонормированными.
- (2) Определение и свойства ортогональных матриц.
- (3) Процесс ортогонализации Грама — Шмидта и его интерпретация как нового способа разложения матриц.

#### ПРОЕКЦИИ И НАИМЕНЬШИЕ КВАДРАТЫ: ОРТОНОРМИРОВАННЫЙ СЛУЧАЙ

Пусть  $A$  — матрица размера  $m \times n$ , и пусть ее столбцы являются ортонормированными. Тогда эти столбцы являются линейно независимыми, и мы знаем матрицу проектирования на пространство столбцов, а также решение  $\bar{x}$  по методу наимень-



Сопрягающий член  $(A^T A)^{-1}$  исчезает, так что  $p$  превращается в сумму  $n$  отдельных проекций.

Таким образом, мы получаем пять основных формул этой главы, которые полезно выписать вместе:

1.  $Ax = b$  — исходная система (возможно, несовместная);
2.  $A^T A \bar{x} = A^T b$  — нормальные уравнения для  $\bar{x}$ ;
3.  $p = A \bar{x}$  — проекция вектора  $b$  на пространство столбцов матрицы  $A$ ;
4.  $P = A(A^T A)^{-1} A^T$  — матрица проектирования, дающая вектор  $p = Pb$ ;
5.  $\bar{x} = A^T b$  и  $P = AA^T = a_1 a_1^T + \dots + a_n a_n^T$  — частный случай, когда матрица  $A$  имеет ортонормированные столбцы.

**Пример 1.** Рассмотрим следующий простой, но типичный случай. Пусть точка  $b = (x, y, z)$  проектируется на плоскость  $xy$ . Ясно, что проекцией будет  $p = (x, y, 0)$ , и она представляет собой сумму отдельных проекций на оси  $x$  и  $y$ :

$$a_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad a_1 a_1^T b = \begin{bmatrix} x \\ 0 \\ 0 \end{bmatrix}; \quad a_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad a_2 a_2^T b = \begin{bmatrix} 0 \\ y \\ 0 \end{bmatrix}.$$

Таким образом, матрица проектирования имеет вид

$$P = a_1 a_1^T + a_2 a_2^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \text{или} \quad P \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}.$$

**Пример 2.** Существует один случай линейного приближения данных, когда столбцы матрицы оказываются ортогональными. Если измерения  $b_1$ ,  $b_2$  и  $b_3$  производятся в моменты времени, симметричные относительно точки  $t = 0$  (скажем, при  $t_1 = -1$ ,  $t_2 = 0$  и  $t_3 = 1$ ), то попытка построения приближения  $y = C + Dt$  приводит к трем уравнениям с двумя неизвестными:

$$\begin{aligned} C + Dt_1 &= b_1, & \text{или} & & \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} &= \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}. \\ C + Dt_2 &= b_2, \\ C + Dt_3 &= b_3, \end{aligned}$$

Столбцы этой матрицы взаимно ортогональны. Они не являются ортонормированными, но легко могут быть сделаны таковыми — мы вычисляем их длины, равные  $\sqrt{3}$  и  $\sqrt{2}$ , и производим замену

неизвестных  $c = C\sqrt{3}$ ,  $d = D\sqrt{2}$ :

$$Ax = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} \\ 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

Теперь  $A^T A$  является единичной матрицей и решение по методу наименьших квадратов выписывается непосредственно:

$$\bar{x} = A^T b = \begin{bmatrix} (b_1 + b_2 + b_3)/\sqrt{3} \\ (-b_1 + b_3)/\sqrt{2} \end{bmatrix} = \begin{bmatrix} \bar{c} \\ \bar{d} \end{bmatrix}. \quad (34)$$

Другими словами, коэффициенты  $C$  и  $D$  оптимальной прямой имеют вид

$$C = \frac{c}{\sqrt{3}} = \frac{b_1 + b_2 + b_3}{3}, \quad D = \frac{d}{\sqrt{2}} = \frac{-b_1 + b_3}{2}.$$

Каждая из компонент  $C$  и  $D$  важна сама по себе.  $C$  есть *среднее арифметическое* данных, и этот коэффициент дает наилучшее приближение горизонтальной прямой, а  $D$  дает наилучшее приближение прямой, проходящей через начало координат. *Поскольку столбцы матрицы ортогональны, сумма этих двух слагаемых дает наилучшее приближение в классе всех прямых линий.*

*Замечание.* Изучение специальных свойств ортонормированного базиса часто дает полезный «побочный эффект». Обычный базис легче описать с помощью свойств, которыми он не обладает. Воображение ошибается, представляя каждый базис ортонормированным. То же самое справедливо относительно специального свойства проекций, которое только что было нами отмечено, а именно, что в ортонормированном случае они представляют собой сумму  $n$  одномерных проекций. Интуиция подсказывает, что вектор можно представить в виде суммы его компонент по координатным осям, но что наличие этого представления *зависит от ортогональности осей*. Если ось  $x$  заменяется прямой  $y=x$  и вектор  $(0, 1)$  проектируется на эту прямую и на ось  $y$ , то сумма двух проекций отнюдь не совпадает с исходным вектором  $(0, 1)$ .

**Упражнение 3.3.1.** (а) Выписать четыре уравнения для линейного приближения  $y = C + Dt$  следующих данных:

$$\begin{array}{llll} y = -4 & \text{при } t = -2, & y = -3 & \text{при } t = -1, \\ y = -1 & \text{при } t = 1, & y = 0 & \text{при } t = 2. \end{array}$$

Показать, что столбцы соответствующей матрицы являются ортогональными и нормализовать их. Какой вид имеют неизвестные  $c$  и  $d$  в новой задаче  $Ax = b$ ?



Последнее утверждение об ортонормированности строк немедленно получается из соотношения  $QQ^T = I$ . В произведении  $QQ^T$  участвуют скалярные произведения каждой строки матрицы  $Q$  на каждую другую строку, и, поскольку в результате получается единичная матрица, эти строки являются ортонормированными. Необходимо обратить внимание на этот замечательный факт. Вектор-строки имеют совершенно иные направления, чем вектор-столбцы, но они каким-то образом автоматически оказываются взаимно перпендикулярными при перпендикулярности столбцов.

### Пример 1

$$Q = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, \quad Q^T = Q^{-1} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}.$$

Матрица  $Q$  поворачивает каждый вектор на угол  $\theta$ , а матрица  $Q^T$  поворачивает его в обратном направлении на угол  $-\theta$ .

**Пример 2.** Любая матрица перестановки  $P$  является ортогональной, поскольку ее вектор-столбцы имеют единичную длину и ортогональны друг другу (в силу того, что единицы в столбцах расположены на различных местах):

$$\text{если } P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \text{то } P^{-1} = P^T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

и

$$\text{если } P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \text{то } P^{-1} = P^T = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

В первом случае, где матрица получена путем перестановки лишь двух строк единичной матрицы, будет справедливо дополнительное свойство  $P^{-1} = P$ . В общем случае это не выполняется, как показывает второй пример, где переставлены местами все три строки матрицы. Отметим также, что матрица  $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$  не является матрицей вращения  $Q$  из первого примера, т. е. ее нельзя получить ни при каком значении  $\theta$ . На самом деле матрица  $P$  отображает каждую точку  $(x, y)$  в ее зеркальный образ  $(y, x)$  относительно прямой  $y = x$ , наклоненной под углом  $45^\circ$  к оси  $x$ . Таким образом, нельзя утверждать, что каждая ортогональная матрица  $Q$  представляет собой чистое вращение.

Однако остается одно свойство, которым обладают и матрицы вращений  $Q$ , и матрицы перестановок  $P$ , и вообще все ортогональные матрицы. Причем это свойство является как наиболее важным, так и наиболее характерным для ортогональных матриц:

**3N.** Умножение на ортогональную матрицу  $Q$  сохраняет длины:

$$\|Qx\| = \|x\| \quad \text{для любого вектора } x, \quad (37)$$

а также скалярные произведения:

$$(Qx)^T(Qy) = x^T y \quad \text{для всех векторов } x \text{ и } y. \quad (38)$$

Доказательство получается непосредственно:  $(Qx)^T(Qy) = x^T Q^T Q y = x^T I y = x^T y$ . Если  $y = x$ , то это соотношение принимает вид  $\|Qx\|^2 = \|x\|^2$ , так что длины сохраняются, раз сохраняются скалярные произведения.

**Пример.** Для описанных выше вращений плоскости имеем

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

Длина вектора сохраняется, поскольку

$$(x \cos \theta - y \sin \theta)^2 + (x \sin \theta + y \cos \theta)^2 = x^2 + y^2.$$

**Упражнение 3.3.4.** Пусть  $Q_1$  и  $Q_2$  — ортогональные матрицы, так что для них выполняются соотношения (36). Показать, что матрица  $Q_1 Q_2$  также является ортогональной.

**Упражнение 3.3.5.** Пусть  $u$  — единичный вектор. Показать, что матрица  $Q = I - 2uu^T$  является ортогональной. (Она называется преобразованием Хаусхолдера). Найти матрицу  $Q$ , если  $u = (1, 1, 1)/\sqrt{3}$ .

**Упражнение 3.3.6.** Подобрать третий столбец так, чтобы матрица

$$Q = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} \\ 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & -1/\sqrt{2} \end{bmatrix}$$

была ортогональной. Вектор, ортогональный остальным столбцам, должен быть единичным; в таком случае, насколько мы свободны в его выборе? Проверить, что одновременно становятся ортонормированными и строки матрицы.

**Упражнение 3.3.7.** Записав в явном виде выражение  $v^T v$ , показать, что теорема Пифагора выполняется для любой комбинации  $v = x_1 q_1 + \dots + x_n q_n$  ортонормированных векторов:

$$\|v\|^2 = x_1^2 + \dots + x_n^2.$$

В терминах матриц это означает, что  $v = Qx$ , так что мы получаем новое и более наглядное доказательство того, что длины векторов сохраняются:  $\|Qx\|^2 = \|x\|^2$ .

Упражнение 3.3.8. Является ли ортогональной следующая матрица

$$Q = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} ?$$

### ОРТОГОНАЛИЗАЦИЯ ГРАМА — ШМИДТА

Мы уже видели, какие преимущества несет в себе ортонормированность столбцов матрицы: для обращения матрицы  $Q$  нужно лишь транспонировать ее, и решение задачи о наименьших квадратах  $Ax = b$  сразу получается в виде  $\bar{x} = A^T b$ . Но во всех случаях мы были вынуждены начинать со слов: «Если столбцы матрицы являются ортонормированными...». Теперь мы хотим показать, как *сделать* их ортонормированными.

Это можно сделать предварительно, до начала решения самой задачи. Для иллюстрации рассмотрим простейший пример. Пусть заданы два независимых вектора  $a$  и  $b$ , и мы хотим получить из них два взаимно перпендикулярных вектора  $v_1$  и  $v_2$ . Естественно, что в качестве первого из них можно взять вектор  $a$ , т. е.  $v_1 = a$ . Тогда задача состоит в отыскании перпендикулярного вектора. Но ведь именно с этого и начиналась данная глава: вектор  $b - p$  был перпендикулярен  $a$ , поскольку проекция вектора  $b$  на направление вектора  $a$  (вектор  $p$ ) вычиталась из  $b$ . Таким образом, направление второй оси задается вектором

$$v_2 = b - p = b - \frac{a^T b}{a^T a} a = b - \frac{v_1^T b}{v_1^T v_1} v_1. \quad (39)$$

Легко проверяется, что  $v_1^T v_2 = v_1^T b - v_1^T b = 0$ .

Чтобы сделать процесс Грама — Шмидта совершенно очевидным, предположим, что у нас имеется еще и третий линейно независимый вектор  $c$ . Идея остается в точности такой же. Мы *вычитаем компоненты в разложении вектора  $c$  по направлениям  $v_1$  и  $v_2$ , которые уже нами получены:*

$$v_3 = c - \frac{v_1^T c}{v_1^T v_1} v_1 - \frac{v_2^T c}{v_2^T v_2} v_2. \quad (40)$$

Тогда вектор  $v_3$  автоматически оказывается перпендикулярен векторам  $v_1$  и  $v_2$ , например:

$$v_1^T v_3 = v_1^T c - \frac{v_1^T c}{v_1^T v_1} v_1^T v_1 - \frac{v_2^T c}{v_2^T v_2} v_1^T v_2 \stackrel{0}{=} 0.$$

Фактически мы вычитаем из вектора  $c$  его проекцию на плоскость векторов  $a$  и  $b$ , но нам удобнее использовать уже имеющиеся взаимно перпендикулярные направления  $v_1$  и  $v_2$ . Проекция вектора  $c$  на плоскость равна сумме проекций этого вектора на направления  $v_1$  и  $v_2$ , и именно поэтому формула (40) оказывается столь простой. Отметим, что  $v_3$  не может быть нулевым вектором, поскольку в этом случае вектор  $c$  лежал бы в плоскости векторов  $a$  и  $b$ , что противоречит линейной независимости всех трех векторов. Разумеется, полученные векторы  $v_i$  являются лишь ортогональными, а не ортонормированными. Для превращения их в единичные векторы разделим каждый вектор на его длину:

$$q_1 = \frac{v_1}{\|v_1\|}, \quad q_2 = \frac{v_2}{\|v_2\|}, \quad q_3 = \frac{v_3}{\|v_3\|}.$$

Этот процесс ортогонализации, дающий каждый новый вектор  $v_i$  путем вычитания кратных векторов  $v_1, \dots, v_{i-1}$ , формулируется следующим образом:

**30.** Произвольный набор независимых векторов  $a_1, \dots, a_n$  может быть преобразован в набор взаимно ортогональных векторов при помощи процесса Грама—Шмидта: в качестве первого вектора берем  $v_1 = a_1$  и затем строим каждый последующий вектор  $v_i$ , ортогональный к векторам  $v_1, \dots, v_{i-1}$ , по формуле

$$v_i = a_i - \frac{v_1^T a_i}{v_1^T v_1} v_1 - \dots - \frac{v_{i-1}^T a_i}{v_{i-1}^T v_{i-1}} v_{i-1}. \quad (41)$$

Для любого номера  $i$  подпространство, порожденное исходными векторами  $a_1, \dots, a_i$ , совпадает с подпространством, порожденным векторами  $v_1, \dots, v_i$ . В результате векторы  $q_i = v_i / \|v_i\|$  оказываются ортонормированными.

**Пример.** Пусть заданы векторы

$$a_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

Тогда  $v_1 = a_1$ , а вектор  $v_2$  вычисляется согласно (39):

$$\frac{a_2^T v_1}{v_1^T v_1} = \frac{1}{2}, \quad v_2 = a_2 - \frac{1}{2} v_1 = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix}.$$

Третья перпендикулярная ось находится из уравнения (40):

$$\frac{a_3^T v_1}{v_1^T v_1} = \frac{1}{2}, \quad \frac{a_3^T v_2}{v_2^T v_2} = \frac{\frac{1}{2}}{\frac{6}{4}} = \frac{1}{3}, \quad v_3 = a_3 - \frac{1}{2} v_1 - \frac{1}{3} v_2 = \begin{bmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}.$$

Окончательная система ортонормированных векторов имеет вид

$$q_1 = \frac{v_1}{\|v_1\|} = \sqrt{\frac{1}{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad q_2 = \frac{v_2}{\|v_2\|} = \sqrt{\frac{2}{3}} \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix},$$

$$q_3 = \frac{v_3}{\|v_3\|} = \sqrt{\frac{3}{4}} \begin{bmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}.$$

Поскольку описанный алгоритм Грама—Шмидта является весьма простым и очевидным, должен существовать в равной степени простой и очевидный способ записи окончательного его результата. Поясним, как это можно сделать. Ситуация во многом близка к той, которая встречалась уже в методе исключения Гаусса, где элементарные операции выбирались естественным образом в процессе исключения, но выписывать заранее всю последовательность было бы очень неудобно. Для метода исключения удобная форма записи результата заключалась в разложении матрицы  $A = LU$ . В нашем же случае процесс Грама—Шмидта будет выражаться другим разложением матрицы  $A$ .

Так же как и для метода исключения, прием состоит в отыскании формул, выражающих исходные столбцы  $a_i$  матрицы через векторы  $q_i$ . В рассмотренном выше примере можно записать

$$a_1 = v_1, \quad \text{или} \quad a_1 = \sqrt{2} q_1,$$

$$a_2 = \frac{1}{2} v_1 + v_2, \quad \text{или} \quad a_2 = \sqrt{\frac{1}{2}} q_1 + \sqrt{\frac{3}{2}} q_2,$$

$$a_3 = \frac{1}{2} v_1 + \frac{1}{3} v_2 + v_3, \quad \text{или} \quad a_3 = \sqrt{\frac{1}{2}} q_1 + \sqrt{\frac{1}{6}} q_2 + \sqrt{\frac{4}{3}} q_3.$$

Эта система уравнений так и просится быть записанной в матричном виде:

$$\begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ 0 & \sqrt{\frac{3}{2}} & \sqrt{\frac{1}{6}} \\ 0 & 0 & \sqrt{\frac{4}{3}} \end{bmatrix}. \quad (42)$$

*Исходная матрица  $A$  разлагается в произведение ортогональной матрицы  $Q$  и верхней треугольной матрицы  $R$ .* Столбцы матрицы  $Q$  являются искомыми ортонормированными векторами, а элементы матрицы  $R$  получаются из соотношения (41). Оно дает выражение векторов  $a_i$  через комбинацию векторов  $v_1, \dots, v_i$ , и нам нужно лишь заменить  $v_1$  на  $\|v_1\| q_1$ ,  $v_2$  на  $\|v_2\| q_2$  и т. д. Тогда  $a_i$  будет комбинацией векторов  $q_1, \dots, q_i$ , а оставшиеся векторы  $q_{i+1}, \dots, q_n$  в нее не входят. Именно поэтому матрица  $R$  является верхней треугольной. Коэффициенты этой линейной комбинации записываются в  $i$ -м столбце матрицы  $R$ , причем элементы главной диагонали матрицы не равны нулю, поскольку вектор  $v_i$  заменяется в (41) на  $\|v_i\| q_i$ , и на диагонали будет находиться ненулевой коэффициент  $\|v_i\|^2$ . Таким образом, матрица  $R$  обратима. Это дает основной результат данного параграфа:

**ЗР.** Любая матрица  $A$  с линейно независимыми столбцами может быть разложена в произведение  $A = QR$ , где матрица  $Q$  имеет ортогональные столбцы, а  $R$  является верхней треугольной и обратимой матрицей. Если исходная матрица  $A$  была квадратной, то и матрицы  $Q$  и  $R$  будут квадратными, так что  $Q$  окажется ортогональной матрицей.

Используя разложение  $A = QR$ , легко решить задачу о наименьших квадратах  $Ax = b$ . Из (21) имеем

$$\bar{x} = (A^T A)^{-1} A^T b = (R^T Q^T Q R)^{-1} R^T Q^T b.$$

Но  $Q^T Q = I$ , поскольку столбцы матрицы  $Q$  являются ортонормированными. Следовательно,

$$\bar{x} = (R^T R)^{-1} R^T Q^T b = R^{-1} Q^T b. \quad (43)$$

Таким образом, вычисление решения  $\bar{x}$  требует лишь подсчета произведения  $Q^T b$  и обратной подстановки в системе с треуголь-

<sup>1)</sup> Нетрудно заметить, что в (42) на диагонали расположены длины векторов  $\|v_i\|$ .

ной матрицей  $R\bar{x} = Q^T b$ . Предварительно выполненная ортогонализация избавляет нас от необходимости строить матрицу  $A^T A$  и решать систему нормальных уравнений  $A^T A \bar{x} = A^T b$ <sup>1)</sup>.

**Упражнение 3.3.9.** Применить процесс Грама—Шмидта к векторам

$$a_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

и записать результат в виде произведения  $A = QR$ .

**Упражнение 3.3.10.** Представить матрицу  $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$  в виде  $QR$ .

**Упражнение 3.3.11.** Записать ортогонализацию Грама—Шмидта для векторов

$$a_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$$

в виде  $A = QR$ . Если заданы  $n$  векторов  $a_i$ , каждый из которых имеет  $m$  компонент, то как будут выглядеть матрицы  $A$ ,  $Q$  и  $R$ ?

**Упражнение 3.3.12.** Для той же самой матрицы  $A$  и вектора  $b = (1, 1, 1)^T$  решить задачу о наименьших квадратах  $Ax = b$  (используя разложение  $A = QR$ ).

**Упражнение 3.3.13.** Пусть  $A = QR$ . Найти простое выражение для матрицы  $P$ , осуществляющей проектирование на пространство столбцов матрицы  $A$ .

**Упражнение 3.3.14.** Показать, что два шага

$$w = c - \frac{v_1^T c}{v_1^T v_1} v_1, \quad v_3 = w - \frac{v_2^T w}{v_2^T v_2} v_2$$

дают то же самое третье направление  $v_3$ , что и формула (40). Это пример модифицированного алгоритма Грама—Шмидта, где для повышения устойчивости счета на каждом шаге вычитается лишь одна проекция.

## ФУНКЦИОНАЛЬНЫЕ ПРОСТРАНСТВА И РЯДЫ ФУРЬЕ

Хотя этот короткий раздел и является дополнительным, он преследует важные цели, а именно:

(1) ввести (причем впервые) понятие бесконечномерного пространства,

<sup>1)</sup> При этом мы выигрываем, но не в числе операций, которое на самом деле при использовании предварительной ортогонализации будет больше, а в численной устойчивости. По крайней мере это будет так, если перейти к модифицированному алгоритму Грама—Шмидта, описанному в упр. 3.3.14, или (лучше) к алгоритму Хаусхолдера из § 7.3.

(2) перенести понятия длины и скалярного произведения с векторов  $v$  на функции  $f(x)$ ,

(3) представить ряд Фурье функции  $f$  как сумму одномерных проекций; при этом ортогональными «столбцами», которые порождают пространство, будут синусы и косинусы,

(4) применить ортогонализацию Грама — Шмидта к функциям  $1, x, x^2, \dots$ ,

(5) найти наилучшую линейную аппроксимацию функции  $f(x)$ .

Попытаемся систематически рассмотреть эти вопросы, которые открывают новые возможности приложений линейной алгебры.

1. После изучения всех конечномерных пространств  $\mathbf{R}^n$  естественно рассмотреть пространство  $\mathbf{R}^\infty$ , содержащее все векторы вида  $v = (v_1, v_2, v_3, \dots)$ , имеющие бесконечное число компонент. Такое пространство окажется слишком «большим» для того, чтобы его можно было использовать, если на компоненты  $v_j$  не наложить никаких ограничений. Лучше всего сохранить уже привычное определение длины как корня из суммы квадратов компонент и ограничиться лишь теми векторами, которые имеют конечную длину: бесконечный ряд

$$\|v\|^2 = v_1^2 + v_2^2 + v_3^2 + \dots \quad (44)$$

должен сходиться, т. е. иметь конечную сумму. При этом мы сохраняем бесконечномерность векторов и включаем, например, вектор  $(1, 1/2, 1/3, \dots)$ , но исключаем  $(1, 1/\sqrt{2}, 1/\sqrt{3}, \dots)$ . Векторы конечной длины можно складывать друг с другом (поскольку  $\|v+w\| \leq \|v\| + \|w\|$ ), и умножать на скаляры, так что они образуют векторное пространство. Это знаменитое *гильбертово пространство*.

Понятие гильбертова пространства дает наиболее естественный способ введения бесконечного числа измерений при сохранении геометрии обычных евклидовых пространств. Эллипсы превращаются в бесконечномерные эллипсоиды, параболы переходят в параболоиды, и перпендикулярность прямых вводится совершенно аналогичным образом: векторы  $v = (v_1, v_2, \dots)$  и  $w = (w_1, w_2, \dots)$  являются ортогональными, если их скалярное произведение равняется нулю,

$$v^T w = v_1 w_1 + v_2 w_2 + v_3 w_3 + \dots = 0. \quad (45)$$

Эта сумма, естественно, конечна, и для любых двух векторов она удовлетворяет неравенству Шварца  $|v^T w| \leq \|v\| \|w\|$ . Значения косинуса даже в гильбертовом пространстве никогда не превосходят единицу.

Это пространство замечательно еще в одном отношении: оно является очень распространенным. Векторы гильбертова прост-

ранства могут превратиться в функции, и это приводит нас ко второму пункту.

2. Рассмотрим функцию  $f(x) = \sin x$  на интервале  $0 \leq x \leq 2\pi$ . Ее можно воспринимать как вектор с континуальным числом компонент, равных значениям синуса на всем отрезке. Естественно, что применить обычное правило суммирования квадратов для отыскания длины такого вектора невозможно. Поэтому суммирование естественным образом заменяется *интегрированием*:

$$\|f\|^2 = \int_0^{2\pi} (f(x))^2 dx = \int_0^{2\pi} (\sin x)^2 dx = \pi. \quad (46)$$

Наше гильбертово пространство превратилось в некоторое функциональное пространство: векторами являются функции, мы можем измерять их длину, и пространство содержит все функции, имеющие конечную длину, как в формуле (44). Это пространство не содержит функции  $F(x) = 1/x$ , поскольку интеграл от  $1/x^2$  расходится.

Та же самая идея замены суммирования интегрированием дает *скалярное произведение двух функций*: если  $f(x) = \sin x$  и  $g(x) = \cos x$ , то

$$f^T g = \int_0^{2\pi} f(x) g(x) dx = \int_0^{2\pi} \sin x \cos x dx = 0.$$

Скалярное произведение по-прежнему связано с длиной соотношением  $f^T f = \|f\|^2$ , и выполняется неравенство Шварца:  $|f^T g| \leq \|f\| \|g\|$ . Естественно, что две функции, скалярное произведение которых равняется нулю (как  $\sin x$  и  $\cos x$ ), будут называться ортогональными. Если разделить их на длину  $\sqrt{\pi}$ , то они будут даже ортонормированными.

3. *Рядом Фурье* функции  $y(x)$  называется разложение этой функции по синусам и косинусам:

$$y(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \dots$$

Для вычисления любого коэффициента, скажем  $b_1$ , умножим обе части на соответствующую функцию (здесь на  $\sin x$ ) и проинтегрируем от 0 до  $2\pi$ . Другими словами, мы вычисляем скалярное произведение обеих частей на  $\sin x$ :

$$\int_0^{2\pi} y(x) \sin x dx = a_0 \int_0^{2\pi} \sin x dx + a_1 \int_0^{2\pi} \cos x \sin x dx + b_1 \int_0^{2\pi} (\sin x)^2 dx + \dots$$

Все интегралы в правой части, кроме одного, будут равны нулю. Единственный ненулевой интеграл получается при умножении  $\sin x$  на себя. Это означает, что *синусы и косинусы попарно ортогональны*. Таким образом, коэффициент при функции  $f(x) = \sin x$  равен

$$b_1 = \frac{\int_0^{2\pi} y(x) \sin x \, dx}{\int_0^{2\pi} (\sin x)^2 \, dx} = \frac{y^T f}{f^T f}, \quad \text{или} \quad b_1 \sin x = \frac{y^T f}{f^T f} f. \quad (47)$$

Эти вычисления производятся для того, чтобы увидеть аналогию с проекциями. Компонента в разложении вектора  $b$ , лежащая на прямой, порожденной вектором  $a$ , была вычислена в начале главы:

$$\bar{x} = \frac{b^T a}{a^T a}, \quad \text{или} \quad p = \bar{x} a = \frac{b^T a}{a^T a} a.$$

Для ряда Фурье мы проектируем функцию  $y$  на функцию  $\sin x$ , и компонента  $p$  по этому направлению в точности совпадает с членом  $b_1 \sin x$ . Коэффициент  $b_1$  является решением по методу наименьших квадратов для несовместного уравнения  $b_1 \sin x = y$ . Другими словами, мы выбираем функцию  $b_1 \sin x$  как можно ближе к функции  $y$  (см. упр. 3.3.16). То же самое верно и для остальных членов ряда: каждый из них является проекцией функции  $y$  на синус или косинус. Поскольку эти синусы и косинусы являются взаимно ортогональными, *ряд Фурье дает координаты «вектора»  $y$  по набору (бесконечного числа) взаимно перпендикулярных осей.*

4. Существует масса других полезных функций, отличных от синуса и косинуса, и они не всегда являются ортогональными. Простейшими из них являются многочлены, и, к сожалению, не существует интервала, на котором даже первые три координатные оси, т. е. функции  $1, x, x^2$ , были бы взаимно перпендикулярными. (Скалярное произведение  $1$  на  $x^2$  всегда положительно, поскольку оно равняется интегралу от  $x^2$ .) Поэтому ближайшая к функции  $y(x)$  парабола не будет суммой проекций на  $1, x$  и  $x^2$ . Здесь будет присутствовать сопрягающий член, аналогичный  $(A^T A)^{-1}$  для матричного случая, и фактически это сопряжение осуществляется при помощи плохо обусловленной матрицы Гильберта,

рассмотренной в упр. 1.5.11. На отрезке  $0 \leq x \leq 1$  имеем

$$A^T A = \begin{bmatrix} 1^T 1 & 1^T x & 1^T x^2 \\ x^T 1 & x^T x & x^T x^2 \\ (x^2)^T 1 & (x^2)^T x & (x^2)^T x^2 \end{bmatrix} = \begin{bmatrix} \int 1 & \int x & \int x^2 \\ \int x & \int x^2 & \int x^3 \\ \int x^2 & \int x^3 & \int x^4 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}.$$

Элементы обратной матрицы будут велики, поскольку оси  $1$ ,  $x$ ,  $x^2$  далеко не перпендикулярны. Даже для современной ЭВМ ситуация становится безнадежной, если мы добавим еще несколько осей; *решить систему нормальных уравнений  $A^T A \bar{x} = A^T b$  для отыскания многочлена наилучшего приближения десятой степени практически невозможно.*

Точнее говоря, невозможно решить эту систему методом исключения Гаусса, поскольку все ошибки округления будут возрастать более чем в  $10^{13}$  раз. С другой стороны, мы не можем отказаться от возможности осуществлять аппроксимацию при помощи многочленов. Правильный подход состоит в переходе к взаимно ортогональным осям при помощи ортогонализации Грама—Шмидта: мы ищем комбинации функций  $1$ ,  $x$ ,  $x^2$ , являющиеся ортогональными.

Нам удобнее перейти к симметрично расположенному отрезку, например  $-1 \leq x \leq 1$ , поскольку при этом нечетные степени  $x$  становятся ортогональными к четным:

$$1^T x = \int_{-1}^1 x dx = 0, \quad x^T x^2 = \int_{-1}^1 x^3 dx = 0.$$

Следовательно, можно начать процесс Грама—Шмидта, положив  $v_1 = 1$ ,  $v_2 = x$  в качестве первых двух перпендикулярных осей, так что остается лишь скорректировать угол между  $1$  и  $x^2$ . Пользуясь формулой (40), получаем третий ортогональный многочлен в виде

$$v_3 = x^2 - \frac{1^T x^2}{1^T 1} 1 - \frac{x^T x^2}{x^T x} x = x^2 - \frac{\int_{-1}^1 x^2 dx}{\int_{-1}^1 1 dx} = x^2 - \frac{1}{3}.$$

Построенные таким способом многочлены называются *многочленами Лежандра*, и они взаимно ортогональны на отрезке  $-1 \leq x \leq 1$ .

*Проверка.*

$$1^T \left( x^2 - \frac{1}{3} \right) = \int_{-1}^1 \left( x^2 - \frac{1}{3} \right) dx = \left[ \frac{x^3}{3} - \frac{x}{3} \right]_{-1}^1 = 0.$$

Теперь нетрудно вычислить многочлен наилучшего приближения десятой степени, осуществляя проектирование на каждый из первых десяти (или одиннадцати) многочленов Лежандра.

5. Пусть на отрезке  $0 \leq x \leq 1$  задана функция  $f(x) = \sqrt{x}$ , и мы хотим приблизить ее прямой линией  $y = C + Dx$ . Существуют по крайней мере три способа отыскания этой наилучшей прямой, и, если сравнить их друг с другом, это может прояснить содержание всей главы!

(i) Минимизировать

$$E^2 = \int_0^1 (\sqrt{x} - C - Dx)^2 dx = \frac{1}{2} - \frac{4}{3}C - \frac{4}{5}D + C^2 + CD + \frac{D^2}{3}.$$

Приравняв нулю производные, получаем

$$\frac{\partial E^2}{\partial C} = -\frac{4}{3} + 2C + D = 0, \quad \frac{\partial E^2}{\partial D} = -\frac{4}{5} + C + \frac{2D}{3} = 0.$$

Отсюда получаем решение  $C = 4/15$ ,  $D = 4/5$ .

(ii) Решить систему  $\begin{bmatrix} 1 & x \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \sqrt{x} \end{bmatrix}$ , или  $Ay = f$ , методом наименьших квадратов.

$$A^T A = \begin{bmatrix} 1^T 1 & 1^T x \\ x^T 1 & x^T x \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}, \quad (A^T A)^{-1} = \begin{bmatrix} 4 & -6 \\ -6 & 12 \end{bmatrix},$$

$$A^T f = \begin{bmatrix} 1^T \sqrt{x} \\ x^T \sqrt{x} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{2}{5} \end{bmatrix}.$$

Решение нормальной системы имеет вид

$$\bar{y} = (A^T A)^{-1} A^T f, \quad \text{или} \quad \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 4 & -6 \\ -6 & 12 \end{bmatrix} \begin{bmatrix} \frac{2}{3} \\ \frac{2}{5} \end{bmatrix} = \begin{bmatrix} \frac{4}{15} \\ \frac{4}{5} \end{bmatrix}.$$

(iii) Применить процесс Грама — Шмидта для получения второго столбца  $x - 1/2$ , ортогонального первому столбцу:

$$1^T \left( x - \frac{1}{2} \right) = \int \left( x - \frac{1}{2} \right) dx = 0.$$

Тогда проекция функции  $V\bar{x}$  равна

$$y = \frac{V\bar{x}^T 1}{1^T 1} 1 + \frac{V\bar{x}^T \left(x - \frac{1}{2}\right)}{\left(x - \frac{1}{2}\right)^T \left(x - \frac{1}{2}\right)} \left(x - \frac{1}{2}\right) = \\ = \frac{2}{3} + \frac{4}{5} \left(x - \frac{1}{2}\right) = \frac{4}{15} + \frac{4}{5} x.$$

**Упражнение 3.3.15.** Найти длину вектора  $v = (1/\sqrt{2}, 1/\sqrt{4}, 1/\sqrt{8}, \dots)$  и функции  $f(x) = e^x$  (на отрезке  $0 \leq x \leq 1$ ). Чему равняется скалярное произведение функций  $e^x$  и  $e^{-x}$  на этом отрезке?

**Упражнение 3.3.16.** Приравнявая нулю производную, найти значение  $b_1$ , минимизирующее

$$\|b_1 \sin x - y\|^2 = \int_0^{2\pi} (b_1 \sin x - y(x))^2 dx.$$

Сравнить с коэффициентом Фурье (47). Если  $y(x) = \cos x$ , то чему будет равен  $b_1$ ?

**Упражнение 3.3.17.** Найти коэффициенты Фурье  $a_0, a_1, b_1$  для ступенчатой функции  $y(x)$ , равной 1 на отрезке  $0 \leq x \leq \pi$  и 0 на интервале  $\pi < x < 2\pi$ :

$$a_0 = \frac{y^T 1}{1^T 1}, \quad a_1 = \frac{y^T \cos x}{(\cos x)^T \cos x}, \quad b_1 = \frac{y^T \sin x}{(\sin x)^T \sin x}.$$

**Упражнение 3.3.18.** Найти следующий многочлен Лежандра, т. е. кубический многочлен, ортогональный к 1,  $x$  и  $x^2 - 1/3$ .

**Упражнение 3.3.19.** Как выглядит прямая наилучшего приближения для параболы  $y = x^2$  на отрезке  $-1 \leq x \leq 1$ ?

### § 3.4. ПСЕВДООБРАЩЕНИЕ И СИНГУЛЯРНОЕ РАЗЛОЖЕНИЕ

Чему равно оптимальное решение  $\bar{x}$  несовместной системы  $Ax = b$ ? Мы еще не получили окончательного ответа на этот основной вопрос всей главы. Теперь мы собираемся сделать это. Наша цель состоит в указании правила, с помощью которого мы находим вектор  $\bar{x}$  по заданной матрице коэффициентов  $A$  и правой части  $b$ .

Эта цель уже наполовину достигнута, поскольку мы знаем, чему равен вектор  $A\bar{x}$ . Для произвольного  $x$  вектор  $Ax$  обязательно лежит в пространстве столбцов матрицы  $A$ ; он является комбинацией столбцов с компонентами вектора  $x$  в качестве весов. Поэтому оптимальный выбор  $A\bar{x}$  представляет собой точку  $p$  в пространстве столбцов, ближайшую к заданной точке  $b$ . Этот выбор минимизирует ошибку  $E = \|Ax - b\|$ . Другими словами, наи-

лучшая стратегия заключается в проектировании вектора  $b$  на пространство столбцов:

$$A\bar{x} = p = Pb. \quad (48)$$

В большом числе случаев (я думаю, что они включают большинство практических задач) этого уравнения достаточно для отыскания вектора  $\bar{x}$ . (Оно является другой формой записи нормального уравнения  $A^T A \bar{x} = A^T b$ , которое возникало в §1 при минимизации ошибки  $E$ .) Естественно, что вектор  $\bar{x}$  будет определен лишь в том случае, когда существует только одна комбинация столбцов матрицы  $A$ , дающая вектор  $p$ ; веса этой комбинации и определяют вектор  $\bar{x}$ . Это соответствует «единственности», описанной в гл. 2, и мы знаем несколько эквивалентных условий для того, чтобы уравнение  $A\bar{x} = p$  имело единственное решение:

- (i) столбцы матрицы  $A$  линейно независимы,
- (ii) нуль-пространство матрицы  $A$  содержит только нулевой вектор,
- (iii) ранг матрицы  $A$  равен  $n$ ,
- (iv) квадратная матрица  $A^T A$  является обратимой.

Для этого случая единственное решение уравнения (48) уже изучалось нами в течение всей главы:

$$\bar{x} = (A^T A)^{-1} A^T b. \quad (49)$$

Эта сравнительно простая формула включает в себя наиболее простой случай — когда матрица  $A$  действительно обратима. Тогда вектор  $\bar{x}$  совпадает с единственным решением исходной системы  $Ax = b$ :  $\bar{x} = A^{-1} (A^T)^{-1} A^T b = A^{-1} b$ .

Это обстоятельство дает другой способ для описания нашей цели: мы пытаемся определить *псевдообратную*  $A^+$  к матрице, которая может не быть обратимой<sup>1)</sup>. Если матрица обратима, то, естественно,  $A^+ = A^{-1}$ . Когда исходная матрица удовлетворяет перечисленным выше условиям (i) — (iv), псевдообратной является левая обратная матрица из формулы (49):  $A^+ = (A^T A)^{-1} A^T$ . Но если условия (i) — (iv) не выполняются и вектор  $\bar{x}$  определяется из системы  $Ax = p$  не единственным образом, понятие псевдообратной матрицы еще подлежит определению. Мы должны выбрать один из многих векторов, удовлетворяющих системе  $A\bar{x} = p$ , и

<sup>1)</sup> Матрица  $A^+$  также называется *обратной матрицей Мура—Пенроуза*, поскольку именно они ввели это понятие, или *обобщенной обратной к  $A$* . Однако название «обобщенная обратная» часто применяется к другим матрицам, которые обладают не всеми свойствами матрицы  $A^+$ . Именно поэтому мы и пользуемся термином «псевдообратная».

этот выбор по определению будет считаться оптимальным решением  $\bar{x} = A^+b$  несовместной системы  $Ax = b$ .

Выбор осуществляется по следующему правилу: *оптимальным среди всех решений системы  $A\bar{x} = p$  является решение с минимальной длиной*. Для его отыскания вспомним, что пространство строк и нуль-пространство матрицы  $A$  представляют собой ортогональные дополнения друг к другу в пространстве  $\mathbb{R}^n$ . Это означает, что любой вектор может быть разложен на две перпендикулярные составляющие — проекцию на пространство строк и проекцию на нуль-пространство. Применим это разложение к одному из решений (обозначим его через  $\bar{x}_0$ ) системы  $A\bar{x} = p$ . Тогда  $\bar{x}_0 = \bar{x}_r + \omega$ , где  $\bar{x}_r$  принадлежит пространству строк, а  $\omega$  — нуль-пространству. Здесь возникают три существенных момента:

(1) Компонента  $\bar{x}_r$  сама является решением системы  $A\bar{x} = p$ : поскольку  $A\omega = 0$ , имеем

$$A\bar{x}_r = A(\bar{x}_r + \omega) = A\bar{x}_0 = p.$$

(2) Все решения системы  $A\bar{x} = p$  имеют одну и ту же компоненту  $\bar{x}_r$  в пространстве строк и отличаются лишь компонентой из нуль-пространства. «Общее решение равняется сумме частного решения (в данном случае это  $\bar{x}_r$ ) и произвольного решения однородного уравнения».

(3) Длина такого решения  $\bar{x}_r + \omega$  удовлетворяет теореме Пифагора, поскольку две компоненты взаимно ортогональны:

$$\|\bar{x}_r + \omega\|^2 = \|\bar{x}_r\|^2 + \|\omega\|^2.$$

Отсюда мы заключаем, что *решением с минимальной длиной является вектор  $\bar{x}_r$* . Нам следует приравнять компоненту из нуль-пространства нулю, оставив решение, целиком принадлежащее пространству строк.

**3Q.** *Оптимальным решением по методу наименьших квадратов произвольной системы  $Ax = b$  является вектор  $\bar{x}_r$  (или  $\bar{x}$ , если вернуться к исходным обозначениям), определяемый двумя условиями:*

(1) *вектор  $A\bar{x}$  равняется проекции вектора  $b$  на пространство столбцов матрицы  $A$ ,*

(2) *вектор  $\bar{x}$  принадлежит пространству строк матрицы  $A$ . Матрица, «решающая» систему  $Ax = b$ , является псевдообратной к  $A$  матрицей  $A^+$ , и она определяется уравнением  $\bar{x} = A^+b$ .*

Легче всего понять смысл псевдообратной матрицы из геометрических соображений. Для этого мы вновь обратимся к четырем основным подпространствам матрицы (рис. 3.7).

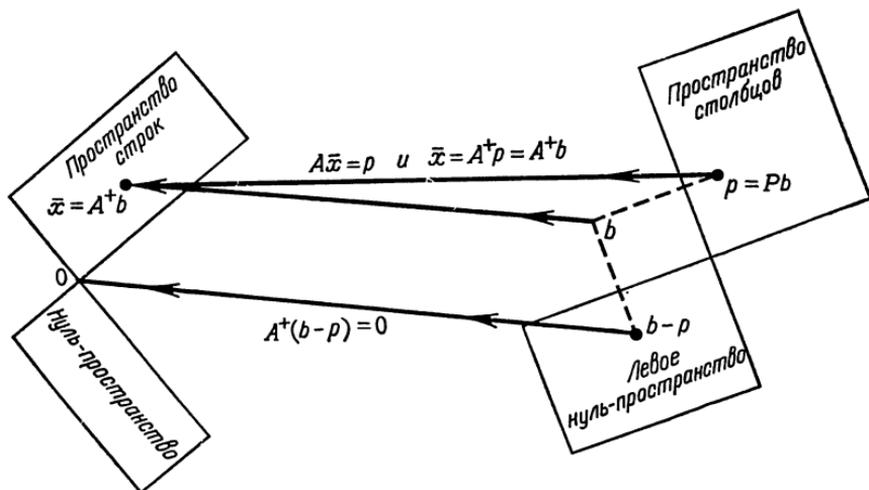


Рис. 3.7. Матрица  $A$  и ее псевдообратная  $A^+$ .

Матрица  $A^+$  сочетает в себе результат применения двух отдельных шагов: она проектирует  $b$  на  $p$  и выбирает единственный вектор  $\bar{x}$  из пространства строк матрицы, являющийся решением системы  $A\bar{x} = p$ . Один крайний случай возникает, когда вектор  $b$  перпендикулярен пространству столбцов матрицы, т. е. когда вектор  $b$  лежит в левом нуль-пространстве. Тогда  $p = 0$  и  $\bar{x} = 0$ : матрица  $A^+$  переводит все векторы из  $\mathcal{N}(A^T)$  в нуль. Другая крайность состоит в том, что вектор  $b$  может лежать целиком в пространстве столбцов; тогда  $p = b$  и решение  $\bar{x}$  находится путем «обращения» матрицы  $A$ . (Мы уже отмечали в  $\mathbb{Z}$ , что матрица  $A$  обратима, если рассматривать ее просто как отображение из ее пространства строк в пространство столбцов; обратной будет матрица  $A^+$ .)

Произвольный вектор  $b$  разбивается в соответствии с этими двумя предельными случаями: компонента  $p$  обращается, давая вектор  $\bar{x}$ , а другая компонента  $b-p$  аннулируется. Из этого описания и из рис. 3.7 можно вывести некоторые основные свойства псевдообратной матрицы:

- (1) Матрица  $A^+$  имеет размер  $n \times m$ . Она применяется к вектору  $b$  из  $\mathbb{R}^m$  и дает вектор  $\bar{x}$  из  $\mathbb{R}^n$ .
- (2) Пространство столбцов матрицы  $A^+$  совпадает с пространством строк матрицы  $A$ , а пространство строк матрицы  $A^+$  сов-

падает с пространством столбцов  $A$ . (В качестве важного следствия получаем, что  $r(A^+) = r(A)$ .)

(3) Псевдообратная к матрице  $A^+$  есть матрица  $A$ .

(4) В общем случае  $AA^+ \neq I$ , поскольку матрица  $A$  может не иметь правой обратной, но матрица  $AA^+$  всегда совпадает с матрицей  $P$ , осуществляющей проектирование на пространство столбцов:

$$AA^+b = A\bar{x} = p = Pb, \quad \text{или} \quad AA^+ = P. \quad (50)$$

**Пример 1.** Матрица

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

не является обратимой. Ее пространство столбцов и пространство строк совпадают с плоскостью  $xy$  в  $\mathbb{R}^3$ , которая содержит все векторы вида  $(x, y, 0)$ . Нуль-пространством является ось  $z$ , и, разумеется, оно ортогонально пространству строк. Отыскание вектора  $\bar{x}$  начинается с проектирования  $b$  на пространство столбцов:

$$\text{если} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \quad \text{то} \quad p = Pb = \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix}.$$

Затем мы решаем систему  $A\bar{x} = p$ , т. е.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix}. \quad (51)$$

Решение выписывается легко:  $\bar{x}_1 = b_1$ ,  $\bar{x}_2 = b_2$  и  $\bar{x}_3$  произвольно. Из этого бесконечного семейства решений мы выбираем одну, имеющее минимальную длину, полагая третью компоненту  $\bar{x}_3$  равной нулю. Это дает вектор  $\bar{x} = (b_1, b_2, 0)^T$ , лежащий в пространстве строк (т. е. в плоскости  $xy$ ), как и должно было быть, согласно теории. Наконец, мы находим псевдообратную матрицу  $A^+$ :

$$A^+ = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A^+b = \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix} = \bar{x}.$$

В нашем случае матрица  $A^+$  совпадает с  $A$ . Это происходит потому, что матрица  $A$  действует подобно единичной матрице

при отображении своего пространства строк в пространство столбцов, а все остальное для псевдообратной матрицы несущественно.

Подведем итог сказанному выше. Оптимальное решение несовместной системы  $Ax = b$

$$\begin{aligned}x_1 + 0x_2 + 0x_3 &= b_1, \\0x_1 + x_2 + 0x_3 &= b_2, \\0x_1 + 0x_2 + 0x_3 &= b_3\end{aligned}$$

заключается в том, что мы находим решение первых двух уравнений и полагаем  $x_3 = 0$ .

**Пример 2.** Пусть

$$A = \begin{bmatrix} \mu_1 & 0 & 0 & 0 \\ 0 & \mu_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

где  $\mu_1 > 0$ ,  $\mu_2 > 0$ . Пространство столбцов этой матрицы вновь совпадает с плоскостью  $xy$ , так что проектирование вектора  $b$  опять приводит к аннулированию его компоненты вдоль оси  $z$ :  $p = Pb = (b_1, b_2, 0)^T$ . Система  $A\bar{x} = p$  принимает вид

$$\begin{bmatrix} \mu_1 & 0 & 0 & 0 \\ 0 & \mu_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \bar{x}_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix}.$$

Первые две компоненты вновь оказываются определенными:  $\bar{x}_1 = b_1/\mu_1$ ,  $\bar{x}_2 = b_2/\mu_2$ . Оставшиеся две компоненты должны равняться нулю в силу требования минимальности длины, так что оптимальное решение имеет вид

$$\bar{x} = \begin{bmatrix} b_1/\mu_1 \\ b_2/\mu_2 \\ 0 \\ 0 \end{bmatrix}, \quad \text{или} \quad A^+b = \begin{bmatrix} \mu_1^{-1} & 0 & 0 \\ 0 & \mu_2^{-1} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

Эти два примера являются типичными для всего семейства матриц специального вида, имеющих в качестве первых  $r$  элементов главной диагонали положительные числа  $\mu_1, \dots, \mu_r$  и остальные элементы которых равняются нулю. Если обозначить такую матрицу через  $\Sigma$ , то псевдообратная к ней вычисляется

в точности так же, как и в приведенных примерах:

$$\Sigma = \begin{bmatrix} \mu_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \mu_r \end{bmatrix} \quad \text{и} \quad \Sigma^+ = \begin{bmatrix} \mu_1^{-1} & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \mu_r^{-1} \end{bmatrix}.$$

Если матрица  $\Sigma$  имеет размер  $m \times n$ , то  $\Sigma^+$  будет иметь размер  $n \times m$ . Легко видеть, что ранг матрицы  $\Sigma^+$  равен рангу  $\Sigma$  (т. е.  $r$ ) и что псевдообратная к  $\Sigma^+$  совпадает с  $\Sigma$ :  $(\Sigma^+)^+ = \Sigma$ .

Эти простые матрицы гораздо полезнее, чем может показаться, поскольку они дают новый способ для разложения матрицы  $A$ . Такое разложение называется *сингулярным разложением*, и оно отнюдь не столь известно, как того заслуживает.

**3R.** Произвольная матрица размера  $m \times n$  может быть представлена в виде  $A = Q_1 \Sigma Q_2^T$ , где  $Q_1$  есть ортогональная матрица размера  $m \times m$ ,  $Q_2$  — ортогональная матрица размера  $n \times n$ , а матрица  $\Sigma$  имеет специальную диагональную форму, описанную выше.

Числа  $\mu_i$  называются *сингулярными числами* матрицы  $A$ . Это разложение немедленно приводит к явной формуле для псевдообратной матрицы — результат, которого нам так не хватало. Для этого частного случая (но не для общего — см. формулу (60) ниже) мы собираемся «псевдообратить» все три сомножителя по отдельности и затем перемножить их, как обычно, в обратном порядке. Поскольку обратная к ортогональной матрице совпадает с транспонированной, формула будет иметь вид

$$A^+ = Q_2 \Sigma^+ Q_1^T. \quad (52)$$

Эта формула легко доказывается непосредственно из принципа наименьших квадратов. Умножение на ортогональную матрицу  $Q_1^T$  не изменяет длину вектора (согласно 3N), так что минимизируемая ошибка имеет вид

$$\|Ax - b\| = \|Q_1 \Sigma Q_2^T x - b\| = \|\Sigma Q_2^T x - Q_1^T b\|. \quad (53)$$

Введем новый неизвестный вектор  $y = Q_2^T x = Q_2^{-1} x$ , длина которого совпадает с длиной вектора  $x$ . Тогда мы будем минимизировать величину  $\|\Sigma y - Q_1^T b\|$ , и оптимальное решение  $\bar{y}$  (т. е. решение с минимальной длиной) имеет вид  $\bar{y} = \Sigma^+ Q_1^T b$ . Следова-

тельно,

$$\bar{x} = Q_2 \bar{y} = Q_2 \Sigma + Q_1^T b, \quad \text{или} \quad A^+ = Q_2 \Sigma + Q_1^T 1. \quad (54)$$

Доказательство справедливости приведенного сингулярного разложения основано на главном результате из гл. 5: для симметрической матрицы вида  $A^T A$  существует ортонормированный набор «собственных векторов»  $x_1, \dots, x_n$  (которые образуют столбцы матрицы  $Q_2$ ):

$$A^T A x_i = \lambda_i x_i, \quad \text{причем} \quad x_i^T x_i = 1 \quad \text{и} \quad x_j^T x_i = 0 \quad \text{для} \quad j \neq i. \quad (55a)$$

Умножая скалярно на  $x_i$ , получаем, что  $\lambda_i \geq 0$ :

$$x_i^T A^T A x_i = \lambda_i x_i^T x_i, \quad \text{или} \quad \|A x_i\|^2 = \lambda_i. \quad (55b)$$

Предположим, что числа  $\lambda_1, \dots, \lambda_r$  положительны, а остальные  $n - r$  векторов  $A x_i$  и чисел  $\lambda_i$  равны нулю. Для положительных  $\lambda_i$  положим  $\mu_i = \sqrt{\lambda_i}$  и  $y_i = A x_i / \mu_i$ . Из (55b) следует, что  $y_i$  являются единичными векторами, причем ортогональными (что вытекает из (55a)):

$$y_j^T y_i = \frac{x_j^T A^T A x_i}{\mu_j \mu_i} = \frac{\lambda_i x_j^T x_i}{\mu_j \mu_i} = 0 \quad \text{для} \quad j \neq i.$$

Следовательно, мы имеем систему  $r$  ортонормированных векторов, которая может быть расширена до полного ортонормированного базиса  $y_1, \dots, y_r, \dots, y_m$ , образующего столбцы матрицы  $Q_1$ , при помощи процедуры Грама—Шмидта. Тогда элементами матрицы  $Q_1^T A Q_2$  будут числа  $y_j^T A x_i$ , равные нулю при  $i > r$  (поскольку в этом случае  $A x_i = 0$ ) и равные  $y_j^T \mu_i y_i$  при  $i \leq r$ . Последние равняются нулю при  $j \neq i$  и  $\mu_i$  при  $j = i$ . Другими словами, матрица  $Q_1^T A Q_2$  в точности совпадает со специальной матрицей  $\Sigma$ , у которой на главной диагонали расположены числа  $\mu_i$ . Следовательно,  $A = Q_1 \Sigma Q_2^T$ . Матрицы вращения  $Q_1$  и  $Q_2$  поворачивают пространство столбцов матрицы  $A$  до совпадения с пространством строк, так что матрица  $A$  переходит в диагональную матрицу  $\Sigma$ .

**Пример 3.** Если  $A$  есть вектор-столбец  $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ , то  $A^T A$  будет матрицей первого порядка [25], и ее единственный собственный вектор равен  $x = [1]$ . Это и есть матрица  $Q_2$ . Затем мы находим  $\lambda = 25$ ,  $\mu = 5$  и  $y_1 = A x / \mu = \begin{bmatrix} 3/5 \\ 4/5 \end{bmatrix}$ . Второй столбец матрицы  $Q_1$

<sup>1</sup>) На практике бывает полезно выделить очень маленькие числа  $\mu_i$  и заменить их нулями. В противном случае ошибки округления или эксперимента легко могут дать, например, число  $\mu_r = 10^{-6}$ , так что  $\mu_r^{-1} = 10^6$  и псевдообратная матрица окажется «плохой». Это свидетельствует о крайней неустойчивости ранга и, следовательно, псевдообратной матрицы.

должен быть ортогонален к  $y_1$ , так что

$$Q_1^T A Q_2 = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} [1] = \begin{bmatrix} 5 \\ 0 \end{bmatrix} = \Sigma.$$

Это и есть требуемая диагональная форма. Псевдообратной к столбцу  $\Sigma$  является строка  $\Sigma^+ = [1/5 \ 0]$ , так что окончательно имеем  $A^+ = Q_2 \Sigma^+ Q_1^T = [3/25 \ 4/25]$ . Отметим, что  $A^+ A = I$ ; в данном примере ранг равен  $r = n = 1$ , так что псевдообратная является левой обратной матрицей.

Разумеется, сингулярное разложение требует отыскания ортогональных матриц  $Q_1$  и  $Q_2$ , что, как правило, невозможно осуществить при помощи операций над строками. Ближе по духу к нашему изложению было бы попытаться вычислить вектор  $x$  непосредственно из сомножителей  $L$  и  $U$  метода Гаусса. Результат напоминает  $U^{-1}L^{-1}$ -разложение матрицы  $A^{-1}$ , и, поскольку найти его где-либо невозможно, мы приведем полное доказательство, которое, разумеется, при чтении можно опустить<sup>1)</sup>.

Идея состоит в отбрасывании нулевых строк матрицы  $U$ , как это уже делалось на стр. 120. Если матрица  $A$  имеет ранг  $r$ , то таких строк будет  $m - r$ . Одновременно мы отбрасываем также и последние  $m - r$  столбцов матрицы  $L$ , или матрицы  $P^{-1}L$ , если в процессе исключения придется делать перестановки строк матрицы. Обозначим через  $\bar{U}$  оставшуюся от  $U$  матрицу размера  $r \times n$ , а через  $\bar{L}$  — оставшуюся от  $L$  матрицу размера  $m \times r$ .

**Пример 4** (когда ранг  $r$  равен 2).

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 3 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \bar{L}\bar{U}.$$

Последний столбец матрицы  $L$  — который умножается лишь на последнюю строку матрицы  $U$  — устраняется без изменения скалярного произведения:  $LU = \bar{L}\bar{U}$ . Отметим два существенных обстоятельства:

(1)  $\bar{U}$  есть матрица размера  $r \times n$  и ранга  $r$ . Ее пространство строк совпадает с пространством строк матрицы  $U$ , поскольку мы удаляем лишь нулевые строки. Следовательно, пространства строк матриц  $\bar{U}$  и  $A$  совпадают.

<sup>1)</sup> Мы не считаем, что новая и более «элементарная» формула (56) будет удобнее для вычислений, чем (52) или процесс Грама — Шмидта. Скорее всего, как раз наоборот.

(2)  $\bar{L}$  есть матрица размера  $m \times r$ , и она также имеет ранг  $r$ ; матрица  $L$  (и матрица  $P$  так же) обратима, так что ее первые  $r$  столбцов являются линейно независимыми. Как отмечалось в § 2.6, *пространства столбцов матриц  $\bar{L}$  и  $A$  совпадают.*

При помощи разложения  $A = \bar{L}\bar{U}$  мы можем, наконец, явным образом построить матрицу  $A^+$ :

**3S.** Псевдообратная к произвольной матрице размера  $m \times n$  дается формулой

$$A^+ = \bar{U}^T (\bar{U} \bar{U}^T)^{-1} (\bar{L}^T \bar{L})^{-1} \bar{L}^T. \quad (56)$$

Наша цель — проверить, что вектор  $A^+b$  является оптимальным решением  $\bar{x}$  задачи о наименьших квадратах  $Ax = b$ . Это и будет служить доказательством правильности формулы (56). Как отмечалось выше, вектор  $\bar{x}$  должен удовлетворять двум условиям: он должен лежать в пространстве строк матрицы  $A$ , и вектор  $A\bar{x}$  должен равняться проекции  $p$  вектора  $b$  на пространство столбцов. Поэтому доказательство заключается в сопоставлении этих свойств с двумя свойствами (для матриц  $\bar{U}$  и  $\bar{L}$ ), приведенными выше.

(1') Если записать вектор  $A^+b$  в соответствии с (56), то он будет произведением матрицы  $\bar{U}^T$  (в левой части формулы) и некоторого вектора  $y$ . Другими словами, он будет комбинацией столбцов матрицы  $\bar{U}^T$ . Поскольку эти столбцы являются строками матрицы  $\bar{U}$ , вектор  $A^+b$  принадлежит пространству строк матрицы  $\bar{U}$ , т. е. пространству строк матрицы  $A$ . Таким образом, первое условие выполняется.

(2') Умножая  $A^+b$  на матрицу  $A$ , мы надеемся получить вектор  $p$ :

$$AA^+b = (\bar{L}\bar{U}) [\bar{U}^T (\bar{U} \bar{U}^T)^{-1} (\bar{L}^T \bar{L})^{-1} \bar{L}^T b] = \bar{L} (\bar{L}^T \bar{L})^{-1} \bar{L}^T b. \quad (57)$$

Но  $\bar{L} (\bar{L}^T \bar{L})^{-1} \bar{L}^T$  в точности совпадает с матрицей проектирования на пространство столбцов матрицы  $\bar{L}$ . Следовательно, формула (57) дает вектор  $p = Pb$ . Таким образом, вектор  $A^+b$  удовлетворяет обоим условиям оптимальности решения, и справедливость формулы (56) установлена.

**Упражнение 3.4.1.** Найти псевдообратную к нулевой матрице  $A=0$  размера  $m \times n$  и объяснить ход рассуждений при этом.

**Упражнение 3.4.2.** Для матрицы ранга 1 разложения  $A = \bar{L}\bar{U}$  и  $A = uv^T$  совпадают: оба они являются произведениями вектор-столбца на вектор-строку. Для

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{и} \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix} .$$

вычислить матрицу  $A^+$  и вектор  $\bar{x}$ . Спроектировав вектор  $b$  на пространство столбцов, показать, что  $A\bar{x} = p$ .

**Упражнение 3.4.3.** Если матрица  $A$  имеет ортонормированные столбцы, то что можно сказать об псевдообратной к ней?

**Упражнение 3.4.4.** Если матрица  $AA^T$  обратима, то  $A^+ = A^T(AA^T)^{-1}$  и  $\bar{x} = A^+b$ .

(а) Проверить, что  $A\bar{x} = b$ .

(б) Для матрицы  $A = \begin{bmatrix} 1 & 1 \end{bmatrix}$  найти оптимальное решение уравнения  $u + v = 3$ .

**Упражнение 3.4.5.** Первоначально матрица  $A^+$  была введена Пенроузом из других соображений: он доказал, что для произвольной матрицы  $A$  существует, и притом единственная, матрица  $A^+$ , удовлетворяющая четырем условиям

(i)  $AA^+A = A$ , (ii)  $A^+AA^+ = A^+$ ,

(iii)  $(AA^+)^T = AA^+$ , (iv)  $(A^+A)^T = A^+A$ .

Мы предпочли дать более геометрическое и более интуитивное определение (для той же самой матрицы  $A^+$ ), но алгебраичность только что данного определения упрощает проверку того, что данная матрица является псевдообратной.

(а) Показать, что если  $A$  является матрицей проектирования (т. е.  $A^2 = A$  и  $A = A^T$ ), то при  $A^+ = A$  все четыре требования (i)–(iv) будут выполнены. Поэтому псевдообратная совпадает с исходной матрицей  $A$ , как в примере 1 или в утверждении 3.4.1.

(б) Показать, что матрица  $(A^+)^T$  удовлетворяет требованиям для псевдообратной к матрице  $A^T$ , так что  $(A^T)^+ = (A^+)^T$ . (Следствие: если  $A$  — симметрическая матрица, то  $A^+$  также будет симметрической.)

(с) Показать, что псевдообратная к матрице  $A^+$  совпадает с  $A$ , как это было показано выше для матрицы  $\Sigma$  и справедливость чего утверждалась для произвольной матрицы  $A$ .

## § 3.5. ВЗВЕШЕННЫЕ НАИМЕНЬШИЕ КВАДРАТЫ

Вернемся на время к простейшему случаю задачи о наименьших квадратах и оценим средний вес пациента  $\bar{x}$  на основе двух измерений  $x = b_1$  и  $x = b_2$ . При несовпадении результатов измерений мы приходим к несовместной системе двух уравнений относительно одного неизвестного:

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} [x] = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

До сих пор мы считали эти два наблюдения равнозначными и искали величину  $\bar{x}$ , которая давала бы минимум функции  $E^2 = (x - b_1)^2 + (x - b_2)^2$ :

$$\frac{dE^2}{dx} = 0 \quad \text{при} \quad \bar{x} = \frac{b_1 + b_2}{2}.$$

Оптимальная величина  $\bar{x}$  равняется среднему арифметическому измерений.

Предположим теперь, что эти два измерения неравнозначны. Значение  $x = b_1$  могло быть получено в результате более точных измерений (или, в статистической задаче, из большей выборки), чем значение  $x = b_2$ . Но мы все-таки не хотим целиком полагаться на значение  $x = b_1$ , так как в измерении  $x = b_2$  также содержится некоторая информация. Простейший компромисс состоит во введении различных весов  $w_i^2$ , с которыми будут братья наши измерения, и в выборе числа  $x$ , минимизирующего взвешенную сумму квадратов:

$$E^2 = w_1^2(x - b_1)^2 + w_2^2(x - b_2)^2. \quad (58)$$

Если  $w_1 > w_2$ , то доминирующим будет первое измерение, т. е. процесс минимизации будет сильнее уменьшать  $(x - b_1)^2$ . Вычисляя

$$\frac{dE^2}{dx} = 2[w_1^2(x - b_1) + w_2^2(x - b_2)]$$

и приравнявая это выражение нулю, получаем новое решение  $\bar{x}_W$ :

$$\bar{x}_W = \frac{w_1^2 b_1 + w_2^2 b_2}{w_1^2 + w_2^2}. \quad (59)$$

Вместо среднего арифметического чисел  $b_1$  и  $b_2$  для случая  $w_1 = w_2 = 1$   $\bar{x}_W$  будет взвешенным средним значением данных. Это среднее ближе к  $b_1$ , чем к  $b_2$ .

Взвешенные наименьшие квадраты можно трактовать двояко. С одной стороны, можно посмотреть, какая обычная задача о наименьших квадратах будет приводить к функции  $E^2$  вида (58). Эта задача имеет вид

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} [x] = \begin{bmatrix} w_1 b_1 \\ w_2 b_2 \end{bmatrix} \quad \text{с} \quad E^2 = (w_1 x - w_1 b_1)^2 + (w_2 x - w_2 b_2)^2.$$

Это в точности совпадает со взвешенной ошибкой  $E^2$  для исходной задачи. При введении весов мы заменяем исходную систему  $Ax = b$  на новую систему  $WAx = Wb$ , где

$$A = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{и} \quad W = \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \end{bmatrix}.$$

Взвешенное решение  $\bar{x}_W$  представляет собой решение обычной задачи о наименьших квадратах  $WAx = Wb$ .

Такой взгляд приводит к очевидному, но важному следствию: мы должны ожидать изменения решения по методу наименьших квадратов при замене исходной системы на систему  $WAx = Wb$ . Отметим, что при обратимой матрице  $A$  никаких изменений не происходит: в исходном случае имеем  $x = A^{-1}b$ , а с весами  $x = (WA)^{-1}Wb = A^{-1}W^{-1}Wb = A^{-1}b$ . Точные решения совпадают,

в то время как для задачи о наименьших квадратах решения  $\bar{x} = A^+b$  и  $\bar{x}_W = (WA)^+Wb$  различны. Это обстоятельство, вообще говоря, не удивительно, но оно приводит к неприятному следствию.

ЗТ. Основное правило обращения произведения матриц,  $(BA)^{-1} = A^{-1}B^{-1}$ , не выполняется для псевдообратных матриц, т. е. в общем случае  $(WA)^+ \neq A^+W^+$ . (60)

Доказательство. Если правило обращения выполняется для некоторой обратимой матрицы  $W$ , такой, что псевдообратная к ней совпадает с обратной, то взвешенное решение  $\bar{x}_W$  будет иметь вид

$$\bar{x}_W = (WA)^+Wb = A^+W^{-1}Wb = A^+b = \bar{x}.$$

Но  $\bar{x}_W$  не равно  $\bar{x}$ , так что правило действительно не выполняется.

Вторая точка зрения на взвешенные наименьшие квадраты состоит в том, что вместо перехода от системы  $Ax = b$  к  $W Ax = Wb$  мы можем сохранить исходные уравнения, но изменить определение длины. Естественным определением длины для «взвешенной» задачи является взвешенная сумма квадратов:

вектор  $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$  имеет длину  $\|v\|_W = (\omega_1^2 v_1^2 + \omega_2^2 v_2^2)^{1/2}$ . (61)

Сделав такое изменение, будем минимизировать ошибку  $E = \|Ax - b\|_W$  обычным методом. Поскольку  $E^2$  в точности является взвешенной суммой квадратов, ее минимум достигается в точке  $\bar{x}_W$ . При новом определении длины ближайшая к  $b$  точка переместится от  $A\bar{x}$  к  $A\bar{x}_W$ . Вероятно, наилучшим способом иллюстрации этого является перемасштабирование координатных осей (рис. 3.8). Проекция сдвигается вниз по прямой, заменяя

$$\bar{x} = \frac{b_1 + b_2}{2} = \frac{3}{4} \quad \text{на} \quad \bar{x}_W = \frac{\omega_1^2 b_1 + \omega_2^2 b_2}{\omega_1^2 + \omega_2^2} < \frac{3}{4}.$$

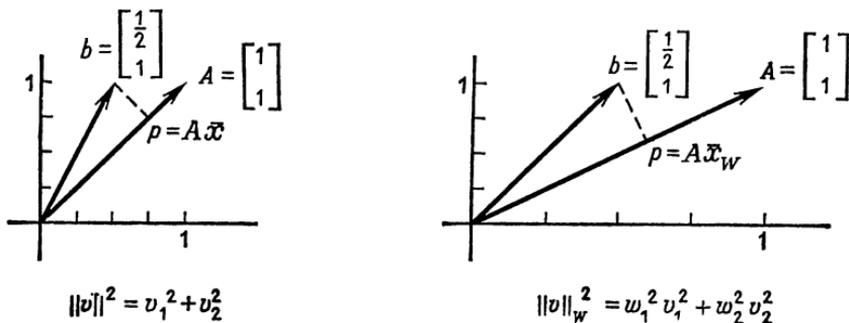


Рис. 3.8. Изменение длины и изменение вектора  $\bar{x}$ .

Поскольку взвешенное среднее оказывается меньше  $\frac{3}{4}$ , эта точка будет ближе к первому измерению  $b_1 = \frac{1}{2}$ , чем ко второму  $b_2 = 1$ . Растягивая одну координатную ось по сравнению с другой, мы тем самым придаем ей большее значение.

### ВЕСА, ДЛИНЫ И СКАЛЯРНЫЕ ПРОИЗВЕДЕНИЯ

В следующих абзацах мы хотим показать, что рассмотренный пример является типичным для всех задач о взвешенных наименьших квадратах. Одновременно мы рассмотрим все возможные определения скалярного произведения двух векторов и соответствующие им определения длины. К счастью, такое глобальное обобщение очень легко осуществить.

Начнем с задачи о наименьших квадратах  $Ax = b$ . Если  $m$  наблюдений  $b_1, \dots, b_m$  неравнозначны, то с  $m$  уравнениями будут связаны  $m$  различных весов  $w_1, \dots, w_m$ . Это обобщение исходного примера, но на самом деле здесь имеется и более глубокое обобщение. Кроме неравнозначности, наблюдения *могут не быть независимыми*. В этом случае мы также вводим коэффициенты  $w_{ij}$ , характеризующие связь  $i$ -го и  $j$ -го наблюдений. Тогда система

$$Ax = b \quad \text{заменяется на} \quad W Ax = W b. \quad (62)$$

Числа  $w_1, \dots, w_n$  располагаются на главной диагонали матрицы  $W$  (теперь мы будем обозначать их через  $w_{11}, \dots, w_{nn}$ ), а связывающие коэффициенты, если таковые вообще имеются, являются внедиагональными элементами матрицы.

*Замечание.* В статистике это приблизительно соответствует введению *коэффициентов регрессии*, или, после нормализации, *коэффициентов корреляции*. Точная связь между матрицей весов  $W$  и статистической матрицей ковариации  $V$  приводится в конце этой главы.

Для решения новой задачи  $W Ax = W b$  нужно лишь вернуться к нормальным уравнениям  $A^T A \bar{x} = A^T b$  исходной задачи и сделать соответствующие изменения, т. е. заменить  $A$  на  $W A$ , а  $b$  на  $W b$ .

**3U.** Если столбцы матрицы  $A$  линейно независимы (ее ранг равен  $n$ ) и матрица  $W$  обратима, то решение по методу наименьших квадратов для задачи  $W Ax = W b$  находится из взвешенных нормальных уравнений:

$$(A^T W^T W A) \bar{x}_W = A^T W^T W b. \quad (63)$$

Если обозначить матрицу  $W^T W$  через  $H$ , то

$$\bar{x}_W = (A^T H A)^{-1} A^T H b.$$

Этим завершается первый подход к методу взвешенных наименьших квадратов, и теперь мы переходим ко второму подходу. Вспомним, что в примере, с которого мы начинали этот параграф, основная идея заключалась в изменении определения длины. Прделаем то же самое и для более общей матрицы весов  $W$ :

**3V.** Для произвольной обратимой матрицы  $W$  можно определить новую длину и скалярное произведение по следующим правилам: новая длина вектора  $x$  равняется старой длине вектора  $Wx$ , а новое скалярное произведение векторов  $x$  и  $y$  равняется старому скалярному произведению векторов  $Wx$  и  $Wy$ :

$$\|x\|_W = \|Wx\| \quad \text{и} \quad (x, y)_W = (Wx)^T (Wy) = x^T W^T W y \quad (64)$$

Длина и скалярное произведение связаны обычным правилом

$$\|x\|_W^2 = (x, x)_W > 0 \quad \text{для всех ненулевых векторов } x. \quad (65)$$

Это определение фактически описывает все возможные скалярные произведения, т. е. все способы получения числа, которое линейно зависит от векторов  $x$  и  $y$  и положительно при  $x = y \neq 0$ . Именно это условие положительности приводит к требованию обратимости матрицы  $W$ , поскольку в противном случае в нуль-пространстве этой матрицы найдется ненулевой вектор  $x$ , и длина этого вектора будет равняться  $\|x\|_W = \|Wx\| = \|0\| = 0$ . Мы не допускаем, чтобы ненулевой вектор имел нулевую длину.

Вернемся к задаче о наименьших квадратах  $Ax = b$  с матрицей весов  $W$ . Можно получить геометрическое истолкование этой задачи, используя новые определения длины и скалярного произведения. Оптимальный выбор вектора  $Ax$ , минимизирующего ошибку  $Ax - b$ , — это вновь точка  $p$  в пространстве столбцов матрицы  $A$ , ближайшая к  $b$ . Но слово «ближайшая» здесь нужно понимать в терминах нового определения длины. Теперь мы минимизируем величину  $\|Ax - b\|_W = \|W Ax - W b\|$ . Эта точка  $p$  (или, вернее,  $p_W$ , поскольку она зависит от матрицы  $W$ ) вновь оказывается проекцией точки  $b$  на пространство столбцов, а вектор ошибки  $b - p_W$  ортогонален этому пространству. Имеется, однако, и изменение: ортогональность теперь означает равенство нулю нового скалярного произведения. Поэтому точка  $p_W = \bar{A}x_W$  определяется из условия, что вектор  $b - p_W$  перпендикулярен всем векторам  $Ay$  из пространства столбцов:

$$(Ay)^T W^T W (b - \bar{A}x_W) = 0 \quad \text{для всех } y.$$

<sup>1)</sup> Матрица  $H = W^T W$  при обратимой матрице  $W$  — это в точности положительно определенная матрица из гл. 6.

Вектор, на который умножается  $y^T$ , должен равняться нулю,

$$A^T W^T W (b - A \bar{x}_W) = 0, \quad \text{или} \quad A^T W^T W A \bar{x}_W = A^T W^T W b,$$

и мы вновь приходим к взвешенным нормальным уравнениям (63).

Будем пессимистами и, не веря всему сказанному, попытаемся изобразить (рис. 3.9), как выглядит эта новая перпендикулярность. Мы можем либо растягивать пространство в соответствии с матрицей  $W$  (рис. 3.8), либо сохранить исходные оси и стра-

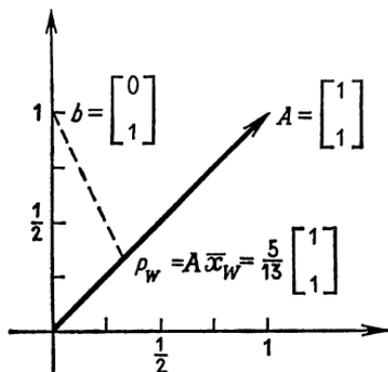


Рис. 3.9. Взвешенные наименьшие квадраты и  $W$ -перпендикулярность.

дать от созерцания углов, которые не выглядят прямыми, но считаются таковыми.

Задача, соответствующая этому рисунку, такова:

$$W A x = W b, \quad \text{или} \quad \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} [x] = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Взвешенные нормальные уравнения имеют вид

$$(W A)^T W A \bar{x}_W = (W A)^T W b,$$

или

$$\begin{bmatrix} 3 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} \bar{x}_W = \begin{bmatrix} 3 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \bar{x}_W = \frac{5}{13}.$$

В качестве последнего замечания о методе наименьших квадратов сравним наш подход с подходом, который обычно применяется в статистике. Предполагается, что ошибки для  $m$  уравнений  $Ax = b$  в среднем дают нуль и матрицу ковариации  $V$ . Тогда оптимальная оценка  $\bar{x}$  является решением системы нормальных уравнений  $A^T V^{-1} A \bar{x} = A^T V^{-1} b$  и эти уравнения совпадают.

с (63) при условии, что в качестве  $H = W^T W$  выбрана обратная к  $V$  матрица.

**Упражнение 3.5.1.** Для матрицы  $W = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$  найти оптимальный вектор  $\bar{x}_W$ , когда наблюдение  $x=4$  берется с вдвое большим весом, чем наблюдение  $x=1$ :  $\omega_1=2$  и  $\omega_2=1$ .

**Упражнение 3.5.2.** При той же самой матрице  $W$  какие векторы  $W$ -перпендикулярны к вектору  $b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ? Чему равняется взвешенная длина  $\|b\|_W$  вектора  $b$ ?

**Упражнение 3.5.3.** Если  $W = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ , то чему равняется оптимальное среднее  $\bar{x}_W$  наблюдений  $x=1$  и  $x=0$ ?

**Упражнение 3.5.4.** Для какой матрицы  $W$  новая длина и новое скалярное произведение совпадают с исходными?

## ОБЗОРНЫЕ УПРАЖНЕНИЯ

3.1. Как связано произведение  $(ABC)^T$  с каждой из матриц  $A^T$ ,  $B^T$  и  $C^T$ ?

3.2. Записать матрицу ранга 1

$$A = \begin{bmatrix} 2 & 3 \\ 6 & 9 \end{bmatrix}$$

и транспонированную  $A^T$  в виде  $uv^T$ .

3.3. Чему равен угол между векторами  $a=(2, -2, 1)$  и  $b=(1, 2, 2)$ ?

3.4. Чему равна проекция  $p$  вектора  $b=(1, 2, 2)$  на вектор  $a=(2, -2, 1)$ ?

3.5. Вычислить косинус угла между векторами  $(3, 4)$  и  $(4, 3)$ .

3.6. Чему равна проекция вектора  $b=(1, 1, 1)$  на плоскость, порожденную векторами  $(1, 0, 0)$  и  $(1, 1, 0)$ ?

3.7. Какой вид имеют нормальные уравнения для системы  $u=0, v=2, u+v=4$ ? Чему равны  $\bar{u}$  и  $\bar{v}$ ?

3.8. Построить матрицу проектирования  $P$  на пространство, порожденное векторами  $(1, 1, 1)$  и  $(0, 1, 3)$ .

3.9. Какой вид имеет прямая наилучшего приближения для данных:  $y=0$  при  $x=0, y=0$  при  $x=1, y=12$  при  $x=3$ ?

3.10. Найти прямую вида  $y=Dt$ , дающую наилучшее приближение для данных:  $y=1, t=1$  и  $y=1, t=2$ .

3.11. Какая константа наилучшим образом приближает функцию  $y=x^4$  в смысле метода наименьших квадратов на отрезке  $0 \leq x \leq 1$ ?

3.12. Если матрица  $Q$  ортогональна, то будет ли ортогональна матрица  $Q^3$ ?

3.13. Выписать все ортогональные матрицы третьего порядка, состоящие только из нулей и единиц.

3.14. Какое кратное вектора  $a_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$  нужно вычесть из  $a_2 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$ , чтобы получившийся в результате вектор был ортогонален к  $a_1$ ? Изобразить это на рисунке.

3.15. Нормализовать ортогональные векторы, полученные в предыдущем упражнении. Это завершит процесс Грама—Шмидта. Представить матрицу

$$A = \begin{bmatrix} 1 & 4 \\ 1 & 0 \end{bmatrix}$$

в виде  $A = QR$ .

3.16. Представить матрицу

$$\begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & 0 \end{bmatrix}$$

в виде  $QR$ , учитывая, что ее первый столбец уже является вектором единичной длины.

3.17. Найти псевдообратную к матрице  $A = [3 \ 0]$ .

3.18. Чему равняется скалярное произведение векторов  $(1, 0)$  и  $(0, 1)$  при матрице весов

$$W = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}.$$

3.19. Если каждый из элементов ортогональной матрицы равен  $1/4$  либо  $-1/4$ , то каков размер этой матрицы?

# Глава 4

## ОПРЕДЕЛИТЕЛИ

### § 4.1. ВВЕДЕНИЕ

Трудно решить, что именно следует сказать об определителях. Лет семьдесят назад казалось, что определители важнее матриц, из которых они получаются, и была опубликована четырехтомная «История определителей» Мура. Направление развития математики, однако, меняется, и сейчас теория определителей достаточно далека от основных проблем линейной алгебры. Действительно, одно число слишком слабо характеризует матрицу.

Одна из точек зрения состоит в том, что определитель дает явную «формулу», сжатое и конкретное выражение в замкнутой форме для таких величин, как  $A^{-1}$ . На самом деле эта формула не используется для вычисления  $A^{-1}$  или  $A^{-1}b$ . Более того, сам определитель находится с помощью метода исключения, который может трактоваться как наиболее эффективный способ для подстановки элементов заданной матрицы в формулу для определителя. Общая формула важна потому, что она показывает зависимость величины определителя от каждого из  $n^2$  элементов матрицы.

Перечислим основные применения определителей.

1. Определитель дает критерий для проверки обратимости матрицы. *Если определитель матрицы  $A$  равен нулю, то эта матрица вырожденна, а если  $\det A \neq 0$ , то  $A$  обратима.* Этот критерий наиболее важен применительно к семейству матриц  $A - \lambda I$  (чем и объясняется существенность данной главы для всей книги в целом). Параметр  $\lambda$  вычитается из всех элементов главной диагонали, и наша задача заключается в отыскании тех значений  $\lambda$  (*собственных значений*), при которых матрица  $A - \lambda I$  будет вырожденной. Приведенный критерий показывает, что для этого достаточно выяснить, когда определитель матрицы  $A - \lambda I$  будет равен нулю. В дальнейшем мы увидим, что этот определитель является многочленом степени  $n$  от  $\lambda$  и, следовательно, с учетом кратности, имеет в точности  $n$  корней; таким образом, матрица

(порядка  $n$ ) имеет  $n$  собственных значений. Это один из фактов, получаемых с помощью формулы для определителя, а не непосредственным вычислением.

2. Определитель матрицы  $A$  равен объему параллелепипеда  $P$  в  $n$ -мерном пространстве, построенного на векторах, совпадающих со строками матрицы  $A$  (рис. 4.1)<sup>1</sup>). Необходимость в вычисле-

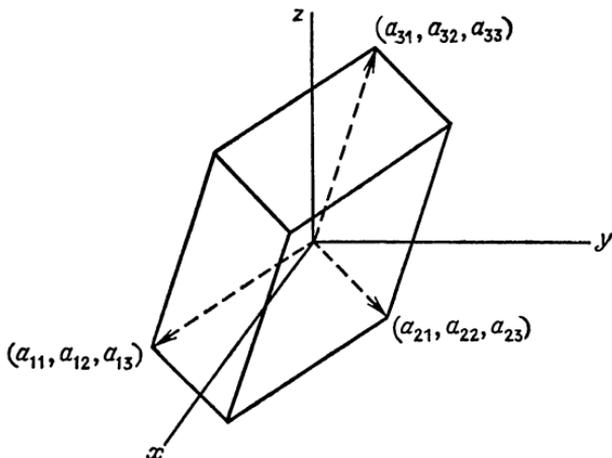


Рис. 4.1. Параллелепипед, образованный строками матрицы  $A$ .

нии такого рода объемов может показаться странной, но на самом деле  $P$  обычно является элементом объема в кратном интеграле. В простейшем случае это кубик  $dV = dx dy dz$  в интеграле  $\iiint f(x, y, z) dV$ . Если для упрощения подынтегрального выражения мы осуществим замену переменных  $x, y, z$  на  $x', y', z'$  по формулам  $x = x' \cos y', y = x' \sin y', z = z'$ , то элемент объема  $dx dy dz$  перейдет в  $J dx' dy' dz'$  (подобно тому, как это происходит в одномерном случае, когда дифференциал  $dx$  преобразуется в  $(dx/dx') dx'$ ). Здесь  $J$  есть *определитель Якоби (якобиан)*, являющийся трехмерным аналогом коэффициента подобия  $dx/dx'$ ,

$$J = \begin{vmatrix} \partial x/\partial x' & \partial x/\partial y' & \partial x/\partial z' \\ \partial y/\partial x' & \partial y/\partial y' & \partial y/\partial z' \\ \partial z/\partial x' & \partial z/\partial y' & \partial z/\partial z' \end{vmatrix} = \begin{vmatrix} \cos y' & -x' \sin y' & 0 \\ \sin y' & x' \cos y' & 0 \\ 0 & 0 & 1 \end{vmatrix},$$

и он оказывается равным  $x'$ . В полярных, или, точнее, в цилиндрических координатах, где элемент объема равен  $r dr d\theta dz$ , это  $r$ . (Если мы попытаемся нарисовать «кубик»  $r dr d\theta dz$ , то он будет

<sup>1</sup>) В качестве векторов для построения параллелепипеда можно взять и столбцы матрицы  $A$ , в результате чего получится совсем другой параллелепипед того же объема.

выглядеть искривленным, но при уменьшении длин сторон будет «выпрямляться».)

3. Определитель дает явную формулу для ведущих элементов матрицы. Теоретически мы могли бы пользоваться им, чтобы выяснить заранее, когда ведущий элемент будет равен нулю и потребуются перестановка строк матрицы. Гораздо важнее, однако, что из формулы

$$\det A = \pm (\text{произведение ведущих элементов})$$

вытекает следующее свойство: *независимо от порядка шагов в методе исключения произведение ведущих элементов неизменно с точностью до знака*. Раньше из этого делали вывод о бесполезности перестановки строк матрицы для устранения слишком маленького ведущего элемента, считая, что он все равно должен появиться. Но на самом деле (и это часто происходит), если мы не исключаем «ненормально» малый ведущий элемент, то за ним вскоре следует «ненормально» большой; в результате их произведение приходит в норму, но численное решение оказывается неверным.

4. Определитель показывает зависимость величины  $A^{-1}b$  от каждого элемента вектора  $b$ . Если в ходе эксперимента изменяется один из параметров или уточняется одно из наблюдений, то «коэффициентом влияния» этих изменений на  $x = A^{-1}b$  является отношение определителей.

Имеется еще одна проблема, связанная с определителями. Трудно не только решить, насколько они важны и каково их место в линейной алгебре, но и выбрать их определение. Ясно, что  $\det A$  не будет очень простой функцией от  $n^2$  переменных; в противном случае найти  $A^{-1}$  было бы значительно проще, чем это есть на самом деле. Явная формула, приведенная в § 4.3, потребует тщательного разъяснения, и ее связь с обратной матрицей далеко не очевидна.

*Из связанных с определителями вопросов наиболее простым оказывается выяснение их свойств, а не нахождение явного выражения для них.* Поэтому естественно начать изложение следующим образом. Определитель может быть (и будет) введен при помощи трех основных его свойств, и задача состоит в том, как, используя эти свойства, вычислить значение определителя. Такой подход вернет нас к методу исключения Гаусса и к произведению ведущих элементов. Теоретически наиболее трудным здесь будет доказательство того факта, что независимо от порядка использования упомянутых выше свойств, результат всегда будет один и тот же и эти свойства определителя образуют непротиворечивую систему.

В следующем параграфе перечисляются основные свойства определителя и наиболее важные следствия из них. В § 4.3 дается

несколько возможных формул для определителя: явная формула с  $n!$  членами, рекуррентная формула и формула, использующая ведущие элементы (именно она и применяется на практике при вычислении определителей больших матриц). В § 4.4 определитель используется для вычисления  $A^{-1}$  и решения уравнения  $x = A^{-1}b$  (*правило Крамера*). И, наконец, в дополнительном замечании о перестановках мы докажем, что выбранные свойства определителя непротиворечивы, откуда будет следовать корректность такого выбора.

## § 4.2. СВОЙСТВА ОПРЕДЕЛИТЕЛЯ

Список этих свойств занимает довольно много места, однако они весьма просты и легко иллюстрируются на матрицах второго порядка. Поэтому будем проверять, что известная формула для определителя

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

обладает каждым из перечисленных ниже свойств. (Существует два общепринятых обозначения для определителя матрицы  $A$ :  $\det A$  и  $|A|$ .) Мы покажем также, что все свойства определителей, начиная с четвертого, можно получить из предыдущих; таким образом, каждое из этих свойств будет следствием первых трех.

1. *Определитель является линейной функцией своей первой строки.* Это означает, что если имеются матрицы  $A$ ,  $B$  и  $C$ , отличающиеся друг от друга только элементами своих первых строк, и первая строка матрицы  $A$  является линейной комбинацией первых строк матриц  $B$  и  $C$ , то  $\det A$  есть линейная комбинация с теми же коэффициентами определителей  $\det B$  и  $\det C$ . Поскольку термин «линейная комбинация» подразумевает как умножение векторов на числа, так и сложение векторов, мы получаем две формулы:

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

$$\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = tad - tbc = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

Подчеркнем, что эта формула в корне отлична от неверной формулы  $\det(B+C) = \det B + \det C$ ; изменения допускаются лишь в одной строке матрицы.

2. При перестановке двух строк матрицы ее определитель меняет знак.

$$\begin{vmatrix} c & d \\ a & b \end{vmatrix} = cb - ad = - \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

Отсюда следует, что свойство 1 относится не только к первой, но и к любой другой строке матрицы, так что определитель является линейной функцией каждой своей строки.

3. Определитель единичной матрицы равен 1.

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1.$$

Это свойство является нормализующим; два предыдущих дают «одномерное пространство» определителей, а это выделяет среди них один.

4. Если матрица  $A$  имеет две одинаковые строки, то  $\det A = 0$ .

$$\begin{vmatrix} a & b \\ a & b \end{vmatrix} = ab - ba = 0.$$

Доказательство следует из свойства 2; при перестановке двух строк матрицы знак определителя должен измениться на противоположный, но, с другой стороны, он должен остаться прежним, поскольку матрица не изменится. Следовательно,  $\det A = 0$ . (Рассуждение теряет силу в булевой алгебре, где  $1 = -1$ . В этом случае основное свойство 2 должно быть заменено свойством 4.)

5. Элементарное преобразование, состоящее в вычитании строки, умноженной на постоянную, из другой строки, не изменяет величины определителя.

$$\begin{vmatrix} a - lc & b - ld \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

Используя свойство 1, мы получаем дополнительный член  $l \begin{vmatrix} c & d \\ c & d \end{vmatrix}$ , но из свойства 4 следует, что он равен нулю.

6. Если одна из строк матрицы  $A$  нулевая, то  $\det A = 0$ .

$$\begin{vmatrix} 0 & 0 \\ c & d \end{vmatrix} = 0.$$

Одно из доказательств состоит в том, что к нулевой строке прибавляется какая-либо другая строка матрицы. При этом определитель по свойству 5 остается неизменным, но в матрице появляются две одинаковые строки и по свойству 4  $\det A = 0$ .

7. Если матрица  $A$  треугольная, то  $\det A$  равен произведению  $a_{11}a_{22}\dots a_{nn}$  элементов главной диагонали. В частности, если главная диагональ матрицы состоит из единиц, то  $\det A = 1$ .

$$\begin{vmatrix} a & b \\ 0 & d \end{vmatrix} = ad, \quad \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} = ad.$$

Доказательство. Пусть диагональные элементы матрицы отличны от нуля. Тогда элементарными преобразованиями можно исключить все внедиагональные элементы, не изменяя при этом величины определителя (по свойству 5). В результате получим диагональную матрицу

$$D = \begin{bmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{nn} \end{bmatrix}.$$

(Это обычный метод исключения, если  $A$  — нижняя треугольная матрица; если же  $A$  — верхняя треугольная матрица, то каждая умноженная на некоторый коэффициент строка вычитается из строк, стоящих выше нее). Для вычисления определителя матрицы  $D$  нужно последовательно применять свойство 1; вынося каждый элемент  $a_{ii}$ , получим единичную матрицу, определитель которой нам известен:

$$\det D = a_{11}a_{22}\dots a_{nn} \det I = a_{11}a_{22}\dots a_{nn}.$$

Если какой-либо диагональный элемент матрицы  $A$  равен нулю, то эта матрица вырожденная и, согласно следующему свойству, ее определитель равен нулю.

8. Если матрица  $A$  вырожденна, то  $\det A = 0$ . Если  $A$  невырожденна, то  $\det A \neq 0$ .

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ вырожденна тогда и только тогда, когда } ad - bc = 0.$$

Если  $A$  — вырожденная матрица, то элементарными преобразованиями она приводится к матрице  $U$  с нулевой строкой. Тогда по свойствам 5 и 6  $\det A = \det U = 0$ . Обратно, предположим, что матрица  $A$  невырожденна. Тогда элементарными преобразованиями и перестановкой строк она приводится к верхней треугольной матрице  $U$  с ненулевыми ведущими элементами  $d_1, \dots, d_n$  на главной диагонали и по свойству 7  $\det A = \pm \det U = \pm d_1 \dots d_n$ . Знак плюс или минус зависит, согласно свойству 2, от того, является число перестановок строк четным или нечетным.

9. Для любых двух матриц размера  $n \times n$  определитель их произведения равен произведению определителей:  $\det AB = (\det A)(\det B)$ .

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

Это эквивалентно тождеству

$$(ad - bc)(eh - fg) = (ae + bg)(cf + dh) - (af + bh)(ce + dg).$$

В частности, если матрица  $A$  обратима, то  $\det A \det A^{-1} = \det AA^{-1} = \det I = 1$  или  $\det A^{-1} = 1/\det A$ .

Мы приведем два возможных доказательства этого наименее очевидного пока свойства, предполагая, что матрицы  $A$  и  $B$  невырожденные (в противном случае матрица  $AB$  будет вырожденной и равенство  $\det AB = \det A \det B$  легко устанавливается: по свойству 8 оно эквивалентно равенству  $0 = 0$ ).

(i) Можно рассмотреть величину  $d(A) = \det AB / \det B$  и доказать, что она обладает свойствами 1—3; тогда, поскольку определители *вводятся* при помощи этих трех свойств, величина  $d(A)$  должна равняться  $\det A$ . Например, если  $A$  является единичной матрицей, то  $d(I) = \det B / \det B = 1$ ; следовательно, величина  $d(A)$  обладает свойством 3. При перестановке двух строк матрицы  $A$  то же самое произойдет с соответствующими строками матрицы  $AB$  и знак величины  $d$  изменится в соответствии со свойством 2. И, наконец, линейная комбинация в первой строке матрицы  $A$  даст такую же линейную комбинацию в первой строке матрицы  $AB$ . Поэтому свойство 1 для определителя матрицы  $AB$ , деленного на фиксированную величину  $\det B$ , приводит к свойству 1 для их отношения  $d(A)$ . Следовательно,  $d(A)$  совпадает с определителем матрицы  $A$  и  $\det AB / \det B = \det A$ .

(ii) Второе доказательство менее элегантно, чем первое. Рассмотрим сначала случай диагональной матрицы  $D$ . Тогда равенство  $\det DB = \det D \det B$  следует из свойства 1 и получается вынесением каждого диагонального элемента  $d_i$  из своей строки. В случае произвольной матрицы  $A$  мы приводим ее к  $D$  последовательностью исключений по Гауссу — Жордану, сначала переходя от  $A$  к  $U$  в обычном порядке, а затем от  $U$  к  $D$ , действуя каждой строкой на расположенные выше строки. Определитель при этом остается неизменным с точностью до знака, который меняется на противоположный при каждой перестановке строк. Те же шаги переводят матрицу  $AB$  в  $DB$ , точно так же действуя на определитель, а для произведения  $DB$  справедливость свойства 9 уже установлена.

10. При транспонировании матрицы ее определитель не изменяется:  $\det A^T = \det A$ .

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix}.$$

Здесь вновь следует выделить вырожденный случай; матрица  $A$  вырожденна тогда и только тогда, когда  $A^T$  вырожденна, и мы имеем  $0=0$ . Если матрица  $A$  невырожденная, то она допускает разложение  $PA=LDU$ , и можно применить свойство 9 для определителя произведения:

$$\det P \det A = \det L \det D \det U. \quad (1)$$

Транспонирование матрицы  $PA=LDU$  дает  $A^T P^T = U^T D^T L^T$ , и по свойству 9

$$\det A^T \det P^T = \det U^T \det D^T \det L^T. \quad (2)$$

На самом деле это выражение проще, чем может показаться, поскольку матрицы  $L$ ,  $U$ ,  $L^T$  и  $U^T$  являются треугольными матрицами с единицами вдоль главной диагонали. Поэтому по свойству 7 их определители равны единице. Кроме того, любая диагональная матрица совпадает со своей транспонированной, т. е.  $D^T = D$ . Поэтому остается вычислить лишь определитель матрицы перестановки.

Ясно, что определитель матрицы  $P$  равен либо 1, либо  $-1$ , поскольку она получается из единичной матрицы последовательностью перестановок строк. Заметим, что для любой матрицы перестановки мы имеем  $PP^T = I$ . (При перемножении матриц  $P$  и  $P^T$  единица в первой строке матрицы  $P$  умножается на единицу в первом столбце матрицы  $P^T$  и на нули в остальных столбцах.) Следовательно,  $\det P \det P^T = \det I = 1$  и матрицы  $P$  и  $P^T$  должны иметь одинаковые определители, равные либо 1, либо  $-1$ .

Для завершения доказательства заметим, что произведения в (1) и (2) равны между собой и, следовательно,  $\det A = \det A^T$ . Этот факт практически удваивает число свойств определителя, поскольку каждое из перечисленных выше свойств оказывается справедливым и для столбцов матрицы; *определитель меняет знак при перестановке двух столбцов матрицы, наличие двух равных столбцов (или столбца из нулей) дает определитель, равный нулю, и определитель линейно зависит от каждого отдельного столбца*. Доказательство заключается в транспонировании матрицы и проверке свойств для строк.

На наш взгляд, пора остановиться и объявить список свойств определителя полным. Остается лишь найти формулу для определителя и использовать ее на практике.

**Упражнение 4.2.1.** Как связаны  $\det(2A)$ ,  $\det(-A)$  и  $\det(A^2)$  с  $\det A$ ?

**Упражнение 4.2.2.** На примере матрицы второго порядка показать, что перестановка  $i$ -й и  $j$ -й строк может быть осуществлена путем прибавления  $i$ -й строки к  $j$ -й, вычитания новой  $j$ -й строки из  $i$ -й, прибавления новой  $i$ -й строки к  $j$ -й и, наконец, умножения  $i$ -й строки на  $-1$ . Какие свойства тогда можно использовать для получения свойства 2?

**Упражнение 4.2.3.** Применяя преобразования строк для нахождения верхних треугольных матриц  $U$ , вычислить определители

$$\det \begin{bmatrix} 1 & 2 & -2 & 0 \\ 2 & 3 & -4 & 1 \\ -1 & -2 & 0 & 2 \\ 0 & 2 & 5 & 3 \end{bmatrix}, \quad \det \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}.$$

*Замечание.* Многим читателям известна формула для вычисления определителей 3-го порядка, содержащая шесть членов (см. формулу (5) ниже). Естественно ожидать аналогичного правила для определителей 4-го порядка, и такая формула есть. Но она содержит  $4! = 24$  члена, а не 8, как может показаться по аналогии с определителем 3-го порядка. Именно поэтому мы и рекомендуем использовать преобразования строк и приведение к треугольной матрице  $U$  при вычислении определителя.

**Упражнение 4.2.4.** Кососимметрическая матрица удовлетворяет равенству  $A^T = -A$ . Пусть

$$A = \begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}.$$

Требуется показать, что  $\det A = 0$  (сравнивая его с  $\det A^T$  и  $\det(-A)$ ). Показать также, что для кососимметрической матрицы второго порядка  $\det A$  может быть отличен от нуля; для равенства нулю определителя порядок матрицы должен быть нечетным.

**Упражнение 4.2.5.** Вычислить определители:

а) произведения матриц ранга один

$$A = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} [2 \quad -1 \quad 2]$$

б) верхней треугольной матрицы

$$U = \begin{bmatrix} 4 & 4 & 8 & 8 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 0 & 2 \end{bmatrix};$$

в) нижней треугольной матрицы  $U^T$ ;

д) обратной матрицы  $U^{-1}$ ;

е) «обратно-треугольной» матрицы, получаемой перестановкой строк,

$$M = \begin{bmatrix} 0 & 0 & 0 & 2 \\ 0 & 0 & 2 & 6 \\ 0 & 1 & 2 & 2 \\ 4 & 4 & 8 & 8 \end{bmatrix}.$$

**Упражнение 4.2.6.** Найти другой способ получения свойства 6 из предыдущих.

**Упражнение 4.2.7.** Пусть  $Q$  — ортогональная матрица, т. е.  $Q^T Q = I$ . Доказать, что  $\det Q$  равен либо  $+1$ , либо  $-1$ . Какой параллелепипед образуют строки (или столбцы) матрицы  $Q^2$ ?

**Упражнение 4.2.8.** Используя преобразования строк, проверить, что определитель Вандермонда 4-го порядка равняется

$$V_4 = \begin{vmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \\ 1 & x_4 & x_4^2 & x_4^3 \end{vmatrix} = (x_2 - x_1)(x_3 - x_1)(x_4 - x_1)(x_3 - x_2)(x_4 - x_2)(x_4 - x_3).$$

### § 4.3. ФОРМУЛЫ ДЛЯ ОПРЕДЕЛИТЕЛЯ

Первая формула нами практически уже получена:

**4А.** Если матрица  $A$  невырожденная, то  $A = P^{-1}LDU$  и

$$\begin{aligned} \det A &= \det P^{-1} \det L \det D \det U = \\ &= \pm (\text{произведение ведущих элементов}). \end{aligned} \quad (3)$$

Знак плюс или минус дается определителем матрицы  $P^{-1}$  (или  $P$ ) и зависит от того, является число перестановок строк в приведении четным или нечетным. Для треугольных сомножителей имеем  $\det L = \det U = 1$  и  $\det D = d_1 \dots d_n$ .

В двумерном случае стандартное  $LDU$ -разложение имеет вид

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ c/a & 1 \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & (ad - bc)/a \end{bmatrix} \begin{bmatrix} 1 & b/a \\ 0 & 1 \end{bmatrix} \quad (4)$$

и произведение ведущих элементов равно  $ad - bc$ . Если первый шаг состоял в перестановке строк, то

$$PA = \begin{bmatrix} c & d \\ a & b \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ a/c & 1 \end{bmatrix} \begin{bmatrix} c & 0 \\ 0 & (cb - da)/c \end{bmatrix} \begin{bmatrix} 1 & d/c \\ 0 & 1 \end{bmatrix}$$

и произведение ведущих элементов теперь равно  $-\det A$ .

**Пример.** Матрица конечных разностей в § 1.6 представлялась в виде  $A = LDU$ :

$$\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & \cdot & \\ & & \cdot & \ddots & -1 \\ & & & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ -\frac{1}{2} & 1 & & & \\ & -\frac{2}{3} & 1 & & \\ & & \cdot & \cdot & \\ & & & \frac{1-n}{n} & 1 \end{bmatrix} \begin{bmatrix} 2 \\ \frac{3}{2} \\ \frac{4}{3} \\ \cdot \\ \frac{n+1}{n} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & & & \\ & 1 & -\frac{2}{3} & & \\ & & 1 & \cdot & \\ & & & \cdot & \frac{1-n}{n} \\ & & & & 1 \end{bmatrix}$$

Ее определитель равен произведению ведущих элементов:

$$\det A = 2 \left(\frac{3}{2}\right) \left(\frac{4}{3}\right) \dots \left(\frac{n+1}{n}\right) = n+1.$$

Именно таким образом обычно и вычисляются определители, кроме некоторых частных случаев. Фактически понятие ведущих элементов появилось в результате систематического рассмотрения информации, первоначально распределенной на все  $n^2$  элементов матрицы. Однако с теоретической точки зрения такая концентрация информации в ведущих элементах имеет недостаток, заключающийся в невозможности установить, как изменение одного из элементов матрицы влияет на величину определителя. Поэтому мы попытаемся найти явное выражение для определителя через все  $n^2$  элементов матрицы. Для  $n=2$  будет показана справедливость формулы  $ad-bc$ . Для  $n=3$  соответствующая формула также хорошо известна:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = + a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - \quad (5) \\ - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.$$

Наша цель состоит в том, чтобы вывести эти формулы непосредственно из определяющих свойств 1—3 предыдущего параграфа. Если мы сумеем сделать это для  $n=2$  и  $n=3$  достаточно систематическим образом, то это укажет путь для нахождения определителей матриц большего порядка.

Начнем с того, что каждую строку матрицы  $A$  можно разложить на сумму векторов по направлениям координатных осей; для  $n=2$  это разбиение имеет вид

$$[a \ b] = [a \ 0] + [0 \ b] \quad \text{и} \quad [c \ d] = [c \ 0] + [0 \ d].$$

Затем мы воспользуемся основным свойством линейности по каждой из строк матрицы — сначала по строке 1, а затем по строке 2:

$$\begin{aligned} \begin{vmatrix} a & b \\ c & d \end{vmatrix} &= \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} = \\ &= \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix}. \end{aligned} \quad (6)$$

Для матрицы размера  $n \times n$  каждая строка раскладывается по  $n$  направлениям. Тогда разложение, аналогичное (6), будет содержать  $n^n$  членов: в нашем случае  $2^2 = 4$ . Однако большинство из них (как первый и последний в приведенном разложении) автоматически равняются нулю. Если векторы двух строк окажутся направленными одинаково, один из них будет равен другому, умноженному на постоянную, и

$$\begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} = 0, \quad \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix} = 0.$$

Здесь имеются столбцы из нулей, и определители равняются нулю. Поэтому нас будут интересовать *только те определители, у матриц которых вектор-строки имеют разные направления*, т. е. ненулевые члены присутствуют в *разных столбцах*. Пусть в первой строке матрицы ненулевой элемент находится на месте  $\alpha$ , во второй строке — на месте  $\beta$  и, наконец, в  $n$ -й строке ненулевой элемент находится на месте  $\nu$ , причем все номера столбцов  $\alpha, \beta, \dots, \nu$  различные. Тогда этот набор является перенумерацией, или перестановкой, набора чисел  $1, 2, \dots, n$ . Для матрицы размера  $3 \times 3$  имеется шесть членов такого вида:

$$\begin{aligned} \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= \begin{vmatrix} a_{11} & & \\ & a_{22} & \\ & & a_{33} \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ & a_{23} & \\ a_{31} & & \end{vmatrix} + \begin{vmatrix} & & a_{13} \\ a_{21} & & \\ & a_{32} & \end{vmatrix} + \\ &+ \begin{vmatrix} a_{11} & & \\ & a_{23} & \\ & & a_{32} \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ a_{21} & & \\ & a_{33} & \end{vmatrix} + \begin{vmatrix} & & a_{13} \\ & a_{22} & \\ a_{31} & & \end{vmatrix}. \end{aligned} \quad (7)$$

Отметим еще раз, что разложение вида (6) содержало бы  $3^3 = 27$  членов, однако все они, кроме  $3! = 6$ , равны нулю благодаря совпадению направлений двух вектор-строк. В общем случае мы

имеем  $n$  возможностей для выбора позиции  $\alpha$ ,  $n-1$  возможностей для выбора  $\beta$  и только одну возможность для выбора последней позиции  $\nu$  (к этому моменту, если мы производим разложение, аналогичное (7), будут использованы все столбцы, кроме одного), и поэтому разложение будет содержать  $n!$  членов. Другими словами, число перестановок набора чисел  $1, 2, \dots, n$  равняется  $n!$ . Каждой последовательности номеров столбцов матрицы соответствует определенная перестановка; шесть членов в (7) получены из следующих наборов столбцов:

$$(\alpha, \beta, \nu) = (1, 2, 3), (2, 3, 1), (3, 1, 2), (1, 3, 2), (2, 1, 3), (3, 2, 1).$$

Это все  $3! = 6$  перестановок набора  $(1, 2, 3)$ , причем первая перестановка тождественная.

Итак, определитель матрицы  $A$  заменен суммой шести значительно более простых определителей. Вынося  $a_{ij}$ , получаем по одному члену суммы для каждой из перестановок:

$$\begin{aligned} \det A = & a_{11}a_{22}a_{33} \begin{vmatrix} 1 & & \\ & 1 & \\ & & 1 \end{vmatrix} + a_{12}a_{23}a_{31} \begin{vmatrix} & 1 & \\ & & 1 \\ 1 & & \end{vmatrix} + a_{13}a_{21}a_{32} \begin{vmatrix} & & 1 \\ & 1 & \\ 1 & & \end{vmatrix} + \\ & + a_{11}a_{23}a_{32} \begin{vmatrix} 1 & & \\ & & 1 \\ & 1 & \end{vmatrix} + a_{12}a_{21}a_{33} \begin{vmatrix} & 1 & \\ & & 1 \\ 1 & & \end{vmatrix} + a_{13}a_{22}a_{31} \begin{vmatrix} & & 1 \\ & 1 & \\ 1 & & \end{vmatrix}. \quad (7') \end{aligned}$$

Каждый из членов суммы равен произведению  $n=3$  элементов  $a_{ij}$ , причем каждые строка и столбец представлены по одному разу. Другими словами, в разложении имеется член, соответствующий каждому пути, проходящему через всю матрицу и использующему каждый столбец по одному разу; если столбцы используются в порядке  $(\alpha, \dots, \nu)$ , то этот член равен произведению  $a_{1\alpha} \dots a_{n\nu}$ , умноженному на определитель матрицы перестановки  $P_\sigma$ . Определитель всей матрицы равен сумме таких членов; **эта сумма и дает ту явную формулу, которую мы искали:**

$$\| \det A = \sum_{\sigma} (a_{1\alpha} a_{2\beta} \dots a_{n\nu}) \det P_{\sigma}. \quad (8)$$

Для матрицы порядка  $n$  сумма берется по всем  $n!$  перестановкам  $\sigma = (\alpha, \dots, \nu)$  набора чисел  $(1, \dots, n)$ . Перестановка дает последовательность номеров столбцов, когда мы движемся вниз по строкам матрицы, и она также определяет матрицу перестановки  $P_\sigma$ : единицы находятся в ней на тех же местах, где находились элементы  $a$  в  $A$ .

Остается вычислить определитель матрицы  $P_\sigma$ . Поскольку перестановками строк  $P_\sigma$  преобразуется в единичную матрицу и каждая из перестановок изменяет знак определителя,  $\det P_\sigma$  равен

либо  $+1$ , либо  $-1$ . Знак зависит от того, является число перестановок строк четным или нечетным. Рассмотрим два примера:

$$P_{\sigma} = \begin{bmatrix} 1 & & \\ & & 1 \\ & 1 & \end{bmatrix}$$

с последовательностью столбцов  $(\alpha, \beta, \nu) = (1, 3, 2)$  и

$$P_{\sigma} = \begin{bmatrix} & & 1 \\ 1 & & \\ & 1 & \end{bmatrix}$$

с последовательностью столбцов  $\sigma = (3, 1, 2)$ .

В первом случае требуется одна перестановка строк (второй и третьей), так что  $\det P_{\sigma} = -1$ . Во втором для получения единичной матрицы нужны две перестановки строк (первой и второй, а затем второй и третьей), так что  $\det P_{\sigma} = (-1)^2 = 1$ . Это два из шести знаков  $\pm$ , имеющихя в формуле (5).

Формула (8) представляет собой явное выражение для определителя, и нетрудно проверить формулу для определителя 2-го порядка:  $2! = 2$  перестановки есть  $\sigma = (1, 2)$  и  $\sigma = (2, 1)$ , и поэтому

$$\begin{aligned} \det A &= a_{11}a_{22} \det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + a_{12}a_{21} \det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \\ &= a_{11}a_{22} - a_{12}a_{21} = ad - bc. \end{aligned}$$

Нельзя сказать, что явная формула (8) отличается особой простотой. Тем не менее можно показать, что она обладает свойствами 1—3. Свойство 3, утверждающее, что  $\det I = 1$ , является наиболее простым: все произведения  $a_{ij}$  будут равны нулю, кроме произведений, соответствующих специальной последовательности столбцов  $\sigma = (1, 2, \dots, n)$ , т. е. тождественной перестановке. Этот член дает  $\det I = 1$ . Свойство 2 будет проверено в следующем параграфе, поскольку сейчас нас больше интересует свойство 1: определитель должен быть линейной функцией первой строки  $a_{11}, a_{12}, \dots, a_{1n}$ . Для того чтобы убедиться в этом, рассмотрим члены формулы (8), содержащие элемент  $a_{11}$ . Такие члены появляются, когда для первой позиции мы берем  $\alpha = 1$ , в результате чего остается некоторая перестановка  $\sigma' = (\beta, \dots, \nu)$  из оставшихся номеров столбцов  $(2, \dots, n)$ . Запишем все эти члены в виде  $a_{11}A_{11}$ , где коэффициент при  $a_{11}$  равен

$$A_{11} = \sum_{\sigma'} (a_{2\beta} \dots a_{n\nu}) \det P_{\sigma'}. \quad (9)$$

Подобным же образом элемент  $a_{i2}$  умножается на некоторое выражение  $A_{i2}$ . Складываем все члены с  $a_{1j}$ , и формула (8) прини-

мает вид

$$\det A = a_{11}A_{11} + a_{12}A_{12} + \dots + a_{1n}A_{1n}. \quad (10)$$

Ясно, что  $\det A$  линейно зависит от элементов  $a_{11}, \dots, a_{1n}$  первой строки. Поэтому свойство 1 выполняется безотносительно к виду коэффициентов  $A_{1j}$ .

**Пример.** Для матрицы 3-го порядка этот метод дает

$$\det A = a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{13}(a_{21}a_{32} - a_{23}a_{31}).$$

«Алгебраические дополнения»  $A_{11}, A_{12}, A_{13}$  записаны в скобках.

#### РАЗЛОЖЕНИЕ $\det A$ НА АЛГЕБРАИЧЕСКИЕ ДОПОЛНЕНИЯ

Нам нужна еще одна формула для определителя. При этом нет необходимости вновь повторять все рассуждения, *поскольку фактически она уже нами получена*—это формула (10); *нужно лишь определить алгебраические дополнения  $A_{1j}$ .*

Мы знаем, что это число  $A_{1j}$  зависит от строк  $2, \dots, n$ ; 1-я строка уже учтена при помощи множителя  $a_{1j}$ . Более того, множитель  $a_{1j}$  учитывает также и вклад  $j$ -го столбца, так что алгебраическое дополнение  $A_{1j}$  должно зависеть только от оставшихся столбцов. Ни одна строка или столбец не могут присутствовать дважды в одном и том же члене, и фактически мы занимаемся разложением определителя на следующую сумму:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & & \\ & a_{22} & a_{23} \\ & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ a_{21} & & \\ a_{31} & & \end{vmatrix} + \begin{vmatrix} & & a_{13} \\ a_{21} & a_{22} & \\ a_{31} & a_{32} & \end{vmatrix}.$$

Это разложение сводит определитель порядка  $n$  к нескольким определителям (называемым *минорами*) порядка  $n-1$ ; в данном случае это определители подматриц 2-го порядка в правой части разложения. Подматрица  $M_{1j}$  получается выбрасыванием 1-й строки и  $j$ -го столбца, и каждый из членов в правой части равен произведению элемента  $a_{1j}$  на определитель подматрицы  $M_{1j}$ , взятый с нужным знаком. Эти знаки изменяются при переходе от одного столбца к следующему, и в результате алгебраические дополнения определяются как

$$A_{1j} = (-1)^{1+j} \det M_{1j}. \quad (11)$$

Например второе алгебраическое дополнение  $A_{12}$  равно  $a_{23}a_{31} - a_{21}a_{33}$ , что совпадает с величиной  $(-1)^{1+2} \det M_{12}$ . Эти рассуждения остаются в силе для квадратных матриц любого порядка и внимательное рассмотрение формулы (9) подтверждает, что  $A_{11}$  есть определитель подматрицы  $M_{11}$ , расположенной в правом нижнем углу исходной матрицы  $A$ .

Аналогичное разложение справедливо для любой другой строки, например  $i$ -й, что можно доказать перестановкой  $i$ -й и 1-й строк:

**4В.** *Определитель матрицы  $A$  может быть вычислен разложением по алгебраическим дополнениям  $i$ -й строки:*

$$\det A = a_{i1}A_{i1} + a_{i2}A_{i2} + \dots + a_{in}A_{in}. \quad (12)$$

**Алгебраическое дополнение  $A_{ij}$**  есть определитель подматрицы  $M_{ij}$ , взятый с нужным знаком:

$$A_{ij} = (-1)^{i+j} \det M_{ij}.$$

Подматрица  $M_{ij}$  образуется вычеркиванием  $i$ -й строки и  $j$ -го столбца матрицы  $A$ .

Эти формулы выражают  $\det A$  в виде линейной комбинации определителей порядка  $n-1$ . Поэтому можно ввести определитель индукцией по  $n$ . Для матрицы порядка 1 полагаем  $\det A = a_{11}$  и затем используем формулу (12) последовательного введения определителей матриц порядка 2, 3 и так далее. Мы предпочли ввести определитель через его свойства, которые значительно легче объяснить, и затем вывести явную формулу (8) и индуктивную формулу (12) из этих свойств.

Наконец, имеется еще одно следствие из того факта, что  $\det A = \det A^T$ . Это свойство позволяет нам производить разложение по алгебраическим дополнениям столбца, а не строки; для каждого  $j$

$$\det A = \sum_{i=1}^n a_{ij}A_{ij}. \quad (13)$$

Доказательство состоит в разложении  $\det A^T$  по алгебраическим дополнениям  $j$ -й строки, которая является  $j$ -м столбцом матрицы  $A$ .

**Пример 1.** Разложения по алгебраическим дополнениям особенно полезны в тех случаях, когда некоторые строки матрицы почти целиком состоят из нулей, как, например, для треугольной матрицы

$$A = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 1 & 2 & -1 & 0 \\ 6 & 6 & 7 & 2 \end{bmatrix}.$$

Единственный член, получающийся при разложении по первой строке, есть  $a_{11}A_{11}$ , поскольку  $a_{12} = a_{13} = a_{14} = 0$ :

$$a_{11} = 4, \quad A_{11} = \det \begin{bmatrix} 3 & 0 & 0 \\ 2 & -1 & 0 \\ 6 & 7 & 2 \end{bmatrix}.$$

Раскладывая  $A_{11}$  по первой строке, приходим к соотношению

$$\det A = 4A_{11} = 4 \cdot 3 \cdot \det \begin{bmatrix} -1 & 0 \\ 7 & 2 \end{bmatrix}.$$

Проводя редукцию еще раз, получаем

$$\det A = 4 \cdot 3 \cdot (-1) \det [2] = -24.$$

Тот же самый результат получится при разложении по алгебраическим дополнениям последнего столбца — и этим подтверждается тот факт, что определитель любой треугольной матрицы равен произведению диагональных элементов.

**Пример 2.** Вновь рассмотрим матрицу конечных разностей

$$A = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & \ddots & \\ & & \ddots & \ddots & -1 \\ & & & -1 & 2 \end{bmatrix}.$$

Пусть  $\det A$  вычисляется разложением по первой строке. Единственными ненулевыми элементами в ней являются  $a_{11} = 2$  и  $a_{12} = -1$ , так что

$$\det A = a_{11}A_{11} + a_{12}A_{12} + \dots + a_{1n}A_{1n} = 2A_{11} - A_{12}.$$

$A_{11}$  вычисляется путем вычеркивания первых строки и столбца, что дает подматрицу в точности такого же вида, как и сама  $A$ , лишь порядок ее равен  $n-1$ . Другое алгебраическое дополнение получается в результате вычеркивания 1-й строки и 2-го столбца:

$$A_{12} = (-1)^{1+2} \det \begin{bmatrix} -1 & -1 & & \\ & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix}.$$

Теперь производим разложение по первому столбцу, в котором содержится лишь один ненулевой элемент:

$$A_{12} = (-1)^{1+2} (-1) \det \begin{bmatrix} 2 & \ddots & \\ \ddots & \ddots & -1 \\ & -1 & 2 \end{bmatrix}.$$

Мы вновь получили разностную матрицу того же вида, но уже порядка  $n-2$ . Если  $D_n$  — это определитель матрицы  $A$ , содержащей  $n$  строк и столбцов, то нами получено рекуррентное соотношение

$$D_n = 2A_{11} - A_{12} = 2D_{n-1} - D_{n-2}.$$

Для матриц малой размерности, при  $n=1$  и  $n=2$ , легко вычислить

$$D_1 = |2| = 2, \quad D_2 = \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3.$$

Рекуррентное соотношение дает последовательно  $D_3 = 2D_2 - D_1 = 4$ ,  $D_4 = 5$  и в общем случае  $D_n = n + 1$ :

$$D_n = 2D_{n-1} - D_{n-2} \quad \text{дает} \quad n + 1 = 2(n) - (n - 1).$$

Это совпадает с произведением ведущих элементов, вычисленным в начале этого параграфа, и определитель матрицы  $A$  равен  $n + 1$ .

**Упражнение 4.3.1.** Для матрицы

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

найти единственный ненулевой член формулы (8), соответствующий единственному способу выбора четырех ненулевых элементов, находящихся в разных строках и столбцах. Вычислить  $\det A$ , определив, является эта перестановка четной или нечетной.

**Упражнение 4.3.2.** Произвести разложение указанной выше матрицы  $A$  по алгебраическим дополнениям первой строки, сведя  $\det A$  к определителю 3-го порядка. Прodelать ту же операцию с этим определителем (следя за знаком  $(-1)^{i+j}$ ) и с получающимся из него определителем 2-го порядка. Вычислить  $\det A$ .

**Упражнение 4.3.3.** Найти матрицу 4-го порядка, для которой

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} \neq \det A \det D - \det B \det C.$$

Здесь  $A$ ,  $B$ ,  $C$  и  $D$  — подматрицы 2-го порядка и равенство будет иметь место в случае, когда  $B$  или  $C$  есть нулевая матрица.

**Упражнение 4.3.4.** Вычислить определитель матрицы

$$A_4 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix},$$

используя либо преобразования строк с целью получения нулей, либо разложение по алгебраическим дополнениям первой строки, либо другими методами. Найти также определители меньших матриц  $A_3$  и  $A_2$  того же вида, т. е. с нулями вдоль главной диагонали и единицами на остальных местах. Можете ли вы предсказать, чему равен  $\det A_n$ ?

**Упражнение 4.3.5.** Найти определитель и все девять алгебраических дополнений матрицы

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}.$$

Построить новую матрицу  $B$ , элементы с индексами  $i, j$  которой равны алгебраическим дополнениям  $A_{ji}$ . Проверить, что  $AB$  есть единичная матрица, умноженная на определитель  $\det A$ . Чему равна матрица  $A^{-1}$ ?

#### § 4.4. ПРИМЕНЕНИЯ ОПРЕДЕЛИТЕЛЕЙ

В этом параграфе мы сделаем попытку систематически описать каждое из применений определителей, упомянутых во введении к этой главе.

1. *Вычисление  $A^{-1}$ .* С этой целью воспользуемся разложением определителя по алгебраическим дополнениям  $i$ -й строки:

$$\det A = \sum_{j=1}^n a_{ij} A_{ij}. \quad (14)$$

Составим линейную комбинацию элементов одной строки матрицы с алгебраическими дополнениями другой и покажем, что результат равен нулю:

$$0 = \sum_{j=1}^n a_{kj} A_{ij} \quad \text{для} \quad i \neq k. \quad (15)$$

Это получается потому, что фактически мы вычисляем определитель другой матрицы  $B$ , полученной в результате выбрасывания  $i$ -й строки матрицы  $A$  и замены ее на  $k$ -ю. Таким образом,  $k$ -я строка присутствует в матрице  $B$  дважды и  $\det B = 0$ . Тогда (15) есть обычное разложение матрицы  $B$  по  $i$ -й строке; алгебраические дополнения  $A_{ij}$  не изменялись, поскольку они не зависят от  $i$ -й строки, но умножаются на элементы  $a_{kj}$   $i$ -й строки матрицы  $B$ .

Из (14), (15) получается замечательная формула умножения:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \vdots & \vdots & \dots & \vdots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix} = \begin{bmatrix} \det A & 0 & \dots & 0 \\ 0 & \det A & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \det A \end{bmatrix} \quad (16)$$

**Пример.** Алгебраические дополнения элементов матрицы  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$  суть  $A_{11} = d$ ,  $A_{12} = -c$ ,  $A_{21} = -b$ ,  $A_{22} = a$ . Они образуют «присоединенную матрицу», соответствующую средней матрице формулы (16) и обозначаемую через  $\text{adj } A$ : для матриц 2-го порядка

$$\text{adj } A = \begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{bmatrix} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Тогда формула (16) принимает вид

$$A \text{adj } A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} ad - bc & 0 \\ 0 & ad - bc \end{bmatrix} = (\det A) I.$$

Деля на число  $\det A$  (если оно не равно нулю), получаем искомую обратную матрицу:

$$A \frac{\text{adj } A}{\det A} = I, \quad \text{или} \quad A^{-1} = \frac{\text{adj } A}{\det A}. \quad (17)$$

Каждый элемент обратной матрицы равен *соответствующему алгебраическому дополнению матрицы  $A$ , деленному на  $\det A$* . Это подтверждает тот факт, что матрица  $A$  обратима тогда и только тогда, когда  $\det A \neq 0$ , и дает точную формулу для  $A^{-1}$ .

2. *Решение системы  $Ax = b$ .* Здесь применение определителей сводится к формуле

$$x = A^{-1}b = \frac{(\text{adj } A) b}{\det A},$$

которая представляет собой просто произведение матрицы и вектора, деленное на число  $\det A$ . Имеется, однако, другая форма записи этого соотношения:

**4С. Правило Крамера:**  $j$ -й элемент вектора  $x = A^{-1}b$  равен

$$x_j = \frac{\det B_j}{\det A}, \quad \text{где} \quad B_j = \begin{bmatrix} a_{11} & a_{12} & \dots & b_1 & \dots & a_{1n} \\ \vdots & \vdots & & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & b_n & \dots & a_{nn} \end{bmatrix}. \quad (18)$$

В  $B_j$  вектор  $b$  заменяет собой  $j$ -й столбец матрицы  $A$ .

**Доказательство.** Разлагая  $\det B_j$  по алгебраическим дополнениям  $j$ -го столбца (который равен  $b$ ), из формулы (13) получаем

$$\det B_j = b_1 A_{1j} + b_2 A_{2j} + \dots + b_n A_{nj}.$$

Это есть в точности  $j$ -й элемент произведения  $(\text{adj } A)b$ . Деля на  $\det A$ , получаем  $j$ -й элемент вектора  $x$ .

Таким образом, каждая из компонент вектора  $x$  равна отношению двух определителей, т. е. частному от деления двух многочленов степени  $n$ . Этот факт можно было бы установить и из метода исключения Гаусса, но это не было сделано.

**Пример.** Решение системы

$$\begin{aligned}x_1 + 3x_2 &= 0, \\ 2x_1 + 4x_2 &= 6\end{aligned}$$

есть

$$x_1 = \frac{\begin{vmatrix} 0 & 3 \\ 6 & 4 \\ 1 & 3 \\ 2 & 4 \end{vmatrix}}{\begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix}} = \frac{-18}{-2} = 9, \quad x_2 = \frac{\begin{vmatrix} 1 & 0 \\ 2 & 6 \\ 1 & 3 \\ 2 & 4 \end{vmatrix}}{\begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix}} = \frac{6}{-2} = -3.$$

**3. Объем параллелепипеда.** Связь между определителем и объемом не очевидна, однако мы можем предположить для начала, что все углы прямые, т. е. грани взаимно перпендикулярны, и мы имеем дело с прямоугольным параллелепипедом. Тогда объем его равен просто произведению длин ребер  $l_1 l_2 \dots l_n$ .

Мы хотим получить ту же самую формулу с помощью определителя. С этой целью вспомним, что ребра параллелепипеда представляются строками матрицы  $A$ . В нашем случае эти строки взаимно ортогональны, так что

$$AA^T = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} a_{11} & a_{n1} \\ \vdots & \vdots \\ a_{1n} & a_{nn} \end{bmatrix} = \begin{bmatrix} l_1^2 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & l_n^2 \end{bmatrix}.$$

Величины  $l_i^2$  суть квадраты длин строк матрицы, т. е. квадраты длин ребер, и нули вне диагонали получаются вследствие ортогональности строк. Переходя к определителям и используя правила 9 и 10, получаем

$$l_1^2 l_2^2 \dots l_n^2 = \det(AA^T) = (\det A)(\det A^T) = (\det A)^2.$$

Извлекая корень, мы и приходим к требуемому соотношению: *определитель равняется объему. Знак при  $\det A$  будет зависеть от того, образуют ребра правостороннюю систему координат вида  $xuz$  или левостороннюю  $ухz$ .*

Если область не прямоугольна, то объем уже не равен произведению длин ребер. В плоском случае (рис. 4.2) «объем» парал-

делограмма равен произведению длины основания на высоту  $h$ . Вектор  $pb$  длины  $h$  есть разность между вектором второй строки  $Ob = (a_{21}, a_{22})$  и его проекцией  $Op$  на вектор первой строки. Основ-ной момент доказательства заключается в том, что по свойству 5

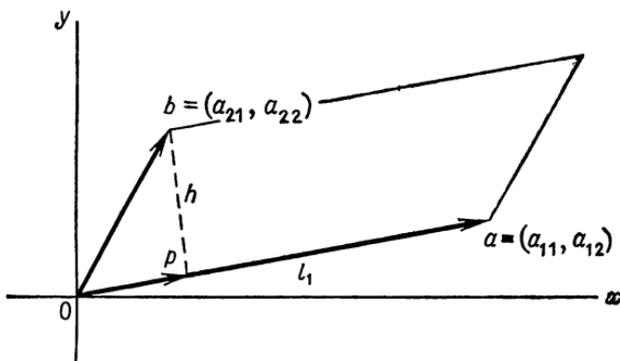


Рис. 4.2. Площадь параллелограмма равна  $\det A$ .

$\det A$  не изменится, если первую строку матрицы умножить на постоянный множитель и вычесть из второй строки. В то же время объем не изменится, если мы перейдем к прямоугольнику с основанием длины  $l_1$  и высотой  $h$ . Таким образом, мы приходим к случаю, для которого уже доказано, что объем равен определителю.

Для доказательства в общем  $n$ -мерном случае используется та же самая идея, усложняются лишь выкладки. Ни объем, ни определитель не будут изменяться, если мы последовательно для 2-й, 3-й, ...,  $n$ -й строк будем вычитать из каждой строки ее проекцию на пространство, порожденное векторами предыдущих строк, что дает в результате «вектор высоты» типа  $pb$ , перпендикулярный к основанию. Результатом этого процесса Грама — Шмидта является набор взаимно ортогональных строк, но определитель и объем при этом не изменяются. Поскольку в прямоугольном случае объем равен определителю, то то же самое должно быть и для исходной матрицы.

Доказательство на этом завершается, однако будет полезно еще раз вернуться к простейшему случаю. Мы знаем, что

$$\det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 1, \quad \det \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix} = 1.$$

Эти определители дают объемы, или площади, поскольку мы рассматриваем двумерный случай, «параллелепипедов», изображенных на рис. 4.3. Первый представляет собой единичный квадрат, и его площадь, разумеется, равна 1. Второй есть паралле-

лограмм с единичными основанием и высотой; его площадь не зависит от «сдвига», даваемого коэффициентом  $c$ , и равна 1.

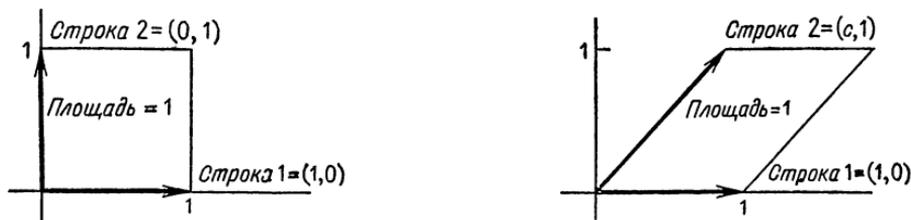


Рис. 4.3. Площади квадрата и параллелограмма.

**4. Формула для ведущих элементов.** Последнее применение (в этом весьма утилитарном обзоре определителей) относится к вопросу о нулевых ведущих элементах; мы можем, наконец, установить, когда исключение Гаусса возможно без перестановки строк. Основной момент состоит в том, что первые  $k$  ведущих элементов полностью определяются подматрицей  $A_k$ , расположенной в левом верхнем углу матрицы  $A$ . *Оставшиеся строки и столбцы матрицы  $A$  не оказывают на них никакого влияния.*

**Пример.**

$$A = \begin{bmatrix} a & b & e \\ c & d & f \\ g & h & i \end{bmatrix} \rightarrow \begin{bmatrix} a & b & e \\ 0 & (ad-bc)/a & (af-ec)/a \\ g & h & i \end{bmatrix}.$$

Очевидно, что первый ведущий элемент зависит только от первых строки и столбца; это  $d_1 = a$ . Второй ведущий элемент получается после одного шага исключения; он равен  $d_2 = (ad-bc)/a$  и зависит только от элементов  $a, b, c$  и  $d$ . Оставшаяся часть матрицы  $A$  не участвует в рассмотрении до исследования третьего ведущего элемента. Фактически левый верхний угол матрицы  $A$  определяет не только ведущие элементы, но и все левые верхние углы матриц  $L, D$  и  $U$ :

$$A = LDU = \begin{bmatrix} 1 & & \\ c/a & 1 & \\ * & * & 1 \end{bmatrix} \begin{bmatrix} a & & \\ & (ad-bc)/a & \\ & & * \end{bmatrix} \begin{bmatrix} 1 & b/a & * \\ & 1 & * \\ & & 1 \end{bmatrix}.$$

То, что мы видим в первых двух строках и столбцах, есть в точности разложение (выписанное ранее в (4)) подматрицы  $A_2 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ . Это общее правило:

4D. Если матрица  $A$  представляется в виде  $LDU$ , то левые верхние углы удовлетворяют соотношению

$$A_k = L_k D_k U_k. \quad (19)$$

Для разных  $k$  разложения подматриц  $A_k$  «согласованы» друг с другом.

Для доказательства можно либо показать, что этот угол можно преобразовать первым, не затрагивая остальные строки матрицы, либо воспользоваться правилами блочного умножения матриц. Эти правила точно такие же, как и обычное поэлементное правило:  $LDU = A$  дает

$$\begin{aligned} \begin{bmatrix} L_k & 0 \\ B & C \end{bmatrix} \begin{bmatrix} D_k & 0 \\ 0 & E \end{bmatrix} \begin{bmatrix} U_k & F \\ 0 & G \end{bmatrix} &= \begin{bmatrix} L_k D_k & 0 \\ B D_k & C E \end{bmatrix} \begin{bmatrix} U_k & F \\ 0 & G \end{bmatrix} = \\ &= \begin{bmatrix} L_k D_k U_k & L_k D_k F \\ B D_k U_k & B D_k F + C E G \end{bmatrix}. \end{aligned}$$

Если матрицы разбиты на блоки таким образом, чтобы их можно было перемножить, т. е. квадратные или прямоугольные матрицы имеют нужные для перемножения размеры, их можно умножать по блокам<sup>1)</sup>. Сравнивая последнюю матрицу с  $A$ , легко видеть, что угол  $L_k D_k U_k$  совпадает с  $A_k$ , что и доказывает формулу (19).

Из этого тотчас же вытекают формулы для ведущих элементов. Переходя к определителям в (19), получаем

$$\det A_k = \det L_k \det D_k \det U_k = \det D_k = d_1 d_2 \dots d_k. \quad (20)$$

Произведение первых  $k$  ведущих элементов равняется определителю  $A_k$ ; это то же самое правило для  $A_k$ , которое мы уже знаем для всей матрицы  $A = A_n$ . Поскольку определитель матрицы  $A_{k-1}$  выражается аналогичным образом через  $d_1 d_2 \dots d_{k-1}$ , мы можем выразить ведущий элемент  $d_k$  как **отношение определителей**:

$$\frac{\det A_k}{\det A_{k-1}} = \frac{d_1 d_2 \dots d_k}{d_1 d_2 \dots d_{k-1}} = d_k. \quad (21)$$

В приведенном выше примере второй ведущий элемент равнялся в точности этому отношению  $(ad-bc)/a$ , т. е. определителю матрицы  $A_2$ , деленному на определитель матрицы  $A_1$ . (Мы полагаем  $\det A_0 = 1$ , так что первый ведущий элемент равен  $a/1 = a$ .) Перемножая все отдельные ведущие элементы, получаем

$$d_1 d_2 \dots d_n = \frac{\det A_1}{\det A_0} \frac{\det A_2}{\det A_1} \dots \frac{\det A_n}{\det A_{n-1}} = \frac{\det A_n}{\det A_0} = \det A.$$

<sup>1)</sup> Это очень полезное правило, хотя в этой книге оно будет использоваться нескоро.

Формула (21) дает нам, наконец, ответ на исходный вопрос: *ведущие элементы отличны от нуля тогда и только тогда, когда все числа  $\det A_k$  отличны от нуля:*

|| 4E. Метод исключения Гаусса без перестановок строк может быть применен к матрице  $A$  тогда и только тогда, когда все главные подматрицы  $A_1, A_2, \dots, A_n$  невырождены.

На этом рассмотрение определителей заканчивается, остается лишь сделать дополнительное замечание, как было обещано в начале главы. Это замечание касается непротиворечивости определяющих свойств 1—3. Основным является свойство 2 об изменении знака при перестановке строк, что приводит к правилу для определителя матрицы перестановки  $P_\sigma$ . Это было единственным сомнительным моментом в явной формуле (8): верно ли, что независимо от выбранной последовательности перестановок строк, приводящей  $P_\sigma$  к единичной матрице, их число является либо всегда четным, либо всегда нечетным? Если это так, то мы вправе называть перестановку «четной» или «нечетной» и соответствующий определитель по правилу 2 равен либо  $+1$ , либо  $-1$ .

Начиная с перестановки  $(3, 2, 1)$ , за один шаг меняя местами 3 и 1, мы получаем естественный порядок  $(1, 2, 3)$ . Тот же самый результат можно получить, меняя местами 3 и 2, затем 3 и 1 и, наконец, 2 и 1. В обеих последовательностях число перемен нечетное. Утверждается, что *четное число перемен никогда не приведет перестановку  $(3, 2, 1)$  к естественному порядку.*

Приведем доказательство. Рассмотрим каждую пару чисел в перестановке и обозначим через  $N$  число пар, в которых большее число стоит первым. Ясно, что  $N=0$  для естественного порядка  $(1, 2, 3)$ , т. е. для тождественной перестановки, и  $N=3$  для перестановки  $(3, 2, 1)$ , поскольку пары  $(3, 2)$ ,  $(3, 1)$  и  $(2, 1)$  неправильные. Теперь остается показать, что перестановка является четной или нечетной в зависимости от того, четным или нечетным является число  $N$ . Другими словами, если мы начинаем с любой перестановки, то каждая перемена местами двух чисел будет изменять  $N$  на нечетное число. Тогда для получения  $N=0$  (естественного порядка) требуется число перемен такой же четности (четное или нечетное), как и исходное  $N$ .

Если переставляются соседние числа, то ясно, что  $N$  изменится либо на  $+1$ , либо  $-1$ ; оба числа нечетные. Поэтому, если заметить, что *перемена местами любых двух чисел может быть осуществлена нечетным числом перемен местами соседних чисел*, то доказательство будет завершено. Это наблюдение легко подтверждается примером; чтобы поменять местами первый и

четвертый элемент в перестановке (это числа 2 и 3), пять раз (нечетное число) меняются местами соседние числа:

$$(2, 1, 4, 3) \rightarrow (1, 2, 4, 3) \rightarrow (1, 4, 2, 3) \rightarrow \\ \rightarrow (1, 4, 3, 2) \rightarrow (1, 3, 4, 2) \rightarrow (3, 1, 4, 2).$$

В общем случае, чтобы поменять местами элементы, находящиеся на местах  $k$  и  $l$ , требуется  $l-k$  раз поменять местами соседние элементы для перемещения элемента с места  $k$  на место  $l$ , и  $l-k-1$  — для перемещения элемента, первоначально находившегося на месте  $l$  (и который теперь находится на месте  $l-1$ ), на место  $k$ . Поскольку число  $(l-k) + (l-k-1)$  нечетное, доказательство закончено. Определитель не только обладает всеми свойствами, указанными ранее, он даже существует.

**Упражнение 4.4.1.** Используя (17), обратить матрицы

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

**Упражнение 4.4.2.** Найти  $x$ ,  $y$  и  $z$  по правилу Крамера:

$$\begin{aligned} x + 4y - z &= 1, \\ x + y + z &= 0, \\ 2x + 3z &= 0. \end{aligned}$$

**Упражнение 4.4.3.** Найти якобиан  $J$  преобразования прямоугольных координат  $x$ ,  $y$ ,  $z$  в сферические координаты  $r$ ,  $\theta$ ,  $\varphi$ :  $x = r \cos \theta \cos \varphi$ ,  $y = r \sin \theta \cos \varphi$ ,  $z = r \sin \varphi$ .

**Упражнение 4.4.4.** (а) Начертить треугольник с вершинами  $A = (2, 2)$ ,  $B = (-1, 3)$ ,  $C = (0, 0)$ . Рассматривая его как половину параллелограмма, объяснить, почему его площадь равняется  $\frac{1}{2} \det \begin{bmatrix} 2 & 2 \\ -1 & 3 \end{bmatrix}$ .

(б) Предположим, что третья вершина вместо начала координат находится в точке  $C = (1, -4)$ . Вычислить площадь и затем проверить формулу

$$\text{площадь } (ABC) = \frac{1}{2} \det \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix} = \frac{1}{2} \det \begin{bmatrix} 2 & 2 & 1 \\ -1 & 3 & 1 \\ 1 & -4 & 1 \end{bmatrix}.$$

Указание: вычитая последнюю строку из предыдущих, получаем

$$\det \begin{bmatrix} 2 & 2 & 1 \\ -1 & 3 & 1 \\ 1 & -4 & 1 \end{bmatrix} = \det \begin{bmatrix} 1 & 6 & 0 \\ -2 & 7 & 0 \\ 1 & -4 & 1 \end{bmatrix} = \det \begin{bmatrix} 1 & 6 \\ -2 & 7 \end{bmatrix}.$$

Как новые вершины  $A' = (1, 6)$ ,  $B' = (-2, 7)$  и  $C' = (0, 0)$  связаны с заданными  $A$ ,  $B$  и  $C$ ? Сделать рисунок.

**Упражнение 4.4.5.** Объяснить в терминах объемов, почему  $\det 3A = 3^n \det A$  для любой матрицы  $A$  порядка  $n$ .

**Упражнение 4.4.6.** Предугадать, чему равны ведущие элементы матрицы

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 5 & 0 \\ 0 & 6 & -4 \end{bmatrix},$$

и найти их методом исключения. Имеется ли перестановка строк, которая позволит избежать появления нулевых ведущих элементов?

**Упражнение 4.4.7.** Перестановка  $(n, n-1, \dots, 1)$  чисел  $(1, 2, \dots, n)$  четна или нечетна?

**Упражнение 4.4.8.** Найти последовательность перемен чисел местами, приводящих перестановку  $\sigma = (5, 3, 1, 2, 4)$  к  $(1, 2, 3, 4, 5)$ . Является ли  $\sigma$  четной или нечетной?

**Упражнение 4.4.9.** Определить, являются векторы  $(1, 3, 2)$ ,  $(2, 4, 4)$  и  $(1, 5, 2)$  линейно независимыми или нет, с помощью образованного ими определителя.

## ОБЗОРНЫЕ УПРАЖНЕНИЯ

**4.1.** Где нужно разместить нули в матрице размера  $4 \times 4$ , чтобы при минимальном их числе гарантировать равенство нулю ее определителя?

**4.2.** Где нужно разместить нули и единицы в матрице размера  $4 \times 4$ , чтобы при минимальном их числе гарантировать равенство единице ее определителя?

**4.3.** Найти определитель матрицы

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 \\ 1 & 1 & 3 & 1 \\ 1 & 4 & 1 & 1 \end{bmatrix}.$$

**4.4.** Если  $B = M^{-1}AM$ , то почему  $\det B = \det A$ ?

**4.5.** Дать контрпример к равенству  $\det(A+B) = \det A + \det B$ .

**4.6.** Умножить первую строку матрицы  $A$  на 3, обозначив новую матрицу через  $B$ , а затем вычесть первую строку матрицы  $B$  из второй строки, обозначив полученную матрицу через  $C$ . Как связаны  $\det C$  и  $\det A$ ?

**4.7.** Решить систему  $3u + 2v = 7$ ,  $4u + 3v = 11$  по правилу Крамера.

**4.8.** Если элементы матрицы  $A$  — целые числа и  $\det A$  равен 1 или  $-1$ , то как показать, что элементы матрицы  $A^{-1}$  — тоже целые числа?

**4.9.** Если все элементы матриц  $A$  и  $A^{-1}$  — целые числа, то как показать, что оба определителя равны 1 или  $-1$ ? Указание: чему равно произведение  $\det A$  на  $\det A^{-1}$ ?

**4.10.** Найти алгебраические дополнения и обратную матрицу для

$$\begin{bmatrix} 3 & 5 \\ 6 & 9 \end{bmatrix}.$$

4.11. Чему равен объем параллелепипеда с четырьмя вершинами в точках  $(0, 0, 0)$ ,  $(-1, 2, 2)$ ,  $(2, -1, 2)$  и  $(2, 2, -1)$ ? Где находятся остальные четыре вершины?

4.12. Сколько членов содержится в формуле для определителя матрицы 5-го порядка и сколько из них равно нулю, если  $a_{21} = 0$ ?

4.13. Если сумма элементов в каждой из строк матрицы  $A$  равняется нулю и вектор-столбец  $x$  состоит из единиц, то чему равен  $Ax$ ? Как показать, что  $\det A = 0$ ?

4.14. Почему имеется четное число перестановок набора  $(1, 2, \dots, 9)$  и почему ровно половину из них составляют нечетные перестановки?

## Глава 5

# СОБСТВЕННЫЕ ЗНАЧЕНИЯ И СОБСТВЕННЫЕ ВЕКТОРЫ

### § 5.1. ВВЕДЕНИЕ

Этой главой начинается «вторая часть» теории матриц. Первая часть почти целиком относилась к линейным системам, которые решались методом исключения. В дальнейшем метод исключения будет играть лишь вспомогательную роль. Новые задачи по-прежнему будут решаться путем упрощения вида матрицы, приведения ее к диагональной или верхней треугольной, но *основным методом уже не будет метод вычитания из одной строки матрицы другой строки, умноженной на постоянную*. Теперь мы заинтересованы в сохранении собственных значений матрицы, а не ее пространства строк, а элементарные преобразования строк для этой цели непригодны.

Глава об определителях служила для перехода от задачи  $Ax=b$  к задаче на собственные значения. В обоих случаях использование определителей приводит к «формальному решению»: к правилу Крамера для задачи  $Ax=b$  и к многочлену  $\det(A-\lambda I)$ , корни которого являются собственными значениями матрицы  $A$ . (Отметим, что все матрицы теперь будут предполагаться квадратными, поскольку говорить о собственных значениях прямоугольной матрицы так же бессмысленно, как и о ее определителе.) Как и в первом случае, определители можно практически использовать для решения задачи на собственные значения, если  $n=2$  или 3; для больших  $n$  вычисление собственных значений представляет собой значительно более трудную задачу, чем решение уравнения  $Ax=b$ , и даже метод Гаусса ничем здесь не поможет. Но об этом позже.

Для начала нужно понять, что представляют собой собственные значения и для чего они служат. Одно из применений собственных значений, а именно то, с помощью которого мы их введем, связано с решением системы обыкновенных дифференциальных уравнений. При этом читателю не нужно быть специалистом в области дифференциальных уравнений, достаточно лишь

уметь дифференцировать обычные функции типа  $x^n$ ,  $\sin x$  и  $e^x$ . В качестве примера рассмотрим систему двух уравнений

$$\begin{aligned} \frac{dv}{dt} &= 4v - 5w, & v &= 8 \text{ при } t=0, \\ \frac{dw}{dt} &= 2v - 3w, & w &= 5 \text{ при } t=0. \end{aligned} \quad (1)$$

Это начальная задача в отличие от краевой задачи из § 1.6. Решение задано лишь в момент времени  $t=0$ , а не на обоих концах некоторого отрезка, т. е. нас интересует неустановившийся режим, а не стационарный. Решение изменяется во времени от заданных начальных значений, и нам требуется проследить за этой эволюцией.

Эту систему легко записать в матричной форме. Обозначим вектор неизвестных через  $u$ , его начальное значение через  $u_0$  и матрицу коэффициентов через  $A$ :

$$u(t) = \begin{bmatrix} v(t) \\ w(t) \end{bmatrix}, \quad u_0 = \begin{bmatrix} 8 \\ 5 \end{bmatrix}, \quad A = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix}.$$

В этих обозначениях система принимает вид

$$\frac{du}{dt} = Au, \quad u = u_0 \text{ при } t=0. \quad (2)$$

Это обычная постановка задачи. Отметим, что уравнение (2) имеет первый порядок, поскольку производные более высокого порядка отсутствуют, и оно является линейным относительно искомой функции. Наиболее общая линейная начальная задача для системы первого порядка будет иметь вид

$$\frac{du}{dt} = A(t)u + b(t), \quad u = u_0 \text{ при } t=0. \quad (3)$$

В нашем случае задача (2) является однородной ( $b=0$ ) и имеет постоянные коэффициенты, т. е. матрица  $A$  не зависит от времени.

Как найти решение этой задачи? Если бы она содержала одну неизвестную функцию вместо двух, то ответить на этот вопрос было бы легко. В этом случае векторное дифференциальное уравнение превращается в скалярное и в случае его однородности и постоянства коэффициентов имеет вид

$$\frac{du}{dt} = au, \quad u = u_0 \text{ при } t=0. \quad (4)$$

Решение этой задачи хорошо известно:

$$u(t) = e^{at}u_0. \quad (5)$$

В начальный момент времени  $t=0$  решение принимает требуемое значение  $u_0$ . Решение неустойчиво при  $a > 0$ , нейтрально устой-

чиво при  $a=0$  и устойчиво при  $a < 0$ ; оно либо уходит в бесконечность, либо остается ограниченным, либо стремится к нулю (при  $t \rightarrow \infty$ ). Если  $a$  — комплексное число,  $a = \alpha + i\beta$ , то те же критерии должны применяться к вещественной части  $\alpha$ . Решение имеет вид  $e^{(\alpha+i\beta)t} u_0$ , и мнимая его часть дает осциллирующий член  $e^{i\beta t} = \cos \beta t + i \sin \beta t$ , в то время как устойчивость определяется множителем  $e^{\alpha t}$ .

Пожалуй, хватит заниматься одним уравнением. Перейдем теперь непосредственно к системам и будем искать их решения с той же экспоненциальной зависимостью от времени, которая была установлена в скалярном случае. Другими словами, будем искать решения в виде

$$\begin{aligned} v(t) &= e^{\lambda t} y, \\ w(t) &= e^{\lambda t} z, \end{aligned} \quad (6a)$$

или, в векторной записи,

$$u(t) = e^{\lambda t} x. \quad (6b)$$

Подставляя это гипотетическое решение в систему дифференциальных уравнений, получаем

$$\begin{aligned} \lambda e^{\lambda t} y &= 4e^{\lambda t} y - 5e^{\lambda t} z, \\ \lambda e^{\lambda t} z &= 2e^{\lambda t} y - 3e^{\lambda t} z. \end{aligned}$$

Множитель  $e^{\lambda t}$  является общим для всех членов, и на него можно сократить. Это сокращение возможно в силу предположения, что для обоих неизвестных показатель  $\lambda$  один и тот же; если бы  $v$  и  $w$  были пропорциональны различным экспоненциальным функциям  $e^{\lambda t}$  и  $e^{\rho t}$ , то эти функции присутствовали бы в каждом из уравнений, и произвести сокращение было бы невозможно. Но в нашем случае это возможно, и мы получаем систему

$$\begin{aligned} 4y - 5z &= \lambda y, \\ 2y - 3z &= \lambda z. \end{aligned} \quad (7)$$

**Подстановка**  $u = e^{\lambda t} x$  в  $du/dt = Au$  дает  $\lambda e^{\lambda t} x = A e^{\lambda t} x$  и, после сокращения,

$$Ax = \lambda x. \quad (8)$$

Это основное уравнение относительно собственного значения  $\lambda$  и собственного вектора  $x$ . Отметим, что это уравнение является нелинейным, поскольку оно содержит произведение обоих неизвестных  $\lambda$  и  $x$ . Если мы знаем число  $\lambda$ , то уравнение относительно вектора  $x$  становится линейным; мы можем записать  $\lambda/x$  вместо  $\lambda x^1$ ) и перенести этот член в левую часть:

$$(A - \lambda I)x = 0. \quad (9)$$

<sup>1</sup>) Единичная матрица вводится для единообразия; выражение  $(A - \lambda)x$  короче, но не является вполне корректным.

Очевидно, что собственный вектор  $x$  лежит в нуль-пространстве матрицы

$$A - \lambda I = \begin{bmatrix} 4 - \lambda & -5 \\ 2 & -3 - \lambda \end{bmatrix}.$$

Здесь имеется одно существенное обстоятельство, а именно: для каждого значения  $\lambda$  вектор  $x = 0$  всегда удовлетворяет уравнению  $Ax = \lambda x$ , т. е. нуль-пространство любой матрицы содержит элемент  $x = 0$ . Но нулевой вектор бесполезен при построении решения  $u(t)$  из экспонент  $e^{\lambda t}x$ . Поэтому нас интересуют только те значения  $\lambda$ , для которых имеется ненулевой собственный вектор  $x$ . Нуль-пространство матрицы  $A - \lambda I$  имеет смысл рассматривать только тогда, когда оно содержит некоторый ненулевой вектор, и поэтому ранг этой матрицы должен быть меньше, чем ее порядок. Другими словами, матрица  $A - \lambda I$  должна быть вырожденной.

Критерий вырожденности дается определителем матрицы.

**5А.** Число  $\lambda$  является **собственным значением** матрицы  $A$  с соответствующим ненулевым собственным вектором тогда и только тогда, когда

$$\det(A - \lambda I) = 0. \quad (10)$$

Это **характеристическое уравнение** для матрицы  $A$ .

В нашем примере

$$\det \begin{bmatrix} 4 - \lambda & -5 \\ 2 & -3 - \lambda \end{bmatrix} = (4 - \lambda)(-3 - \lambda) + 10 = \lambda^2 - \lambda - 2.$$

Характеристический многочлен  $\lambda^2 - \lambda - 2$  разлагается в произведение  $(\lambda + 1)(\lambda - 2)$ , и матрица  $A$  имеет два различных собственных значения  $\lambda_1 = -1$  и  $\lambda_2 = 2$ . Каждому из этих значений соответствует пространство собственных векторов, удовлетворяющих уравнению  $Ax = \lambda x$  или  $(A - \lambda I)x = 0$ . Вычисления проводятся отдельно для  $\lambda_1$  и  $\lambda_2$ :

$$\lambda_1 = -1: \quad (A - \lambda_1 I)x = \begin{bmatrix} 5 & -5 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

и решением является любой вектор, кратный  $x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ;

$$\lambda_2 = 2: \quad (A - \lambda_2 I)x = \begin{bmatrix} 2 & -5 \\ 2 & -5 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

и решением является любой вектор, кратный  $x_2 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ .

Собственные векторы *определены неоднозначно*, поскольку любой вектор из нуль-пространства матрицы  $A - \lambda I$  (которое мы называем *собственным пространством*, соответствующим  $\lambda$ ) является собственным вектором, и поэтому нам нужно ввести в этом пространстве базис. В нашем случае оба собственных пространства являются одномерными и они порождены соответственно векторами  $x_1$  и  $x_2$ .

Отсюда, возвращаясь к дифференциальному уравнению, получаем два чисто экспоненциальных решения:

$$u = e^{\lambda_1 t} x_1 = e^{-t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{и} \quad u = e^{\lambda_2 t} x_2 = e^{2t} \begin{bmatrix} 5 \\ 2 \end{bmatrix}.$$

Поскольку исходное уравнение является линейным и однородным, оно допускает суперпозицию решений, т. е. любая комбинация двух его частных решений

$$u = c_1 e^{\lambda_1 t} x_1 + c_2 e^{\lambda_2 t} x_2 \quad (11)$$

вновь будет его решением. Сюда входят два произвольных параметра  $c_1$  и  $c_2$ , и естественно предполагать, что при надлежащем их выборе будет удовлетворяться начальное условие  $u = u_0$  при  $t = 0$ :

$$c_1 x_1 + c_2 x_2 = u_0, \quad (12)$$

или

$$\begin{bmatrix} 1 & 5 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 8 \\ 5 \end{bmatrix}. \quad (13)$$

Отсюда  $c_1 = 3$  и  $c_2 = 1$ , и *требуемое решение исходного уравнения (1) есть*

$$u(t) = 3e^{-t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + e^{2t} \begin{bmatrix} 5 \\ 2 \end{bmatrix}. \quad (14)$$

Записывая отдельно каждую из двух компонент, получаем

$$v(t) = 3e^{-t} + 5e^{2t}, \quad w(t) = 3e^{-t} + 2e^{2t}.$$

Начальные условия  $v_0 = 8$  и  $w_0 = 5$  проверяются легко.

Из приведенного примера возникает впечатление, что собственные значения и собственные векторы являются ключом к решению уравнений. Однако этот пример не показывает их физического смысла. Собственные значения и собственные векторы важны сами по себе, а не только как часть приема по отысканию решения  $u$ . Вероятно, наиболее известным является пример<sup>1)</sup> с солдатами, идущими по мосту. При переходе моста они пере-

<sup>1)</sup> В истинность которого я, кстати, никогда не верил.

стают идти в ногу. Причина состоит в том, что частота, с которой они маршируют, может совпасть с одной из собственных частот колебаний моста и привести к резонансу. (Так же, как в детских качелях; вы вскоре замечаете естественную частоту качания и, попадая в такт, раскачиваетесь сильнее.) Инженер пытается увести естественные частоты своего моста или ракеты от тех, которые могут быть вызваны ветром или колебаниями топлива. Другую крайность представляет биржевый игрок, который проводит жизнь, пытаясь попасть в такт с естественными колебаниями рынка. Собственные значения являются наиболее важной чертой практически любой динамической системы.

Остановимся теперь и посмотрим, что уже сделано и что остается сделать. Это введение показывает, каким образом собственные значения и собственные векторы матрицы  $A$  появляются естественно и автоматически при решении уравнения  $du/dt = Au$ . Такое уравнение имеет «чисто экспоненциальные решения» вида  $u = e^{\lambda t}x$ ; собственное значение дает скорость, с которой растет или убывает собственный вектор  $x$ . Другие решения будут комбинациями этих «чистых» решений, составленными таким образом, чтобы удовлетворялись начальные условия.

Основным здесь было уравнение  $Ax = \lambda x$ . Большая часть векторов  $x$  не будет удовлетворять такому уравнению независимо от того, является ли  $\lambda$  собственным значением или нет. Обычно  $x$  меняет направление после умножения на  $A$ , так что вектор  $Ax$  не будет кратным  $x$ . Это означает, что *только специальные числа  $\lambda$  являются собственными значениями и только специальные векторы являются собственными векторами*. Разумеется, если  $A$  кратна единичной матрице, то ни один из векторов не будет изменять своего направления и все они будут собственными. Но в обычном случае собственных векторов мало и они направлены в разные стороны.

**Пример.** Представим себе какое-либо вращение земного шара. Тогда существует одно направление, называемое осью вращения, которое остается неизменным. Оно не обязательно является осью, вокруг которой на самом деле происходит вращение Земли, но все равно должны существовать *некоторые* «северный и южный полюсы», остающиеся неподвижными<sup>1)</sup>. Эти полюсы суть собственные векторы с собственным значением, равным 1. В общем случае все остальные точки находятся в движении и не видно никаких других собственных векторов. Единственные исключения возникают, когда вращение происходит на  $360^\circ$  (и тогда все векторы собственные) и на  $180^\circ$ . В последнем случае плоскость экватора заполнена собственными векторами; каждое направление в этой плоскости

<sup>1)</sup> Я думаю, что это не так очевидно, но это правда — нельзя повернуть Землю так, чтобы двигались все точки.

изменяется в точности на противоположное и собственное значение равняется  $-1$ . Эта экваториальная плоскость соответствует случаю «двумерного собственного пространства»; одно собственное значение  $\lambda = -1$  имеет два независимых собственных вектора и, следовательно, целую плоскость собственных векторов.

С одной стороны, этот пример хорош, с другой — нет. Во-первых, собственные значения в общем случае не равны  $\pm 1$ ; векторы обычно либо растягиваются, либо сжимаются. Более того, это было вращение трехмерного пространства  $\mathbb{R}^3$ , но, кроме случаев  $180^\circ$  и  $360^\circ$ , *имелась лишь одна ось собственных векторов, а мы надеялись найти три*. В применении к дифференциальным уравнениям нам будут нужны три различных частных решения, чтобы удовлетворить начальным условиям; одного собственного значения и одного собственного вектора недостаточно. Для того чтобы найти два других, нужно ввести мнимое число  $i$ . Если вещественный мир  $\mathbb{R}^3$  допускает слишком мало решений, то нужно искать их в пространстве  $\mathbb{C}^3$  векторов с комплексными компонентами. Если мы допускаем комплексные числа, то любая матрица размера  $n \times n$  будет иметь  $n$  собственных значений. Тройка собственных значений и собственных векторов для этого примера вращающейся Земли будет найдена в § 5.2, а в § 5.5 мы осуществим полный переход от вещественных векторов и матриц к комплексному случаю.

Истинная цель этой главы состоит, однако, в другом. Важнее всего сейчас объяснить, как система уравнений «распадается» после отыскания собственных векторов. Эти собственные векторы являются «гармониками» системы и независимы друг от друга. Мы можем наблюдать за поведением каждого из собственных векторов в отдельности, а затем комбинировать эти гармоники для отыскания решения. Выражая это же другими словами, *мы диагоналируем исходную матрицу*.

Мы предполагаем посвятить § 5.2 теории диагонализации, а оставшиеся параграфы — ее применениям, сначала к разностным уравнениям, числам Фибоначчи и марковским процессам, а затем к дифференциальным уравнениям. В каждом из примеров мы должны начинать с вычисления собственных значений и собственных векторов; избежать этого не удастся. Но затем примеры пойдут в настолько разных направлениях, что краткое резюме окажется невозможным. Можно лишь отметить, что работать с симметрическими матрицами особенно легко, а с некоторыми «дефектными матрицами» — особенно трудно. Они не обладают полным набором собственных векторов, не диагонализуются и для них не проходит техника «гармоник». Разумеется, ими следует заниматься, но мы не намерены допустить, чтобы это заняло всю оставшуюся часть книги.

Чтобы завершить это введение, дадим обзор некоторых основных фактов, касающихся собственных значений и собственных

векторов, т. е. задачи  $Ax = \lambda x$ . Если задана матрица  $A$  порядка  $n$ , то задача состоит в отыскании тех особых векторов  $x$ , на которые  $A$  действует как простое умножение — направления векторов  $Ax$  и  $x$  совпадают. С этой целью мы сначала находим собственные значения, переписывая  $\lambda x$  в виде  $\lambda Ix$  и перенося этот член в левую часть:  $(A - \lambda I)x = 0$ . Собственный вектор матрицы  $A$  есть вектор из нуль-пространства матрицы  $A - \lambda I$ . Другими словами, основной вопрос состоит в следующем: если  $A$  сдвигается на различные кратные единичной матрицы, то какой сдвиг делает ее вырожденной? Эти сдвиги дадут собственные значения, и мы можем затем вычислить собственные векторы.

Если сама матрица  $A$  вырождена, то одна возможность состоит в том, чтобы вообще не производить сдвига. Одно из собственных значений вырожденной матрицы есть  $\lambda = 0$ , и ее нуль-пространство содержит соответствующие собственные векторы. Но в отличие от первой части книги в вырожденности матрицы теперь нет ничего особенного; это означает лишь, что  $\lambda = 0$  является собственным значением этой матрицы. Является ли  $A$  вырожденной или нет, все ее собственные значения равноправны: комбинация  $A - \lambda I$  вырождена и нуль-пространство этой комбинации есть пространство собственных векторов, соответствующих  $\lambda$ .

Чтобы определить, когда  $A - \lambda I$  вырождена, мы вычисляем ее определитель. Этот определитель есть многочлен от  $\lambda$  степени  $n$ , называемый *характеристическим многочленом* матрицы  $A$ . Уравнение  $\det(A - \lambda I) = 0$  есть *характеристическое уравнение*, и его корни  $\lambda_1, \lambda_2, \dots, \lambda_n$  (которые могут быть, а могут и не быть вещественными числами и среди которых могут оказаться или не оказаться совпадающие) суть собственные значения матрицы  $A$ . Суммируя, получаем:

- 5В.** Каждое из следующих условий является необходимым и достаточным для того, чтобы  $\lambda$  был собственным значением для  $A$ :
- (1) Существует ненулевой вектор  $x$ , такой, что  $Ax = \lambda x$ .
  - (2) Матрица  $A - \lambda I$  вырождена.
  - (3)  $\det(A - \lambda I) = 0$ .

**Пример.** Рассмотрим симметрическую матрицу

$$A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

Ее характеристический многочлен есть

$$\det(A - \lambda I) = \begin{vmatrix} 1 - \lambda & -1 & 0 \\ -1 & 2 - \lambda & -1 \\ 0 & -1 & 1 - \lambda \end{vmatrix} = -\lambda^3 + 4\lambda^2 - 3\lambda.$$

Следовательно, характеристическое уравнение имеет вид

$$-\lambda^3 + 4\lambda^2 - 3\lambda = -\lambda(\lambda - 1)(\lambda - 3) = 0,$$

и все собственные значения вещественны и различны:  $\lambda_1 = 0$ ,  $\lambda_2 = 1$ ,  $\lambda_3 = 3$ . Отдельно для каждого собственного значения  $\lambda_i$  ищем соответствующий собственный вектор  $x_i$ :

$$\begin{aligned} \lambda_1 = 0: \quad (A - 0I)x_1 &= \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} x_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{и} \quad x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \\ \lambda_2 = 1: \quad (A - I)x_2 &= \begin{bmatrix} 0 & -1 & 0 \\ -1 & 1 & -1 \\ 0 & -1 & 0 \end{bmatrix} x_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{и} \quad x_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}. \\ \lambda_3 = 3: \quad (A - 3I)x_3 &= \begin{bmatrix} -2 & -1 & 0 \\ -1 & -1 & -1 \\ 0 & -1 & -2 \end{bmatrix} x_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{и} \quad x_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}. \end{aligned}$$

Это обилие вычислений неизбежно, и, разумеется, нами был выбран хороший пример: большинство кубических многочленов нельзя разложить с такой легкостью, как данный. Нет сомнения, что задача на собственные значения алгебраически и вычислительно значительно более трудна, чем решение системы  $Ax = b$ . Для линейной системы конечное число шагов процесса исключения дает точный ответ за конечное время. (Или, что эквивалентно, правило Крамера дает точную формулу для решения.) В случае с собственными значениями не может существовать такого рода шагов и такой формулы, или Гауа перевернулся бы в гробу: характеристический многочлен матрицы порядка 5 имеет пятую степень, а он доказал, что для нахождения корней многочлена пятой степени не существует алгебраической формулы. Все, что возможно, — это несколько простых тестов для *уже вычисленных* собственных значений, и мы упомянем два из них:

5С. Сумма  $n$  собственных значений равняется сумме  $n$  диагональных элементов матрицы  $A$ :

$$\lambda_1 + \dots + \lambda_n = a_{11} + \dots + a_{nn}.$$

Эта сумма называется **следом** матрицы  $A$ . Кроме того, произведение  $n$  собственных значений равняется определителю матрицы  $A$ .

Эти свойства подтверждаются приведенным выше примером, в котором след  $A$  равен  $0 + 1 + 3 = 1 + 2 + 1 = 4$  и определитель равен  $0 \cdot 1 \cdot 3 = 0$ .

**Упражнение 5.1.1.** Найти собственные значения и собственные векторы матрицы  $A = \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix}$ . Проверить, что ее след равен сумме собственных значений, а определитель — их произведению.

**Упражнение 5.1.2.** При той же самой матрице  $A$  решить дифференциальное уравнение  $du/dt = Au$ ,  $u_0 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$ .

**Упражнение 5.1.3.** Предположим, что мы сдвинули указанную выше матрицу  $A$ , вычтя из нее  $7I$ , т. е.

$$B = A - 7I = \begin{bmatrix} -6 & -1 \\ 2 & -3 \end{bmatrix}.$$

Чему равны собственные значения и собственные векторы матрицы  $B$  и как они связаны с соответствующими величинами для  $A$ ?

**Упражнение 5.1.4.** Используя пример с матрицей 3-го порядка, собственные значения и собственные векторы которой были вычислены выше, подобрать константы  $c_i$  так, чтобы решение  $u = c_1 e^{\lambda_1 t} x_1 + c_2 e^{\lambda_2 t} x_2 + c_3 e^{\lambda_3 t} x_3$  удовлетворяло начальному условию  $u_0 = [1 \ 3 \ 1]^T$ .

Не следует путать собственные значения матрицы и ее диагональные элементы. Обычно они совершенно различные. Тем не менее, рискуя внести некоторую путаницу, мы все же укажем одну ситуацию, в которой эти числа совпадают.

**5D.** Если матрица  $A$  треугольная — она может быть верхней треугольной, нижней треугольной или диагональной, — то собственные значения  $\lambda_1, \dots, \lambda_n$  в точности совпадают с диагональными элементами  $a_{11}, \dots, a_{nn}$ .

Причина этого явления станет ясна из примера. Если

$$A = \begin{bmatrix} 1 & \frac{1}{4} & 0 \\ 0 & \frac{3}{4} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \end{bmatrix},$$

то ее характеристический многочлен есть

$$\det \begin{bmatrix} 1-\lambda & \frac{1}{4} & 0 \\ 0 & \frac{3}{4}-\lambda & \frac{1}{2} \\ 0 & 0 & \frac{1}{2}-\lambda \end{bmatrix} = (1-\lambda) \left( \frac{3}{4}-\lambda \right) \left( \frac{1}{2}-\lambda \right).$$

Определитель в точности равен произведению диагональных элементов. Очевидно, что корни равны  $\lambda = 1$ ,  $\lambda = 3/4$  и  $\lambda = 1/2$ ; собственные значения сразу находились на главной диагонали.

**Упражнение 5.1.5.** Найти собственные векторы этой треугольной матрицы, а также собственные значения и собственные векторы диагональной матрицы

$$A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Эти примеры, в которых собственные значения находятся сразу же по виду матрицы, указывают на главную тему всей главы: привести матрицу  $A$  к диагональной или треугольной форме, *не изменяя ее собственных значений*. Подчеркнем еще раз, что разложение Гаусса  $A=LU$  для этой цели не подходит. Собственные значения матрицы  $U$  расположены на ее диагонали, но это не есть собственные значения матрицы  $A$ .

Имеется еще одна ситуация, в которой легко проводить вычисления. Предположим, что мы уже нашли собственные значения и собственные векторы матрицы  $A$ . Тогда **собственные значения матрицы  $A^2$  в точности равны  $\lambda_1^2, \dots, \lambda_n^2$  и каждый собственный вектор матрицы  $A$  является также и собственным вектором для  $A^2$** . Доказательство типично для математики: если мы попытаемся исследовать  $\det(A^2 - \lambda I)$ , то мы в нем погрязнем, но если начать с  $Ax = \lambda x$ , то все становится ясным. Вновь умножая на  $A$ , получаем

$$A^2x = A \lambda x = \lambda Ax = \lambda^2 x.$$

Отсюда следует, что  $\lambda^2$  является собственным значением для  $A^2$  с тем же собственным вектором  $x$ . Если первое умножение на  $A$  не изменило направления вектора  $x$ , то же самое будет и при втором.

Это рассуждение теряет силу, когда имеются две различные матрицы. Предположим, что  $\lambda$  есть собственное значение матрицы  $A$ , а  $\mu$  — собственное значение  $B$ . Тогда  $\lambda\mu$ , *вообще говоря, не является собственным значением матрицы  $AB$* . Попытка доказательства имела бы такой вид: если  $Ax = \lambda x$  и  $Bx = \mu x$ , то  $ABx = A\mu x = \mu Ax = \mu\lambda x$ . Ошибка заключена в предположении, что  $A$  и  $B$  имеют один и тот же собственный вектор  $x$ . В общем случае это не так.

**Упражнение 5.1.6.** Привести пример, показывающий, что собственные значения могут изменяться, когда кратное одной строки вычитается из другой.

**Упражнение 5.1.7.** Пусть  $\lambda$  — собственное значение матрицы  $A$  и  $x$  — соответствующий собственный вектор;

(а) Показать, что этот же самый  $x$  является собственным вектором матрицы  $B = A - 7I$ , и найти соответствующее собственное значение. Результат должен подтверждать упр. 5.1.3.

(б) Предполагая, что  $\lambda \neq 0$ , показать, что  $x$  также является собственным вектором матрицы  $A^{-1}$ , и вновь найти собственное значение.

**Упражнение 5.1.8.** Показать, что определитель равняется произведению собственных значений, представив характеристический многочлен в виде

$$\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda) \quad (15)$$

и выбирая  $\lambda$  очевидным образом.

**Упражнение 5.1.9.** Показать в два этапа, что след матрицы равняется сумме собственных значений. Во-первых, вычислить коэффициент при  $(-\lambda)^{n-1}$  в правой части (15). Во-вторых, найти все члены в

$$\det(A - \lambda I) = \det \begin{bmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{bmatrix},$$

содержащие  $(-\lambda)^{n-1}$ . Объяснить, почему все они получаются из произведений элементов главной диагонали и найти коэффициент при  $(-\lambda)^{n-1}$  в левой части (15). Сравнить.

**Упражнение 5.1.10.** (а) Построить матрицы  $A, B$  размера  $2 \times 2$ , такие, что собственные значения матрицы  $AB$  не равны произведениям собственных значений матриц  $A$  и  $B$  и собственные значения  $A+B$  не равны суммам их собственных значений.

(б) Проверить, что при этом, однако, сумма собственных значений матрицы  $A+B$  равняется сумме всех собственных значений матриц  $A$  и  $B$  и что то же самое справедливо для произведений. Почему это так?

**Упражнение 5.1.11.** Доказать, что  $A$  и  $A^T$  имеют одинаковые собственные значения, сравнивая их характеристические многочлены.

**Упражнение 5.1.12.** Найти собственные значения и собственные векторы матрицы  $A = \begin{bmatrix} 3 & 4 \\ 4 & -3 \end{bmatrix}$ .

**Упражнение 5.1.13.** Найти собственные значения и проверить, что их сумма равняется следу, для матрицы

$$A = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}.$$

## § 5.2. ДИАГОНАЛЬНАЯ ФОРМА МАТРИЦЫ

Начнем с одного существенного замечания. Оно исключительно простое и будет использоваться в каждом параграфе этой главы.

**5E.** Предположим, что матрица  $A$  размера  $n \times n$  имеет  $n$  линейно независимых собственных векторов. Тогда если взять эти векторы в качестве столбцов матрицы  $S$ , то  $S^{-1}AS$  будет диагональной матрицей  $\Lambda$ , у которой на диагонали стоят соб-

ственные значения матрицы  $A$ :

$$S^{-1}AS = \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}. \quad (16)$$

**Доказательство.** Расположим собственные векторы  $x_i$  в столбцах матрицы  $S$  и вычислим произведение  $AS$ . Умножая поочередно на каждый столбец, получим

$$AS = A \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \dots & x_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1 x_1 & \lambda_2 x_2 & \dots & \lambda_n x_n \\ | & | & & | \end{bmatrix}.$$

Прием состоит в разложении последней матрицы в совершенно другое произведение:

$$\begin{bmatrix} | & | & & | \\ \lambda_1 x_1 & \lambda_2 x_2 & \dots & \lambda_n x_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \dots & x_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

Рассматривая это просто как упражнение по перемножению матриц, важно поставить эти матрицы в нужном порядке. Если  $\Lambda$  будет стоять до  $S$ , то  $\lambda_1$  будет умножаться на элементы первой строки, в то время как этот множитель нужен нам в первом столбце. В данном случае мы имеем правильное произведение  $S\Lambda$ . Поэтому

$$AS = SA, \quad \text{или} \quad S^{-1}AS = \Lambda, \quad \text{или} \quad A = SAS^{-1}. \quad (17)$$

Матрица  $S$  обратима, поскольку ее столбцы (собственные векторы) предполагались линейно независимыми.

Прежде чем перейти к примерам, сделаем четыре замечания.

*Замечание 1.* Если матрица  $A$  не имеет кратных собственных значений — числа  $\lambda_1, \dots, \lambda_n$  различны, — то  $n$  собственных векторов автоматически являются независимыми (см. 5F ниже). Поэтому любая матрица с различными собственными значениями может быть приведена к диагональному виду.

*Замечание 2.* Диагонализирующая матрица  $S$  *неединственна*. Во-первых, собственный вектор  $x$  может умножаться на константу и останется при этом собственным вектором. Поэтому мы можем умножать столбцы матрицы  $S$  на любые ненулевые константы, что дает новую диагонализирующую матрицу  $S$ . Кратные собственные значения дают даже бóльшую свободу, и для тривиального примера  $A = I$  подойдет любая обратимая матрица  $S$ :  $S^{-1}IS$  всегда диагональна (и диагональная матрица  $\Lambda$  есть просто  $I$ ). Это отражает тот факт, что для единичной матрицы все векторы являются собственными.

*Замечание 3.* Равенство  $AS = SA$  выполняется только в том случае, когда столбцы матрицы  $S$  являются собственными векторами матрицы  $A$ . Другие матрицы  $S$  не дадут диагональной матрицы  $\Lambda$ . Причина этого заложена в правилах умножения матриц. Предположим, что первый столбец матрицы  $S$  есть некоторый вектор  $y$ . Тогда первый столбец матрицы  $SA$  есть  $\lambda_1 y$ . Если это совпадает с первым столбцом матрицы  $AS$ , который по правилам умножения матриц равен  $Ay$ , то  $y$  должен быть собственным вектором:  $Ay = \lambda_1 y$ . Фактически порядок расположения собственных векторов в  $S$  и собственных значений в  $\Lambda$  автоматически совпадает.

*Замечание 4.* Не все матрицы имеют  $n$  линейно независимых собственных векторов, и поэтому *не все матрицы диагонализуются*. Стандартный пример «дефектной матрицы» имеет вид

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Ее собственные значения равны  $\lambda_1 = \lambda_2 = 0$ , поскольку она треугольная:

$$\det(A - \lambda I) = \det \begin{bmatrix} -\lambda & 1 \\ 0 & -\lambda \end{bmatrix} = \lambda^2.$$

Если  $x$  есть собственный вектор, то

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \text{или} \quad x = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}.$$

Хотя  $\lambda = 0$  есть собственное значение, *алгебраическая кратность* которого равна 2, оно имеет только одномерное пространство собственных векторов. *Геометрическая кратность* этого собственного значения равна 1, и мы не можем построить  $S$ .

Приведем более простое доказательство того, что  $A$  недиагонализуема. Поскольку  $\lambda_1 = \lambda_2 = 0$ ,  $\Lambda$  должна быть нулевой матрицей. Но если  $S^{-1}AS = 0$ , то, умножая слева на  $S$  и справа

на  $S^{-1}$ , получаем, что  $A=0$ . Поскольку  $A$  не является нулевой матрицей, это противоречие доказывает, что никакая матрица  $S$  не дает  $S^{-1}AS = \Lambda$ .

Некоторые матрицы с кратными собственными значениями могут быть приведены к диагональному виду (например,  $A=I$ ), другие — нет. Единственный тест состоит в вычислении всех собственных векторов и определении, достаточно ли их. Когда все собственные значения различны, все просто: каждая алгебраическая или геометрическая кратность равна 1. Но если собственное значение  $\lambda$  повторяется  $m$  раз, то все определяется нуль-пространством матрицы  $A - \lambda I$ ;  $\lambda$  проходит тест только в том случае, когда есть  $m$  соответствующих собственных векторов. Когда все собственные значения пройдут тест, мы будем иметь полный набор собственных векторов и  $A$  может быть диагонализирована.

Для завершения этого круга идей нам остается доказать полезную, но не очень впечатляющую теорему:

**5F.** Если ненулевые собственные векторы  $x_1, \dots, x_n$  соответствуют различным собственным значениям  $\lambda_1, \dots, \lambda_n$ , то эти собственные векторы линейно независимы.

Предположим сначала, что  $k=2$  и что некоторая комбинация  $x_1$  и  $x_2$  дает нуль:  $c_1x_1 + c_2x_2 = 0$ . Умножая на  $A$ , получаем  $c_1\lambda_1x_1 + c_2\lambda_2x_2 = 0$ . При вычитании произведения  $\lambda_2$  на предыдущее уравнение вектор  $x_2$  исчезает:

$$c_1(\lambda_1 - \lambda_2)x_1 = 0.$$

Поскольку  $\lambda_1 \neq \lambda_2$  и  $x_1 \neq 0$ , обязательно  $c_1 = 0$ . Аналогично  $c_2 = 0$  и эти два вектора независимы, т. е. только тривиальная комбинация дает нуль.

Это рассуждение может быть распространено на произвольное число собственных векторов: мы предполагаем, что некоторая комбинация дает нуль, умножаем на  $A$ , вычитаем произведение  $\lambda_k$  на исходную комбинацию, и вектор  $x_k$  исчезает, оставляя комбинацию  $x_1, \dots, x_{k-1}$ , которая дает нуль. Повторяя эти шаги, мы приходим к кратному  $x_1$ , которое равно нулю. Это непременно дает  $c_1 = 0$  и окончательно каждый  $c_i = 0$ . (Можно сослаться также на математическую индукцию.) Следовательно, собственные векторы, отвечающие различным собственным значениям, автоматически линейно независимы.

**Упражнение 5.2.1.** Для матрицы

$$A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

со стр. 217 найти  $S$  и вычислить  $S^{-1}AS$ . Кроме того, диагонализировать матрицу  $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ .

**Упражнение 5.2.2.** Найти матрицу  $A$ , собственные значения которой равны 1 и 4, а собственные векторы имеют вид  $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$  и  $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$  соответственно. (Указание:  $A = SAS^{-1}$ .)

**Упражнение 5.2.3.** Найти все собственные значения и собственные векторы матрицы

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

и выписать две различные диагонализующие матрицы  $S$ .

**Упражнение 5.2.4.** Транспонированием соотношения  $S^{-1}AS = \Lambda$  найти матрицу, диагонализующую  $A^T$ , и выписать получающуюся диагональную матрицу.

**Упражнение 5.2.5.** Пусть  $A$  и  $B$  имеют одну и ту же матрицу собственных векторов  $S$ , так что  $A = SA_1S^{-1}$  и  $B = SA_2S^{-1}$ . Доказать, что  $AB = BA$ . (Заметить, что  $\Lambda_1\Lambda_2 = \Lambda_2\Lambda_1$  в силу их диагональности.) Для матрицы  $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$  найти подходящую матрицу  $B$ .

Последний пример играет важную роль в квантовой механике: матрицы с одними и теми же собственными векторами должны коммутировать. Обратное также верно и даже более важно: *если  $AB = BA$ , то эти матрицы имеют одинаковые собственные векторы*. Основной момент доказательства состоит в том, что равенство  $Ax = \lambda x$  влечет за собой равенства  $ABx = BAx = B\lambda x = \lambda Bx$ . Поэтому  $x$  и  $Bx$  являются собственными векторами, отвечающими одному и тому же  $\lambda$ , и если мы для удобства предположим, что собственные значения матрицы  $A$  различны, т. е. все собственные пространства одномерны, то вектор  $Bx$  должен быть кратным вектору  $x$ . Другими словами,  $x$  есть собственный вектор матриц  $B$  и  $A$ , и доказательство закончено.

Во введении к этой главе мы говорили о вращениях земного шара. Поворот на  $90^\circ$  оставляет неизменным только одно направление; ось, соединяющая полюсы, была единственным собственным вектором, который мы смогли найти. В плоскости экватора ось  $x$  переходит в ось  $y$ , а ось  $y$  — в ось  $x$ , но с противоположным направлением. Если в исходный момент координаты некоторой точки на Земле были  $(x_0, y_0, z_0)$ , то в новых осях они принимают вид  $(y_0, -x_0, z_0)$ . Легко видеть, что это изменение

координат может быть осуществлено посредством перемножения матриц:

$$\begin{bmatrix} y_0 \\ -x_0 \\ z_0 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}.$$

Мы хотим диагонализировать эту матрицу вращения  $A$ . Первый шаг состоит в отыскании ее собственных значений и собственных векторов, начиная с характеристического многочлена

$$\det(A - \lambda I) = \begin{vmatrix} -\lambda & -1 & 0 \\ 1 & -\lambda & 0 \\ 0 & 0 & 1 - \lambda \end{vmatrix} = (1 - \lambda)(\lambda^2 + 1).$$

Корни этого многочлена суть собственные значения; два из них являются мнимыми, несмотря на то, что элементы самой матрицы — действительные числа:

$$\lambda_1 = 1, \quad \lambda_2 = i, \quad \lambda_3 = -i.$$

Поскольку они различные (хотя и мнимые!), матрица диагонализуется. Таким образом, должен существовать полный набор собственных векторов, вычисляемых обычным способом:

$$\lambda_1 = 1: \quad (A - I)x_1 = \begin{bmatrix} -1 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} x_1 = 0, \quad \text{или} \quad x_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Этот собственный вектор как раз и является осью север—юг, которая остается неподвижной;

$$\lambda_2 = i: \quad (A - iI)x_2 = \begin{bmatrix} -i & -1 & 0 \\ 1 & -i & 0 \\ 0 & 0 & 1 - i \end{bmatrix} x_2 = 0, \quad \text{или} \quad x_2 = \begin{bmatrix} 1 \\ -i \\ 0 \end{bmatrix}.$$

$$\lambda_3 = -i: \quad (A + iI)x_3 = \begin{bmatrix} i & -1 & 0 \\ 1 & i & 0 \\ 0 & 0 & 1 + i \end{bmatrix} x_3 = 0, \quad \text{или} \quad x_3 = \begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix}.$$

Следовательно, диагонализующая матрица  $S$  и диагональная матрица  $\Lambda$  равны соответственно

$$\begin{bmatrix} 0 & 1 & 1 \\ 0 & -i & i \\ 1 & 0 & 0 \end{bmatrix} \quad \text{и} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{bmatrix}.$$

Теория гарантирует, что матрица  $S^{-1}AS$  совпадает с  $\Lambda$ .

В этом примере смысл имеет также и матрица  $A^2$ . Геометрически она соответствует повороту на  $180^\circ$ ; это результат повторного применения матрицы  $A$ . Алгебраически равенство  $A = SAS^{-1}$  дает  $A^2 = SAS^{-1}SAS^{-1} = SA^2S^{-1}$ . Этим подтверждается тот факт, что матрица собственных векторов  $S$  является одной и той же для матриц  $A$  и  $A^2$  и что каждое из собственных значений возводится в квадрат. Собственные значения вращения на  $180^\circ$  равны  $1^2 = 1$ ,  $i^2 = -1$ ,  $(-i)^2 = -1$ ; кратность 2 для собственного значения  $-1$  показывает, что вся экваториальная плоскость обращается. Мы можем продолжить и вычислить матрицу  $A^4$ , которая соответствует полному повороту на  $360^\circ$ :

$$A^4 = (SAS^{-1})(SAS^{-1})(SAS^{-1})(SAS^{-1}) = SA^4S^{-1}.$$

Поскольку  $i^4 = 1$ , матрица собственных значений есть  $\Lambda^4 = I$  и  $A^4 = S I S^{-1} = I$ . Вращение на  $360^\circ$  является тождественным.

**Упражнение 5.2.6.** Можно ли найти квадратный корень  $R$  из матрицы вращения на  $90^\circ$   $A$  и проверить, что  $R^2 = A$ ? Помните, что матрицы  $R$  и  $R^2$  имеют одинаковые собственные векторы, или, другими словами, одну и ту же матрицу  $S$ . Фактически имеется восемь возможных квадратных корней  $R$ , поскольку  $\sqrt{1} = \pm 1$ ,  $\sqrt{i} = \pm(1+i)/\sqrt{2}$ ,  $\sqrt{-i} = \pm(1-i)/\sqrt{2}$ .

**Упражнение 5.2.7.** Выяснить, имеет ли матрица  $A = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$  два собственных вектора и поэтому диагонализуется или же она имеет один собственный вектор и не диагонализуется.

## § 5.3. РАЗНОСТНЫЕ УРАВНЕНИЯ И СТЕПЕНИ $A^k$

Не совсем ясно почему, но разностные уравнения известны не так широко, как дифференциальные. В них используется конечное число конечных шагов, в то время как дифференциальные уравнения требуют бесконечного числа бесконечно малых шагов; но обе теории абсолютно параллельны. Это та же самая аналогия между дискретным и непрерывным, которая вновь и вновь появляется в математике. Наилучшей иллюстрацией, вероятно, будет служить такая, для которой не требуется  $n$ -мерной линейной алгебры. Именно эта ситуация рассматривается нами ниже — ведь деньги в банке являются всего лишь скаляром.

Пусть мы инвестируем 1000 долларов на 5 лет и банк выплачивает нам 6%. Если проценты начисляются банком раз в год, то капитал умножается на 1.06 и  $P_{k+1} = 1.06P_k$ . Это разностное уравнение с шагом по времени в один год. Оно связывает величину капитала после  $k+1$  лет с величиной капитала за год до этого и легко решается: за 5 лет исходный капитал  $P_0 = 1000$

умножался пять раз и

$$P_5 = (1.06)^5 P_0 = (1.06)^5 1000 = 1338 \text{ долларов.}$$

Предположим теперь, что шаг по времени уменьшен до месяца. Новое разностное уравнение будет иметь вид  $p_{k+1} = (1 + 0.06/12) p_k$ . После 5 лет, или 60 месяцев,

$$p_{60} = \left(1 + \frac{0.06}{12}\right)^{60} p_0 = (1.005)^{60} 1000 = 1349 \text{ долларов.}$$

Следующий шаг состоит в ежедневном начислении процентов:

$$\left(1 + \frac{0.06}{365}\right)^{5 \cdot 365} 1000 = 1349.83 \text{ доллара.}$$

Наконец, чтобы окончательно расшевелить вкладчиков, банки предлагают теперь *непрерывное начисление процентов*. Прибыль добавляется в каждый момент и разностное уравнение не годится. Можно ожидать, что банковский служащий не знает дифференциального исчисления и не сможет вычислить, сколько он вам должен. У него имеются две возможности: либо он может исчислять прибыль все чаще и чаще и заметить, что предел равен

$$\left(1 + \frac{0.06}{N}\right)^{5N} 1000 \rightarrow e^{0.30} 1000 = 1349.87 \text{ доллара.}$$

Либо же он может перейти к дифференциальному уравнению, являющемуся пределом разностного уравнения  $p_{k+1} = (1 + 0.06 \Delta t) p_k$ . Переноса  $p_k$  в левую часть и деля на временной шаг  $\Delta t$ , получаем, что  $\frac{p_{k+1} - p_k}{\Delta t} = 0.06 p_k$  сходится к  $\frac{dp}{dt} = 0.06p$ . Решение есть  $p(t) = e^{0.06t} p_0$ , и после 5 лет эта величина вновь равняется 1349.87 доллара. Капитал остается конечным, даже когда начисление производится в каждый момент времени, и разница составляет всего 4 цента.

Этот пример содержал одновременно разностные и дифференциальные уравнения, и одно переходило в другое, когда шаг по времени стремился к нулю. Однако имеется множество разностных уравнений, которые не приводятся к дифференциальным, и наш второй пример возникает из знаменитой **последовательности Фибоначчи**:

$$0, 1, 1, 2, 3, 5, 8, 13, \dots$$

Вероятно вы уловили ее характер: каждое из чисел Фибоначчи равняется сумме двух предшествующих,

$$F_{k+2} = F_{k+1} + F_k. \quad (18)$$

Это разностное уравнение. Оно возникает в фантастическом числе приложений и заслуживает отдельной книги. Колочки и листья

расположены спиралевидно, и на боярышнике, яблоне или дубе вы обнаруживаете пять таких образований на каждые два оборота вокруг ствола. У грушевого дерева их восемь на каждые три оборота, а у ивы еще сложнее — 13 ростков на каждые пять витков. Чемпионом является, по-видимому, подсолнечник Даниеля Т. О'Коннела (Scientific American, ноябрь 1951 г.), семена которого удволяютворяют почти невероятному соотношению  $F_{12}/F_{13} = 144/233$  <sup>1)</sup>.

Каким образом можно найти 1000-е число Фибоначчи, не начиная с  $F_0 = 0$ ,  $F_1 = 1$  и не проходя весь путь до  $F_{1000}$ ? Задача состоит в решении разностного уравнения  $F_{k+2} = F_{k+1} + F_k$ . Для начала оно может быть приведено к «одношаговому уравнению»  $u_{k+1} = Au_k$ . Это уравнение в точности такое же, как и в примере на сложные проценты  $P_{k+1} = 1,06 P_k$ , за исключением того, что теперь неизвестное есть вектор, а множитель  $A$  — матрица: если

$$u_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix},$$

то

$$\begin{aligned} F_{k+2} &= F_{k+1} + F_k, \\ F_{k+1} &= F_{k+1} \end{aligned} \quad \text{дает} \quad u_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} u_k.$$

Этот переход является стандартным приемом для любого уравнения порядка  $s$ ;  $s=1$  тривиальных уравнений вида  $F_{k+1} = F_{k+1}$  сочетаются с заданным уравнением для получения одношаговой системы. В случае чисел Фибоначчи  $s=2$ .

Формально решить разностное уравнение  $u_{k+1} = Au_k$  легко. Поскольку каждый шаг дает умножение на матрицу  $A$ , решение  $u_k$  связано с начальным значением  $u_0$  соотношением  $u_k = A^k u_0$ . Трудность состоит в отыскании быстрого способа вычисления степеней  $A^k$  и нахождения в результате 1000-го числа Фибоначчи. Ключом к этому являются собственные числа и собственные векторы матрицы  $A$ :

**5G.** Если матрица  $A$  приводится к диагональному виду,  $A = SAS^{-1}$ , то автоматически

$$u_k = A^k u_0 = (SAS^{-1})(SAS^{-1}) \dots (SAS^{-1}) u_0 = SA^k S^{-1} u_0. \quad (19)$$

<sup>1)</sup> Относительно этих ботанических приложений см. книгу Д'Арси Томпсона (D'Arcy Thompson, On Growth and Form, Cambridge Univ. Press, London and N. Y., 1942) или прекрасную книгу Питера Стивенса (Peter Stevens, Patterns in Nature, Little, Brown, 1974). Сотни других свойств чисел  $F_n$  были опубликованы в Fibonacci Quarterly. По-видимому Фибоначчи был также первым человеком, введшим арабские цифры в Европе, примерно в 1200 г. н. э.

Все матрицы  $S^{-1}$  сокращаются при умножении на матрицы  $S$ , и остается лишь первая матрица  $S$  и последняя  $S^{-1}$ . Столбцы матрицы  $S$  суть собственные векторы  $x_i$  матрицы  $A$ , и перемножение матриц дает

$$u_k = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{bmatrix} S^{-1} u_0 = c_1 \lambda_1^k x_1 + \dots + c_n \lambda_n^k x_n. \quad (20)$$

Общее решение есть комбинация частных решений  $\lambda_i^k x_i$ , и коэффициенты  $c_i$ , соответствующие начальному условию  $u_0$ , равны  $c_1 \lambda_1^0 x_1 + \dots + c_n \lambda_n^0 x_n = u_0$ , или  $Sc = u_0$ , или  $c = S^{-1} u_0$ . (21)

На самом деле эти формулы дают два различных подхода к одному и тому же решению  $u_k = S \Lambda^k S^{-1} u_0$ . Первая формула (19) показывает, что матрица  $A^k$  совпадает с  $S \Lambda^k S^{-1}$ , и на этом можно остановиться. Но второй подход яснее показывает аналогию с решением дифференциального уравнения: вместо чисто экспоненциальных решений  $e^{\lambda_i t} x_i$  мы теперь имеем чистые степени  $\lambda_i^k x_i$ . «Гармоническими составляющими» вновь являются собственные векторы  $x_i$ , и на каждом шаге они усиливаются благодаря собственным значениям  $\lambda_i$ . Комбинируя эти частные решения так, чтобы получить вектор  $u_0$ , мы приходим к истинному решению  $u_k = S \Lambda^k S^{-1} u_0$ .

В каждом конкретном примере, как и в случае с уравнением Фибоначчи, первый шаг состоит в диагонализации матрицы  $A$ :

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad \det(A - \lambda I) = \lambda^2 - \lambda - 1, \\ \lambda_1 = \frac{1 + \sqrt{5}}{2}, \quad \lambda_2 = \frac{1 - \sqrt{5}}{2}, \\ A = S \Lambda S^{-1} = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} 1 & -\lambda_2 \\ -1 & \lambda_1 \end{bmatrix} \frac{1}{\lambda_1 - \lambda_2}.$$

Как только эти собственные значения и собственные векторы найдены, начинает работать формула (19). Начальные данные  $F_0 = 0$  и  $F_1 = 1$  дают  $u_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ , и тогда

$$\begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = u_k = A^k u_0 = S \Lambda^k S^{-1} u_0 = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1^k & \\ & \lambda_2^k \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \frac{1}{\lambda_1 - \lambda_2}.$$

Число Фибоначчи  $F_k$  равняется второй компоненте этого произведения:

$$F_k = \frac{\lambda_1^k}{\lambda_1 - \lambda_2} - \frac{\lambda_2^k}{\lambda_1 - \lambda_2} = \frac{1}{\sqrt{5}} \left[ \left( \frac{1 + \sqrt{5}}{2} \right)^k - \left( \frac{1 - \sqrt{5}}{2} \right)^k \right].$$

Это и есть искомый ответ. Он является несколько неожиданным, поскольку правило Фибоначчи  $F_{k+2} = F_{k+1} + F_k$  всегда должно давать целые числа, а мы получили дроби и квадратные корни. Они должны сократиться каким-то образом, чтобы осталось целое число. Фактически, поскольку второй член  $[(1 - \sqrt{5})/2]^k / \sqrt{5}$  всегда меньше чем  $1/2$ , он должен превращать первый член в ближайшее целое число. Вычитание оставляет только целую часть, и

$$F_{1000} \text{ равно ближайшему к } \frac{1}{\sqrt{5}} \left( \frac{1 + \sqrt{5}}{2} \right)^{1000} \text{ целому числу.}$$

Разумеется, это число очень велико, а число  $F_{1001}$  должно быть еще больше. Совершенно очевидно, что дробные части становятся пренебрежимо малыми в сравнении с целыми частями; отношение  $F_{1001}/F_{1000}$  должно быть очень близко к величине  $(1 + \sqrt{5})/2 \simeq 1,618$ , которая была названа греками «золотым сечением»<sup>1)</sup>. Другими словами, число  $\lambda_2^k$  становится пренебрежимо малым в сравнении с  $\lambda_1^k$  и отношение  $F_{k+1}/F_k$  стремится к  $\lambda_1^{k+1}/\lambda_1^k = \lambda_1$ .

**Упражнение 5.3.1.** Предположим, что Фибоначчи начинал свою последовательность с  $F_0 = 1$  и  $F_1 = 3$  и следовал затем тому же правилу  $F_{k+2} = F_{k+1} + F_k$ . Найти новый начальный вектор  $u_0$ , новые коэффициенты  $c = S^{-1}u_0$  и новые числа Фибоначчи. Показать, что отношения  $F_{k+1}/F_k$  по-прежнему стремятся к «золотому сечению».

**Упражнение 5.3.2.** Пусть каждое последующее число является *средним* между двумя предыдущими,  $G_{k+2} = \frac{1}{2}(G_{k+1} + G_k)$ . Найти матрицу  $A$  и привести ее к диагональному виду. Начиная с  $G_0 = 0$  и  $G_1 = 1/2$ , найти формулу для  $G_k$  и вычислить предел при  $k \rightarrow \infty$

**Упражнение 5.3.3.** Бернаделли рассматривал виды жуков, «которые живут только три года и размножаются на третий год». Если для удобства рассматривать только самок и при этом первая возрастная группа выживает с вероятностью  $1/2$ , вторая с вероятностью  $1/3$  и затем третья дает шесть самок на каждую самку из этой группы, то соответствующая матрица будет иметь вид

$$A = \begin{bmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix}.$$

Показать, что  $A^3 = I$  и проследить за распространением жуков за шесть лет, начиная с 3000 жуков в каждой возрастной группе.

<sup>1)</sup> Стороны наиболее красивых прямоугольников относятся друг к другу как 1.618 к 1.

## МАРКОВСКИЙ ПРОЦЕСС

В первой главе мы рассматривали пример о въезде и выезде из Калифорнии. Он заслуживает более детального рассмотрения. Мы руководствовались следующими двумя правилами:

*каждый год 1/10 людей, живших вне Калифорнии, въезжают в нее, а 2/10 калифорнийцев выезжают.*

Эта задача сводится к разностному уравнению: мы начинаем с  $y_0$  людей вне и  $z_0$  людей внутри и в конце первого года имеем

$$\begin{aligned} \frac{9}{10}y_0 + \frac{2}{10}z_0 &= y_1 \quad \text{людей вне,} \\ \frac{1}{10}y_0 + \frac{8}{10}z_0 &= z_1 \quad \text{людей внутри,} \end{aligned}$$

или

$$\begin{bmatrix} y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} y_0 \\ z_0 \end{bmatrix}.$$

Разумеется, эта задача взята «с потолка», но она имеет два существенных свойства того, что называют **марковским процессом**: общее число людей остается неизменным и числа людей вне и внутри никогда не становятся отрицательными<sup>1)</sup>. Первое свойство отражается в том факте, что *сумма элементов каждого столбца матрицы равна 1*: все учитываются и нет ни лишних, ни потерянных. Второе свойство отражено в том, что *матрица не имеет отрицательных элементов*: если начальные значения  $y_0$  и  $z_0$  были неотрицательны, то неотрицательными будут  $y_1$  и  $z_1$ ,  $y_2$  и  $z_2$  и т. д. Все степени  $A^k$  неотрицательны.

Мы собираемся сначала решить это разностное уравнение (используя формулу  $S\Lambda^k S^{-1}u_0$ ), затем посмотреть, приходит ли население постепенно к стационарному состоянию, и, наконец, обсудить марковские процессы в целом. Начнем с того, что матрица  $A$  должна быть диагонализирована:

$$\begin{aligned} A &= \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix}, & \det(A - \lambda I) &= \lambda^2 - 1.7\lambda + 0.7, \\ & & \lambda_1 &= 1 \quad \text{и} \quad \lambda_2 = 0.7, \\ A &= SAS^{-1} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & \\ & 0.7 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}. \end{aligned}$$

<sup>1)</sup> Более того, история полностью игнорируется; каждая новая ситуация  $u_{k+1}$  зависит только от текущей  $u_k$  и данные  $u_0, \dots, u_{k-1}$  могут быть выброшены. Возможно, даже наши жизни являют собой примеры марковских процессов, хотя я надеюсь, что это не так.

Теперь мы можем найти матрицу  $A^k$  и распределение через  $k$  лет:

$$\begin{aligned} \begin{bmatrix} y_k \\ z_k \end{bmatrix} &= A^k \begin{bmatrix} y_0 \\ z_0 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 1^k & \\ & 0.7^k \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} y_0 \\ z_0 \end{bmatrix} = \\ &= (y_0 + z_0) \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix} + (y_0 - 2z_0) (0.7)^k \begin{bmatrix} \frac{1}{3} \\ -\frac{1}{3} \end{bmatrix}. \end{aligned}$$

Это и есть искомое решение, и легко заметить, что же произойдет в конце концов: множитель  $(0.7)^k$  становится крайне малым и решение сходится к пределу

$$\begin{bmatrix} y_\infty \\ z_\infty \end{bmatrix} = (y_0 + z_0) \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}.$$

Общее население по-прежнему равно  $y_0 + z_0$ , как и в начальный момент, но в пределе  $2/3$  этого населения будет находиться вне Калифорнии и  $1/3$  внутри. И это будет выполняться независимо от того, каково было начальное распределение. Можно заметить, что такое стационарное состояние представляет собой в точности то распределение, которое требовалось в упр. 1.3.3. Если год начинается с соотношения « $2/3$  вне и  $1/3$  внутри», то он закончится в точности таким же образом:

$$\begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}, \quad \text{или} \quad Au_\infty = u_\infty.$$

Стационарное состояние представляет собой собственный вектор матрицы  $A$ , соответствующий собственному значению  $\lambda = 1$ . Умножение на матрицу  $A$ , дающее переход от одного шага к другому, оставляет вектор  $u_\infty$  неизменным.

Приведенное выше описание марковского процесса является полностью детерминированным: население перемещается в фиксированной пропорции. Но если мы вместо этого будем рассматривать отдельного индивидуума, то правила его передвижения могут быть даны в вероятностной форме. Если индивидуум находится вне Калифорнии, то с вероятностью  $1/10$  он въедет в нее, а если он уже в ней, то с вероятностью  $2/10$  выедет. Его движение становится случайным процессом, и управляющая им матрица  $A$  называется *матрицей перехода*. Мы не знаем теперь точно, где он находится, но каждый год компоненты вектора  $u_k = A^k u_0$  определяют вероятность того, что он находится вне штата, и вероятность того, что он находится в нем. Сумма этих вероятностей

равна 1 (ведь должен же этот человек где-то быть), и они никогда не становятся отрицательными; это вновь приводит нас к двум фундаментальным свойствам матрицы перехода: сумма элементов каждого столбца матрицы равна 1 и каждый элемент удовлетворяет неравенству  $a_{ij} \geq 0$ .

Основной момент теории состоит в том, чтобы понять, почему число  $\lambda = 1$  всегда является собственным значением и почему соответствующий собственный вектор дает стационарное состояние. Первое объяснить легко: каждый столбец матрицы  $A - I$  имеет сумму элементов, равную  $1 - 1 = 0$ . Поэтому строки матрицы  $A - I$  при сложении дают нулевую строку, они линейно зависимы, матрица  $A - I$  вырождена и число  $\lambda_1 = 1$  является собственным значением для  $A$ . Кроме весьма частных случаев<sup>1)</sup>, вектор  $u_k$  будет постепенно приближаться к соответствующему собственному вектору. Это обеспечивается формулой  $u_k = c_1 \lambda_1^k x_1 + \dots + c_n \lambda_n^k x_n$ , в которой ни одно из собственных значений не может превосходить 1; в противном случае вероятности  $u_k$  будут расти, как числа Фибоначчи, что невозможно. Если все остальные собственные значения строго меньше 1, то первый член формулы будет доминирующим, а остальные числа  $\lambda_i^k$  будут быстро сходиться к нулю и  $u_k \rightarrow c_1 x_1 = u_\infty$ . Это стационарное состояние будет обеспечено, если матрица  $A$  не только неотрицательна, но строго положительна, т. е.  $a_{ij} > 0$ . В этом случае вектор  $c_1 x_1$  имеет только положительные компоненты, сумма их равна 1 и они являются предельными значениями вероятностей в марковском процессе.

**Упражнение 5.3.4.** Пусть имеется три главных центра грузовых перевозок. Каждый месяц половина грузовиков Бостона и Лос-Анджелеса едет в Чикаго, другая половина остается на месте, а чикагские грузовики распределяются поровну между Бостоном и Лос-Анджелесом. Выписать матрицу перехода  $A$  размера  $3 \times 3$  и найти стационарное состояние  $u_\infty$ , соответствующее собственному значению  $\lambda = 1$ .

**Упражнение 5.3.5.** Рассмотреть пример эпидемии, при которой каждый месяц половина здоровых людей заболевает, а четверть заболевших умирает. Найти стационарное состояние для соответствующего марковского процесса

$$\begin{bmatrix} d_{k+1} \\ s_{k+1} \\ w_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{4} & 0 \\ 0 & \frac{3}{4} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} d_k \\ s_k \\ w_k \end{bmatrix}.$$

<sup>1)</sup> Если каждый человек, находящийся вне, въезжает, и каждый, находящийся внутри, выезжает, то число проживающих в штате меняется каждый год и стационарное состояние невозможно. Матрица перехода имеет вид  $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ , и собственными значениями являются как  $-1$ , так и  $+1$ .

**Упражнение 5.3.6.** Пусть в отношении цвета глаз мужчины в данном поколении несут доминантные гены, гетерозиготны или несут рецессивные гены (т. е. либо имеют оба гена «кареглазости», либо один ген «кареглазости», а один «голубоглазости», либо оба гена «голубоглазости») с вероятностями  $d_0$ ,  $h_0$ ,  $r_0$ . (Я думаю, что в гетерозиготном случае оба глаза будут карими, но тем не менее оба отцовских гена могут быть унаследованы с одинаковой вероятностью.) Если же женщины предполагаются гетерозиготными<sup>1)</sup>, то наследуемый от матери ген с одинаковой вероятностью может оказаться геном «кареглазости» или геном «голубоглазости». Найти матрицу  $A$ , которая давала бы для сына из следующего поколения вероятности  $d_1$ ,  $h_1$ ,  $r_1$ :

$$\begin{bmatrix} d_1 \\ h_1 \\ r_1 \end{bmatrix} = A \begin{bmatrix} d_0 \\ h_0 \\ r_0 \end{bmatrix}.$$

Каково будет распределение  $u_\infty$  после бесконечного числа поколений?

### УСТОЙЧИВОСТЬ

Имеется явное различие между числами Фибоначчи и марковскими процессами: числа  $F_k$  становятся все больше и больше, в то время как любая «вероятность» по определению находится между 0 и 1. Уравнение Фибоначчи является неустойчивым, равно как и уравнение сложных процентов  $P_{k+1} = 1.06P_k$ ; капитал растет вечно. Если бы марковские вероятности стремились к нулю, то соответствующее уравнение было бы устойчивым; на самом же деле это не так, поскольку на каждом этапе сумма вероятностей равна 1. Поэтому марковский процесс нейтрально устойчив.

Предположим теперь, что у нас имеется произвольное разностное уравнение  $u_{k+1} = Au_k$  и мы хотим исследовать его поведение при  $k \rightarrow \infty$ . Предполагая диагонализуемость матрицы  $A$ , получим решение  $u_k$  в виде комбинации «чистых решений»:

$$u_k = S\Lambda^k S^{-1}u_0 = c_1\lambda_1^k x_1 + \dots + c_n\lambda_n^k x_n.$$

Рост  $u_k$  управляется множителями  $\lambda_i^k$  и, следовательно, **устойчивость зависит от собственных значений матрицы  $A$ .**

**5Н.** Разностное уравнение  $u_{k+1} = Au_k$  устойчиво и  $u_k \rightarrow 0$ , если  $|\lambda_i| < 1$  для всех собственных значений, **нейтрально устойчиво** и вектор  $u_k$  ограничен при всех  $|\lambda_i| \leq 1$  и **неустойчиво** ( $u_k$  неограничен), если хотя бы одно собственное значение матрицы  $A$  удовлетворяет условию  $|\lambda_i| > 1$ .

**Пример 1.** Матрица

$$A = \begin{bmatrix} 0 & 4 \\ 0 & \frac{1}{2} \end{bmatrix},$$

<sup>1)</sup> На самом деле это неверно, но я надеюсь, что мне простят такое предположение.

безусловно, является устойчивой; ее собственные значения равны 0 и  $1/2$ , т. е. числам, расположенным на главной диагонали, поскольку эта матрица  $A$  является треугольной. Начинаясь с произвольного начального вектора  $u_0$  и удовлетворяя соотношению  $u_{k+1} = Au_k$ , решение должно постепенно сходиться к нулю:

$$u_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad u_1 = \begin{bmatrix} 4 \\ \frac{1}{2} \end{bmatrix}, \quad u_2 = \begin{bmatrix} 2 \\ \frac{1}{4} \end{bmatrix}, \quad u_3 = \begin{bmatrix} 1 \\ \frac{1}{8} \end{bmatrix}, \quad u_4 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{16} \end{bmatrix}, \quad \dots$$

Можно заметить, что большее собственное значение  $\lambda = 1/2$  управляет убыванием: после первого шага каждый последующий вектор  $u_k$  будет равен половине предыдущего. Эффект первого шага состоит в расщеплении вектора  $u_0$  на два собственных вектора матрицы  $A$ ,

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 1 \end{bmatrix} + \begin{bmatrix} -8 \\ 0 \end{bmatrix}$$

и аннулировании второго собственного вектора (соответствующего  $\lambda = 0$ ). Первый собственный вектор на каждом шаге умножается на  $\lambda = 1/2$ .

*Замечание.* Несмотря на сильную устойчивость матрицы  $A$ , обусловленную малостью собственных значений, длины векторов  $u_k$  в начале процесса растут. Только начиная с  $u_4$ , эти векторы становятся короче, чем исходный вектор  $u_0$ , и затем убывание происходит со скоростью геометрической прогрессии. Чтобы гарантировать *немедленное убывание*, т. е. выполнение неравенств  $\|u_1\| < \|u_0\|$  или  $\|Au\| < \|u\|$  для всех ненулевых начальных векторов  $u$ , нам требуется условие

$$\|Au\|^2 = u^T A^T A u < \|u\|^2 = u^T u, \quad \text{или} \quad u^T (A^T A - I) u < 0. \quad (22)$$

Это более сильное требование, чем условие  $|\lambda_i| \leq 1$  на собственные значения. В терминах гл. 6 оно означает, что матрица  $A^T A - I$  должна быть *отрицательно определенной* (т. е. все ее собственные значения отрицательны), а в терминах гл. 7 это требование означает, что «норма» матрицы  $A$  меньше 1.

**Упражнение 5.3.7.** Показать, что решение начинает убывать немедленно (т. е. матрица  $A^T A - I$  имеет отрицательные собственные значения), если внедиагональный элемент  $a_{12} = 4$  сделать меньше, чем  $a_{12} = \sqrt{3}/2$ .

**Упражнение 5.3.8.** Какие значения  $\alpha$  приводят к неустойчивости системы  $v_{n+1} = \alpha(v_n + w_n)$ ,  $w_{n+1} = \alpha(v_n + w_n)$ .

**Пример 2.** *Модель расширяющейся экономики фон Неймана.* Рассмотрим простую модель, в которой есть три «товара»: сталь, продукты питания и труд. На производство каждого товара затрачивается часть того, что было произведено годом ранее, и перед

экономистом стоит вопрос, может ли (и с какой скоростью) экономика расширяться. Пусть производство новой единицы стали требует 0.4 единицы существующей стали и 0.5 единицы труда, производство единицы продуктов питания требует 0.1 единицы продуктов питания и 0.7 единицы труда и производство (или содержание) единицы труда обходится в 0.8 единицы продуктов питания и по 0.1 единицы стали и труда. Тогда входные данные  $s_0$ ,  $f_0$  и  $l_0$  связаны с выходными данными  $s_1$ ,  $f_1$  и  $l_1$  уравнением

$$u_0 = \begin{bmatrix} s_0 \\ f_0 \\ l_0 \end{bmatrix} = \begin{bmatrix} 0.4 & 0 & 0.1 \\ 0 & 0.1 & 0.8 \\ 0.5 & 0.7 & 0.1 \end{bmatrix} \begin{bmatrix} s_1 \\ f_1 \\ l_1 \end{bmatrix} = Au_1.$$

Отметим, что это разностное уравнение «в обратную сторону»! Вместо соотношения  $u_1 = Au_0$  мы имеем  $u_0 = Au_1$ , так что расширением будет управлять не матрица  $A$ , а матрица  $A^{-1}$  (точнее, ее собственные значения). В этой задаче имеется и еще одна особенность: количества стали, продуктов питания и труда не могут становиться отрицательными. Фон Нейман искал максимальную скорость  $\alpha$ , с которой экономика может расширяться, *по-прежнему оставаясь неотрицательной*, т. е.  $u_1 \geq \alpha u_0 \geq 0$ .

Если эти неравенства выполняются (т. е. начиная с некоторого вектора стали, продуктов питания и труда  $u_0$ , мы заканчиваем по крайней мере вектором  $\alpha u_0$ ), то, начиная следующий год с  $u_1$ , мы получим по крайней мере  $\alpha u_1$  или  $\alpha^2 u_0$ . Следовательно, экономика может продолжать расширяться со скоростью  $\alpha$ . Разумеется, если наибольшее возможное  $\alpha$  будет меньше 1, то экономика фактически будет не расширяться, а сокращаться, а если  $A$  — «марковская» матрица, то мы будем иметь равновесие, при котором  $\alpha = 1$ . На самом деле теория фон Неймана очень близка к теории марковских процессов, поскольку в обоих случаях матрица  $A$  является неотрицательной, что гарантирует неотрицательность ее наибольшего собственного значения  $\lambda_1$  и компонент соответствующего собственного вектора. Для марковского случая  $\alpha = 1$ , поскольку  $\lambda_1 = 1$ ; в общем случае  $\lambda_1 = 1/\alpha$ , так как фон Нейман доказал, что наискорейшее возможное расширение всегда дается неотрицательным собственным вектором. В нашем примере  $\lambda_1 = 9/10$ , так что коэффициент расширения равен  $10/9$ , а собственный вектор есть

$$x = \begin{bmatrix} 0.2 \\ 1 \\ 1 \end{bmatrix} \quad \text{и} \quad Ax = \begin{bmatrix} 0.4 & 0 & 0.1 \\ 0 & 0.1 & 0.8 \\ 0.5 & 0.7 & 0.1 \end{bmatrix} \begin{bmatrix} 0.2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.18 \\ 0.9 \\ 0.9 \end{bmatrix} = 0.9x.$$

**Упражнение 5.3.9.** Объяснить в экономических терминах, почему увеличение любого элемента в матрице потребления  $A$  должно увеличивать ее наибольшее собственное значение  $\lambda_1$  (и замедлять расширение).

**Пример 3.** Матрица «затраты — выпуск» Леонтьева. В отличие от фон Неймана, Леонтьев рассматривал прежде всего производство и потребление в течение одного года; его система «затраты — выпуск» явилась одним из первых крупных успехов в математической экономике. Для ее иллюстрации рассмотрим ту же самую матрицу потребления  $A$  и будем выяснять, можно ли получить заданный вектор продукции  $y$ : можем ли мы закончить год с  $y_1$  единицами стали,  $y_2$  единицами продуктов питания и  $y_3$  единицами труда? Для этого нужно произвести товары в больших количествах  $x_1$ ,  $x_2$  и  $x_3$ , поскольку некоторая часть продукции поглощается в самом процессе производства. Фактически потребленное количество продукции равно  $Ax$ , и чистая продукция есть  $x - Ax$ .

**Задача.** Найти вектор  $x$ , такой, что

$$x - Ax = y, \quad \text{или} \quad x = (I - A)^{-1} y.$$

На первый взгляд нам нужно лишь выяснить, обратима ли матрица  $I - A$ . Однако и в этой задаче имеется особенность, связанная с неотрицательностью: мы предполагаем, что вектор заказов неотрицателен,  $y \geq 0$ , и требуем, чтобы вектор продукции также был неотрицательным, т. е.  $x \geq 0$ . Поэтому на самом деле вопрос состоит в том, будут ли элементы матрицы  $(I - A)^{-1}$  неотрицательными; если это так, то произведение  $(I - A)^{-1} y$  будет неотрицательным. При заданной матрице потребления  $A$  с наибольшим собственным значением  $\lambda_1$  основной математический результат формулируется следующим образом: матрица  $(I - A)^{-1}$  неотрицательна тогда и только тогда, когда  $\lambda_1 < 1$ . В этом случае экономика может производить товары в любом сочетании: при таких собственных значениях производство превосходит потребление.

**Пример.** Пусть  $A = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$ , так что на производство единицы стали затрачиваются две единицы продуктов питания, и наоборот. Тогда мы ничего не можем произвести! Наибольшее собственное значение равно  $\lambda_1 = 2$ , и все элементы матрицы  $(I - A)^{-1} = -\frac{1}{3} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$  отрицательны.

**Упражнение 5.3.10.** Почленным умножением проверить, что  $(I - A)(I + A + A^2 + \dots) = I$ . Этот бесконечный ряд представляет матрицу  $(I - A)^{-1}$ , и он является неотрицательным, если неотрицательна матрица  $A$  и если его сумма конечна; условием для этого служит неравенство  $\lambda_1 < 1$ . Просуммировать этот бесконечный ряд и показать, что сумма равна  $(I - A)^{-1}$ , для матрицы потребления

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Цель Леонтьева состояла в том, чтобы найти модель, которая использовала бы подлинные данные реальной экономики. Официальная американская матричная таблица межотраслевых связей за 1958 год содержала данные по 83 отраслям производства. Построенная теория выходит за рамки простого исследования матрицы  $(I - A)^{-1}$  и позволяет решать вопросы о естественных ценах рынка и об оптимизации. Труд при этом обычно рассматривается отдельно как наиболее важный товар, количество которого ограничено и должно быть минимизировано. Кроме того, экономика не всегда является линейной.

### § 5.4. ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ И ЭКСПОНЕНТА $e^{At}$

Когда мы встречаемся не с отдельным уравнением, а с системой, появляется возможность использовать теорию матриц. Это было справедливо для разностных уравнений, где решение  $u_k = A^k u_0$  зависело от степеней матрицы  $A$ . Это в равной мере справедливо и для дифференциальных уравнений, где решение  $u(t) = e^{At} u_0$  связано с экспонентой от матрицы  $A$ . Для того чтобы определить эту экспоненту и понять заложенный в ней смысл, рассмотрим пример

$$\frac{du}{dt} = Au = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} u. \quad (23)$$

Первый шаг, как обычно, состоит в отыскании собственных значений и собственных векторов:

$$A \begin{bmatrix} 1 \\ 1 \end{bmatrix} = (-1) \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad A \begin{bmatrix} 1 \\ -1 \end{bmatrix} = (-3) \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Затем перед нами открывается несколько возможностей, и все они приводят к одному и тому же ответу. Наилучший способ, вероятно, выписать общее решение и согласовать его с начальным вектором  $u_0$  при  $t = 0$ :

$$u(t) = c_1 e^{\lambda_1 t} x_1 + c_2 e^{\lambda_2 t} x_2 = c_1 e^{-t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 e^{-3t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad (24)$$

$$u_0 = c_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}.$$

Вы узнаете матрицу собственных векторов  $S$ ? И коэффициенты  $\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = S^{-1} u_0$  те же, которые получались для разностных урав-

нений. Подставляя их в (24), находим решение задачи. В матричной записи это решение выглядит так:

$$u(t) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} e^{-t} & \\ & e^{-3t} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = S \begin{bmatrix} e^{-t} & \\ & e^{-3t} \end{bmatrix} S^{-1}u_0.$$

Вот основная формула этого параграфа:  $Se^{\Lambda t}S^{-1}u_0$  является решением дифференциального уравнения, так же как  $S\Lambda^k S^{-1}u_0$  было решением разностного уравнения. Основную роль здесь играют матрицы

$$\Lambda = \begin{bmatrix} -1 & \\ & -3 \end{bmatrix} \quad \text{и} \quad e^{\Lambda t} = \begin{bmatrix} e^{-t} & \\ & e^{-3t} \end{bmatrix}.$$

Теперь можно позволить себе замедлить темп изложения.

Остается сделать два замечания относительно этого примера. Одно из них относится к завершению математической части теории, для чего требуется дать прямое определение экспоненты  $e^{\Lambda t}$  и связать его с формулой  $Se^{\Lambda t}S^{-1}$ . Второе состоит в том, чтобы дать физическую интерпретацию самого уравнения и его решения. Это дифференциальное уравнение имеет очень полезные применения.

Рассмотрим сначала экспоненту. Наиболее естественно определить ее по аналогии со степенным рядом для  $e^x$ :

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

и

$$e^{At} = I + At + \frac{(At)^2}{2!} + \frac{(At)^3}{3!} + \dots$$

$e^{At}$  есть матрица порядка  $n$  и ее легко можно продифференцировать:

$$\begin{aligned} \frac{d}{dt}(e^{At}) &= A + \frac{A^2(2t)}{2!} + \frac{A^3(3t^2)}{3!} + \dots = \\ &= A \left( I + At + \frac{(At)^2}{2!} + \dots \right) = Ae^{At}. \end{aligned}$$

Таким образом, подтверждается, что  $e^{At}u_0$  является решением рассматриваемого дифференциального уравнения: при  $t=0$  оно равно начальному вектору  $u_0$  и оно удовлетворяет уравнению  $d(e^{At}u_0)/dt = Ae^{At}u_0$ . Следовательно, это решение должно совпадать с другим решением  $Se^{\Lambda t}S^{-1}u_0$ . Чтобы дать более прямое доказательство этого совпадения, вспомним, что степени матрицы  $A = SAS^{-1}$  сворачиваются в  $A^k = (S\Lambda S^{-1}) \dots (S\Lambda S^{-1}) = S\Lambda^k S^{-1}$ .

Поэтому бесконечный ряд для экспоненты принимает вид

$$e^{At} = I + S\Lambda S^{-1}t + \frac{S\Lambda^2 S^{-1}t^2}{2!} + \frac{S\Lambda^3 S^{-1}t^3}{3!} + \dots = \\ = S \left( I + \Lambda t + \frac{(\Lambda t)^2}{2!} + \frac{(\Lambda t)^3}{3!} + \dots \right) S^{-1} = S e^{\Lambda t} S^{-1}.$$

На этом математическая часть заканчивается, и полученные результаты можно сформулировать следующим образом:

51. Если матрица  $A$  приводится к диагональному виду,  $A = S\Lambda S^{-1}$ , то дифференциальное уравнение  $du/dt = Au$  имеет решение

$$u(t) = e^{At}u_0 = S e^{\Lambda t} S^{-1}u_0. \quad (25)$$

Столбцы матрицы  $S$  являются собственными векторами матрицы  $A$ , так что

$$u(t) = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} e^{\lambda_1 t} \\ \vdots \\ e^{\lambda_n t} \end{bmatrix} S^{-1}u_0 = \\ = c_1 e^{\lambda_1 t} x_1 + \dots + c_n e^{\lambda_n t} x_n \quad (26)$$

Общее решение является линейной комбинацией частных решений  $e^{\lambda_i t} x_i$ , и коэффициенты  $c_i$ , удовлетворяющие начальному условию  $u_0$ , равны  $c = S^{-1}u_0$ .

Здесь наблюдается полная аналогия с разностными уравнениями — вы можете сравнить полученный результат с 5G на стр. 229. В обеих ситуациях мы предполагали, что матрица  $A$  диагонализуема. В противном случае она имела бы меньше чем  $n$  собственных векторов, и мы не нашли бы достаточного числа частных решений. Недостающие решения существуют, но они имеют более сложный вид, чем чистые экспоненты  $e^{\lambda t} x$ : они содержат «обобщенные собственные векторы» и множители типа  $te^{\lambda t}$ . Тем не менее формула  $u(t) = e^{At}u_0$  остается справедливой<sup>1)</sup>.

<sup>1)</sup> Этот «дефектный» случай будет рассмотрен в приложении В, однако пример можно привести и сейчас: если  $y' = z$  и  $z' = 0$ , то система имеет вид

$$\frac{du}{dt} = Au = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} u,$$

так что

$$A^2 = 0 \quad \text{и} \quad e^{At} = I + At = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}.$$

Обсудим теперь физический смысл этого примера, который легко объясним и в то же время по-настоящему важен. Рассматриваемое дифференциальное уравнение описывает процесс диффузии, который можно наглядно представить себе, если рассмотреть бесконечную трубку, разделенную на четыре части, из которых две средние конечные, а две по краям полубесконечные (рис. 5.1).

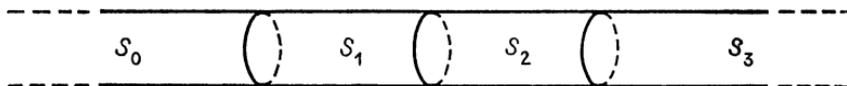


Рис. 5.1. Модель диффузии

В момент времени  $t=0$  две конечные части содержат некоторый химический раствор с концентрациями  $v_0$  и  $w_0$ . В это же время и в дальнейшем концентрация в двух бесконечных частях трубки равна нулю. При бесконечности объема такая модель будет давать верную картину для описания средней концентрации в этих бесконечных частях трубки даже после начала процесса диффузии. Диффузия начинается в момент времени  $t=0$  и подчиняется следующему закону: в любой момент времени  $t$  скорость диффузии между двумя прилегающими частями равняется разности концентраций. Мы считаем, что внутри каждой из частей концентрация во всех точках остается одинаковой. Процесс непрерывен по времени, но дискретен по пространству, и двумя неизвестными являются  $v(t)$  и  $w(t)$  в двух внутренних частях  $S_1$  и  $S_2$ .

Концентрация  $v$  изменяется двумя путями — посредством диффузии в крайний левый участок  $S_0$  и посредством диффузии в  $S_2$  или из  $S_2$ . Общая скорость изменения поэтому равна

$$\frac{dv}{dt} = (w - v) + (0 - v),$$

поскольку концентрация в  $S_0$  тождественно равна нулю. Аналогично

$$\frac{dw}{dt} = (0 - w) + (v - w).$$

Следовательно, система в точности совпадает с нашим примером (23):

$$u = \begin{bmatrix} v \\ w \end{bmatrix}, \quad \frac{du}{dt} = \begin{bmatrix} -2v + w \\ v - 2w \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} u.$$

Собственные значения  $-1$  и  $-3$  будут определять поведение решения. Они дают скорость убывания концентрации, и  $\lambda_1$  является более важным, поскольку лишь в исключительных слу-

чаях выбранные начальные условия могут привести к «сверхубыванию» со скоростью  $e^{-3t}$ . Фактически эти условия должны получаться из собственного вектора  $(1, -1)$ , компоненты которого имеют разные знаки. Если эксперимент допускает только неотрицательные концентрации, сверхубывание невозможно и предельная скорость должна равняться  $e^{-t}$ . Решение, убывающее с этой скоростью, соответствует собственному вектору  $(1, 1)$ , и, следовательно, две концентрации будут становиться почти равными при  $t \rightarrow \infty$ .

Еще одно замечание по поводу этого примера: он является дискретной аппроксимацией лишь с двумя неизвестными для непрерывного диффузионного процесса, описываемого уравнением в частных производных

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \quad u(0) = u(1) = 0.$$

Мы приходим к этому уравнению, если оставить два бесконечных участка с нулевой концентрацией и делить среднюю часть на отрезки все меньшей и меньшей длины  $h = 1/N$ . Дискретная система с  $N$  неизвестными подчиняется уравнению

$$\frac{d}{dt} \begin{bmatrix} u_1 \\ \cdot \\ \cdot \\ \cdot \\ u_N \end{bmatrix} = \begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & & & \\ & \cdot & \cdot & & 1 \\ & & & & \\ & & & 1 & -2 \end{bmatrix} \begin{bmatrix} u_1 \\ \cdot \\ \cdot \\ \cdot \\ u_N \end{bmatrix} = Au.$$

Это в точности совпадает с матрицей конечных разностей, построенной в § 1.6 для аппроксимации второй производной  $d^2/dx^2$ . Более внимательный взгляд на процесс дает масштабирующий множитель  $1/h^2$ , так что мы действительно приходим в пределе при  $h \rightarrow 0$  и  $N \rightarrow \infty$  к уравнению в частных производных  $du/dt = d^2u/dx^2$ . Оно известно как *уравнение теплопроводности*. Его решения по-прежнему могут быть разложены на гармонические составляющие задачи, но они уже оказываются не собственными векторами из  $N$  компонент, а *собственными функциями*. Фактически это в точности функции  $\sin n\pi x$ , и общее решение уравнения теплопроводности имеет вид

$$u(t) = \sum_{n=1}^{\infty} c_n e^{-n^2\pi^2 t} \sin n\pi x.$$

Коэффициенты  $c_n$  определяются, как обычно, из начальных условий, а скорости убывания даются собственными значениями:  $\lambda_n = -n^2\pi^2$ .

**Упражнение 5.4.1.** Вычислить матрицу  $e^{At}$  размера  $2 \times 2$  в рассмотренном выше примере и показать, что ее элементы положительны при  $t > 0$ . Любой

эксперимент, начинающийся при положительных концентрациях, таким и будет оставаться: если  $u_0 > 0$ , то  $e^{At}u_0 > 0^1$ ).

**Упражнение 5.4.2.** Предположим, что в уравнении диффузии направление времени изменено на обратное:  $du/dt = Au$  переходит в  $du/d(-t) = Au$  или

$$\frac{du}{dt} = Bu = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} u_0, \quad u_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Вычислить  $u(t)$  и показать, что решение теперь растет, а не убывает при  $t \rightarrow +\infty$ . (В непрерывном случае это увеличение происходит мгновенно, как только мы отходим от  $t=0$ , т. е. *уравнение теплопроводности необратимо во времени.*)

**Упражнение 5.4.3.** Для получения непрерывного марковского процесса уберем бесконечные участки  $S_0$  и  $S_2$  и перейдем к уравнениям

$$\begin{aligned} \frac{dv}{dt} &= w - v, \\ \frac{dw}{dt} &= v - w, \end{aligned} \quad \text{или} \quad \frac{du}{dt} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} u = Au.$$

Найти общее решение, решение, удовлетворяющее начальным концентрациям  $v=3$  и  $w=1$ , и стационарное состояние при  $t \rightarrow \infty$ . Заметить, что значению  $\lambda=0$  в дифференциальном уравнении соответствует  $\lambda=1$  в разностном уравнении ( $e^{0t}=1$ ), так что соответствующий собственный вектор определяет стационарное состояние.

**Упражнение 5.4.4.** Вывести формулу  $u(t) = Se^{At}S^{-1}u_0$ , используя замену переменных  $v = S^{-1}u$  в дифференциальном уравнении  $du/dt = Au$ . Найти новый вектор  $v_0$ , решить уравнение относительно  $v$  и вернуться к  $u$ .

**Упражнение 5.4.5.** Предполагая диагонализуемость матрицы  $A$ , показать из (25), что  $e^{tA}e^{sA} = e^{(t+s)A}$ . В качестве контрпримера показать для матриц размера  $2 \times 2$ , что в общем случае  $e^B e^C \neq e^{B+C}$ , т. е. правило для скаляров в матричном случае нарушается.

### УСТОЙЧИВОСТЬ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Так же как и для разностных уравнений, поведение решения  $u(t)$  при  $t \rightarrow \infty$  определяется собственными значениями. Если матрица  $A$  диагонализуема, то дифференциальное уравнение будет иметь  $n$  чисто экспоненциальных решений и любое частное решение  $u(t)$  есть некоторая их комбинация:

$$u(t) = Se^{At}S^{-1}u_0 = c_1 e^{\lambda_1 t} x_1 + \dots + c_n e^{\lambda_n t} x_n.$$

Устойчивость определяется множителями  $e^{\lambda t}$ . Если они стремятся к нулю, то  $u(t)$  стремится к нулю; если они остаются ограниченными, то  $u(t)$  остается ограниченным; а если один из этих множителей растет, то решение  $u(t)$  будет тоже расти, за

<sup>1)</sup> Это всегда справедливо при положительных внедиагональных элементах матрицы  $A$ .

исключением случаев с весьма специальными начальными условиями. Более того, поскольку величина (или модуль)  $e^{\lambda t}$  зависит только от вещественной части  $\lambda$ , *устойчивость определяется только вещественными частями*: если  $\lambda = \alpha + i\beta$ , то

$$e^{\lambda t} = e^{\alpha t} e^{i\beta t} = e^{\alpha t} (\cos \beta t + i \sin \beta t) \text{ и } |e^{\lambda t}| = e^{\alpha t}.$$

Эта величина убывает при  $\alpha < 0$ , является постоянной при  $\alpha = 0$  и возрастает при  $\alpha > 0$ , в то время как мнимые части  $\beta$  дают чистые колебания. Таким образом, для любой диагонализуемой матрицы  $A$  справедливо следующее утверждение:

**5J.** Дифференциальное уравнение  $du/dt = Au$  *устойчиво* и  $e^{At} \rightarrow 0$ , если все  $\operatorname{Re} \lambda_i < 0$ ; оно *нейтрально устойчиво* и экспонента  $e^{At}$  ограничена, если все  $\operatorname{Re} \lambda_i \leq 0$ , и оно *неустойчиво* и  $e^{At}$  неограниченна, когда  $\operatorname{Re} \lambda_i > 0$  хотя бы для одного собственного значения  $\lambda_i$  матрицы  $A$ .

**Пример 1.**

$$\frac{du}{dt} = Au = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} u, \quad u_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Собственные значения матрицы  $A$  удовлетворяют уравнению

$$|A - \lambda I| = \begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 + 1 = 0, \quad \text{т. е. } \lambda = \pm i.$$

Они являются чисто мнимыми, так что решение должно быть нейтрально устойчиво. Фактически  $e^{At}$  есть матрица вращения, и каждая компонента решения является простым гармоническим движением:

$$e^{At} = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix} \quad \text{и} \quad u(t) = e^{At} u_0 = \begin{bmatrix} \cos t \\ \sin t \end{bmatrix}. \quad (27)$$

Уравнение описывает движение точки по окружности<sup>1)</sup>.

**Пример 2.** Уравнение диффузии устойчиво, поскольку  $\lambda = -1$  и  $\lambda = -3$ .

**Пример 3.** Непрерывный марковский процесс, описанный в предыдущих упражнениях, лишь нейтрально устойчив, поскольку собственные значения равны  $\lambda = 0$  и  $\lambda = -2$ . Это уравнение диффузии с изолированными концами: ничто не покидает систему и не переходит во внешние бесконечные участки.

<sup>1)</sup> Все экспоненты от кососимметрических матриц являются ортогональными матрицами.

**Пример 4.** В ядерной технике реактор называется *критическим*, если он нейтрально устойчив: расщепление уравнивается поглощением. Более медленное расщепление делает его устойчивым, или *субкритическим*, и постепенно он останавливается. При неустойчивом расщеплении реактор превращается в бомбу.

Так же как и для разностных уравнений, решение задачи может быть устойчивым и при этом допускать кратковременное увеличение нормы  $\|u(t)\|$ , прежде чем начнется убывание. Мне кажется, что этот временный рост нетипичен для реальных задач; убывание, как правило, начинается непосредственно в момент времени  $t=0$ . Записав производную от скалярного произведения

$$\frac{d}{dt} x^T y = \left[ \frac{dx}{dt} \right]^T y + x^T \left[ \frac{dy}{dt} \right], \quad (28)$$

мы можем положить  $x=y=u$  и найти величину, знак которой определяет рост или убывание:

$$\frac{d}{dt} \|u(t)\|^2 = \left[ \frac{du}{dt} \right]^T u + u^T \left[ \frac{du}{dt} \right] = u^T (A^T + A) u. \quad (29)$$

**5К.** Если  $u^T (A^T + A) u < 0$  для всех ненулевых векторов  $u$ , что означает отрицательность всех собственных значений матрицы  $A^T + A$ , то норма решения убывает в каждый момент времени. Если  $A^T + A = 0$ , то длина  $\|u\|^2$  остается постоянной с течением времени, рассеяния энергии не происходит и система консервативна.

Пример 1 был консервативным: если точка движется по окружности, то длина  $\|u\|$ , очевидно, остается постоянной.

Для отыскания условия, которое одновременно являлось бы необходимым и достаточным для устойчивости, другими словами, условия, эквивалентного  $\operatorname{Re} \lambda_i < 0$ , есть две возможности. Одна из них заключается в обращении к Раусу и Гурвицу, которые в девятнадцатом веке получили ряд неравенств на элементы  $a_{ij}$ . Я не думаю, чтобы этот подход был достаточно хорош для матриц большого размера; ЭВМ, вероятно, может быстрее найти все собственные значения матрицы, чем проверить эти неравенства. Другая возможность была открыта Ляпуновым в 1897 году, через два года после работы Гурвица. Она состоит в отыскании матрицы весов  $W$ , такой, чтобы взвешенная длина  $\|Wu(t)\|$  всегда убывала. Если такая матрица  $W$  существует, то решение уравнения  $u' = Au$  должно быть устойчивым. Норма  $\|Wu\|$  будет монотонно убывать к нулю, и после нескольких подъемов и спусков решение  $u$  должно вести себя так же. Настоящую ценность метод Ляпунова приобретает в нелинейном случае, когда уравнение зачастую решить невозможно, но убывание  $\|Wu(t)\|$  по-прежнему

сохраняется, так что устойчивость может быть доказана даже при отсутствии формулы для  $u(t)$ .

**Упражнение 5.4.6.** Проверить, что формула (28) дает производную от скалярного произведения

$$x^T y = x_1(t) y_1(t) + \dots + x_n(t) y_n(t)$$

**Упражнение 5.4.7.** Исследовать на устойчивость уравнение  $du/dt = Au$ , рассматривая собственные значения матрицы  $A = \begin{bmatrix} 0 & 4 \\ -1 & 0 \end{bmatrix}$

**Упражнение 5.4.8.** С помощью собственных значений матриц

$$A = \begin{bmatrix} -1 & 3 \\ 0 & -1 \end{bmatrix} \quad \text{и} \quad A + A^T = \begin{bmatrix} -2 & 3 \\ 3 & -2 \end{bmatrix}$$

определить, должны ли решения уравнения  $du/dt = Au$  убывать по норме с момента времени  $t=0$  и должны ли они убывать при  $t \rightarrow \infty$ .

**Упражнение 5.4.9.** Определить, устойчива или неустойчива система  $dv/dt = w$ ,  $dw/dt = v$ .

#### УРАВНЕНИЯ ВТОРОГО ПОРЯДКА

Законы диффузии приводят к системе уравнений первого порядка  $du/dt = Au$ . То же самое происходит и в большом числе других приложений — в химии, в биологии и т. д., кроме наиболее важных законов физики. Это, например, второй закон Ньютона  $F = ma$ , где ускорение  $a$  есть вторая производная. Инерциальные члены дают уравнения второго порядка (мы должны решать систему  $d^2u/dt^2 = Au$  вместо  $du/dt = Au$ ), и наша цель состоит в том, чтобы понять, как этот переход ко вторым производным изменяет поведение решения<sup>1)</sup>.

Сравнение будет наилучшим, если мы рассмотрим ту же самую матрицу  $A$ :

$$\frac{d^2u}{dt^2} = Au = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} u. \quad (30)$$

Для того чтобы система начала «работать», должны быть заданы два начальных условия — «смещение»  $u = u_0$  и «скорость»  $du/dt = u'_0$ . Чтобы удовлетворить этим условиям, для системы из  $n$  уравнений потребуется уже не  $n$ , а  $2n$  чисто экспоненциальных решений.

Заменим букву  $\lambda$  буквой  $\omega$  и будем записывать частные решения в виде  $u = e^{i\omega t} x$ . Подставляя эту экспоненту в дифферен-

<sup>1)</sup> Четвертые производные также возможны (при изгибе балок), но природа, по-видимому, препятствует появлению производных выше четвертого порядка.

циальное уравнение, получаем условие, которому она должна удовлетворять:

$$\frac{d^2}{dt^2} (e^{i\omega t} x) = A (e^{i\omega t} x), \quad \text{или} \quad -\omega^2 x = Ax. \quad (31)$$

Вектор  $x$  должен быть собственным вектором матрицы  $A$  в точности как и раньше. Соответствующее собственное значение теперь равняется  $-\omega^2$ , так что частота  $\omega$  связана со скоростью убывания  $\lambda$  соотношением  $-\omega^2 = \lambda$ . Каждое частное решение  $e^{\lambda t} x$  уравнения первого порядка дает два частных решения  $e^{i\omega t} x$  уравнения второго порядка, и два их показателя равны  $\omega = \pm \sqrt{-\lambda}$ . Это не так лишь при  $\lambda = 0$ , когда имеется только одно значение корня; если соответствующий этому  $\lambda$  собственный вектор есть  $x$ , то два частных решения будут равны  $x$  и  $tx$ .

Для нашей матрицы диффузии все собственные значения  $\lambda$  отрицательны, и потому все частоты  $\omega$  вещественны: чистая диффузия преобразуется в чистое колебание. Множители  $e^{i\omega t}$  дают нейтральную устойчивость, решение не растет и не затухает и полная энергия остается в точности постоянной: она все время перераспределяется внутри системы. Общее решение уравнения  $d^2u/dt^2 = Au$ , если матрица  $A$  имеет отрицательные собственные значения  $\lambda_1, \dots, \lambda_n$  и если  $\omega_j = \sqrt{-\lambda_j}$ , равно

$$u(t) = (c_1 e^{i\omega_1 t} + d_1 e^{-i\omega_1 t}) x_1 + \dots + (c_n e^{i\omega_n t} + d_n e^{-i\omega_n t}) x_n. \quad (32)$$

Как всегда, произвольные постоянные находятся из начальных условий. Это легче сделать (ценой дополнительной формулы), перейдя от «осциллирующих экспонент» к более привычным синусу и косинусу:

$$u(t) = (a_1 \cos \omega_1 t + b_1 \sin \omega_1 t) x_1 + \dots + (a_n \cos \omega_n t + b_n \sin \omega_n t) x_n. \quad (33)$$

Теперь начальное смещение легко отличить от начальной скорости: из  $t=0$  следует, что  $\sin \omega t = 0$  и  $\cos \omega t = 1$ , в результате чего получаем

$$u_0 = a_1 x_1 + \dots + a_n x_n, \quad u'_0 = Sa, \quad a = S^{-1} u'_0.$$

Смещение согласуется с выбором коэффициентов  $a$ , а скорость — с выбором  $b$ : дифференцируя функцию  $u(t)$  и полагая  $t=0$ , получаем условие для определения  $b$  в виде

$$u'_0 = b_1 \omega_1 x_1 + \dots + b_n \omega_n x_n.$$

Подставляя это в (33), получаем искомое решение уравнения.

Мы хотим применить эти формулы к нашему примеру. Собственные значения равнялись  $\lambda_1 = -1$  и  $\lambda_2 = -3$ , так что частоты равны  $\omega_1 = 1$  и  $\omega_2 = \sqrt{3}$ . Если система начинает движение из состояния покоя (т. е. начальная скорость  $u'_0$  равна нулю), то

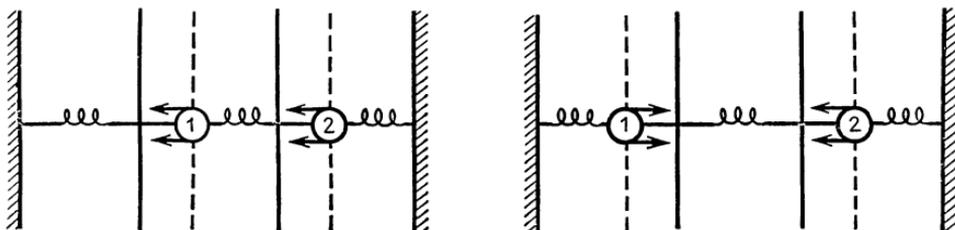
члены  $b \sin \omega t$  исчезнут. И если первый осциллятор получает единичное смещение, то требование  $u_0 = a_1 x_1 + a_2 x_2$  приводит к

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = a_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + a_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \text{т. е.} \quad a_1 = a_2 = \frac{1}{2}.$$

Следовательно, решение равно

$$u(t) = \frac{1}{2} \cos t \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{1}{2} \cos \sqrt{3}t \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Попытаемся дать физическое истолкование полученному решению. Система представляет собой две единичные массы, связанные друг с другом и с неподвижными стенками тремя одинаковыми пружинами (рис. 5.2). Первая масса передвигается в положение  $v_0 = 1$ , вторая масса зафиксирована на месте, и при  $t = 0$  мы их отпускаем. Их движение, описываемое функцией  $u(t)$ , будет средним между двумя чистыми колебаниями, соответствующими двум собственным векторам. На первой гармонике массы движутся в унисон и средняя пружина никогда не растягивается (рис. 5.2a). Частота  $\omega_1 = 1$  оказывается такой же, как и для од-



(a)  $\omega_1 = 1, x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

(b)  $\omega_2 = \sqrt{3}, x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

Рис. 5.2. Две гармоник.

ной пружины с одной массой. Для высокочастотной гармоники  $x_2 = (1, -1)$  с компонентами противоположных знаков и с частотой  $\sqrt{3}$  массы движутся в противоположных направлениях, но с равными скоростями (рис. 5.2b). Общее решение является линейной комбинацией этих двух гармоник, и наше частное решение равняется их полусумме.

Такое движение является, как мы говорим, «почти периодическим» во времени. Если отношение  $\omega_1/\omega_2$  было дробным, то две массы постепенно вернутся к состоянию  $v = 1$  и  $w = 0$  и весь процесс начнется сначала. Комбинация  $\sin 2t$  и  $\sin 3t$  будет иметь период  $2\pi$ . Но, поскольку число  $\sqrt{3}$  иррационально, мы можем

лишь сказать, что массы подойдут произвольно близко к начальному положению. Они также подойдут близко, если у нас хватит терпения дождаться, к противоположному положению  $v=0$  и  $w=1$ . Как и у бильярдного шара, бесконечно отскакивающего от идеально гладкого стола, полная энергия масс остается неизменной, и раньше или позже они подойдут сколь угодно близко к любому состоянию с той же энергией.

И вновь мы не можем не провести параллель с непрерывным случаем, когда вместо двух масс, или  $N$  масс, у нас имеется континуум. При переходе дискретных масс и пружин в сплошной стержень «вторые разности», задаваемые матричными коэффициентами 1,  $-2$ , 1, переходят во вторые производные. Этот предельный случай описывается знаменитым волновым уравнением  $\partial^2 u / \partial t^2 = \partial^2 u / \partial x^2$ .

**Упражнение 5.4.10.** Используя диагонализацию (стр. 218), решить уравнение

$$\frac{d^2 u}{dt^2} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{bmatrix} u \quad \text{при} \quad u_0 = \begin{bmatrix} 2 \\ -2 \\ 0 \end{bmatrix} \quad \text{и} \quad u'_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

**Упражнение 5.4.11.** Решить уравнение

$$\frac{d^2 u}{dt^2} = \begin{bmatrix} -5 & -1 \\ -1 & -5 \end{bmatrix} u \quad \text{при} \quad u_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{и} \quad u'_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

**Упражнение 5.4.12.** При наличии матрицы трения  $F$  две единичные массы будут колебаться по закону  $d^2 u / dt^2 = -F du / dt + Au$ . Подставить чистую экспоненту  $e^{\rho t} x$  в это уравнение и получить для  $\rho$  квадратичную задачу на собственные значения.

**Упражнение 5.4.13.** Решение задачи  $u'' = Au$  по-прежнему дается собственными векторами матрицы  $A$ , образующими столбцы матрицы  $S$ . Показать, что начальным условиям и самому этому дифференциальному уравнению удовлетворяет функция

$$u(t) = S \begin{bmatrix} \cos \omega_1 t & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \cos \omega_n t \end{bmatrix} S^{-1} u_0 + S \begin{bmatrix} (\sin \omega_1 t) / \omega_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & (\sin \omega_n t) / \omega_n \end{bmatrix} S^{-1} u'_0$$

(вместо  $u = S e^{At} S^{-1} u_0$ , как это было в случае  $u' = Au$ ).

**Упражнение 5.4.14.** Найти движение второй массы в рассмотренном перед упражнениями примере, если первая масса получает при  $t=0$  удар:

$$u_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad u'_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

## § 5.5. КОМПЛЕКСНЫЙ СЛУЧАЙ: ЭРМИТОВЫ И УНИТАРНЫЕ МАТРИЦЫ

В дальнейшем становится уже невозможным работать только с вещественными векторами и вещественными матрицами. В первой половине этой книги, когда основная задача состояла в решении системы  $Ax = b$ , было очевидно, что решение  $x$  является вещественным при вещественных  $A$  и  $b$ . Поэтому необходимость в комплексных числах не возникала; их можно было бы допустить, но это не дало бы ничего нового. Теперь же мы не можем обойтись без них. Коэффициенты характеристического многочлена вещественной матрицы вещественны, но собственные значения (как в случае, когда  $\lambda^2 + 1 = 0$ ) могут таковыми не быть.

Мы должны поэтому ввести пространство  $C^n$  векторов с  $n$  комплексными компонентами. Сложение и умножение в этом пространстве осуществляются по тем же правилам, что и ранее. Но формулу  $\|x\|^2 = x_1^2 + \dots + x_n^2$  для длины вектора нужно изменить, иначе вектор с компонентами  $(1, i)$  будет иметь нулевую длину:  $1^2 + i^2 = 0$ . Это изменение в способе вычисления длины вызывает целый ряд других изменений. Скалярное произведение двух векторов, транспонированная матрица, определения симметрических, кососимметрических и ортогональных матриц при введении комплексных чисел нужно модифицировать. Но в каждом случае новое определение будет совпадать со старым, когда векторы и матрицы вещественны.

Все эти изменения перечислены на стр. 267, и этот список служит хорошим словарем для перевода вещественным и комплексным случаем. Надеюсь, что он окажется полезным читателю. Этот список содержит также для каждого класса матриц наиболее полную информацию о расположении их собственных чисел. Нас в особенности интересуют симметрические матрицы: где расположены их собственные значения и чем характеризуются их собственные векторы? С точки зрения практики это наиболее важные вопросы в теории собственных значений и, следовательно, этот параграф является наиболее важным.

### КОМПЛЕКСНЫЕ ЧИСЛА И СОПРЯЖЕННЫЕ К НИМ

Вероятно, читатель уже знаком с комплексными числами, но, поскольку нам нужны лишь основные факты, стоит дать короткий их обзор<sup>1)</sup>. Каждому известно, что  $i$ , чем бы оно ни было, удовлетворяет уравнению  $i^2 = -1$ . Это чисто мнимое число, равно

<sup>1)</sup> Важными понятиями являются комплексная сопряженность и модуль.

как и  $ib$ , где  $b$  вещественно. Сумма вещественного и мнимого чисел есть комплексное число  $a + ib$ , и оно естественным образом изображается на комплексной плоскости (рис. 5.3).

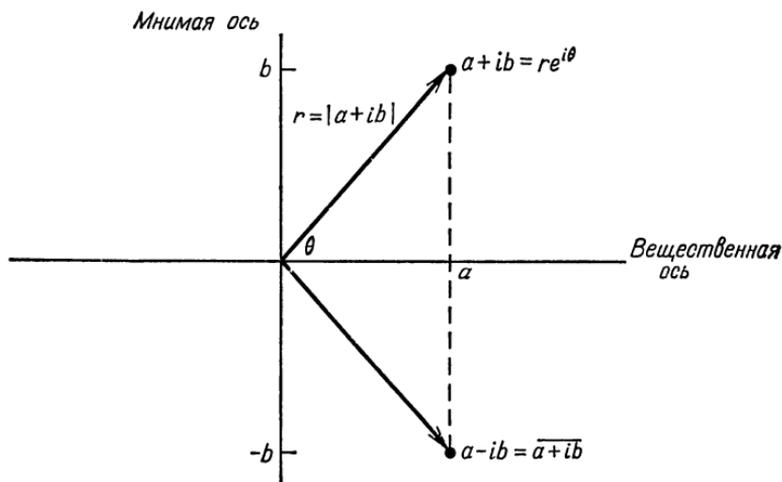


Рис. 5.3. Комплексная плоскость.

Вещественные числа (для которых  $b = 0$ ) и мнимые числа ( $a = 0$ ) являются частными случаями комплексных чисел — они лежат на одной или другой координатной оси. Два комплексных числа складываются по правилу

$$(a + ib) + (c + id) = (a + c) + i(b + d)$$

и перемножаются с использованием соотношения  $i^2 = -1$ :

$$(a + ib)(c + id) = ac + ibc + iad + i^2bd = (ac - bd) + i(bc + ad).$$

Комплексно-сопряженное к  $a + ib$  есть число  $a - ib$ , т. е. знак при мнимой части изменен на противоположный. Геометрически это зеркальное отражение относительно вещественной оси. Любое вещественное число совпадает со своим сопряженным. Сопряженность обозначается чертой,  $\overline{a + ib} = a - ib$ , и имеет три важных свойства:

(1) Сопряженное к произведению равняется произведению сопряженных:

$$\overline{(a + ib)(c + id)} = (ac - bd) - i(bc + ad) = \overline{(a + ib)} \overline{(c + id)}. \quad (34)$$

(2) Сопряженное к сумме равняется сумме сопряженных:

$$\overline{(a + c) + i(b + d)} = (a + c) - i(b + d) = \overline{(a + ib)} + \overline{(c + id)}.$$

(3) Умножение любого числа  $a + ib$  на сопряженное к нему  $a - ib$  дает вещественное число, равное квадрату гипотенузы на

рис. 5.3:

$$(a + ib)(a - ib) = a^2 + b^2 = r^2. \quad (35)$$

Это расстояние  $r$  называется *модулем* исходного числа  $a + ib$  и (как и абсолютное значение вещественного числа, которое также не бывает отрицательным) обозначается вертикальными черточками:  $|a + ib| = r = \sqrt{a^2 + b^2}$ .

Наконец, стороны связаны с гипотенузой тригонометрическими соотношениями

$$a = \sqrt{a^2 + b^2} \cos \theta, \quad b = \sqrt{a^2 + b^2} \sin \theta.$$

Комбинируя эти уравнения, переходим к полярным координатам:

$$a + ib = \sqrt{a^2 + b^2} (\cos \theta + i \sin \theta) = r e^{i\theta}. \quad (36)$$

Имеется важный частный случай, когда модуль  $r$  равняется единице. Тогда комплексное число есть  $e^{i\theta} = \cos \theta + i \sin \theta$  и оно находится на *единичной окружности* комплексной плоскости. При изменении  $\theta$  от 0 до  $2\pi$  это число  $e^{i\theta}$  движется вокруг начала координат на фиксированном расстоянии  $|e^{i\theta}| = \sqrt{\cos^2 \theta + \sin^2 \theta} = 1$ .

**Упражнение 5.5.1.** Для комплексных чисел  $3 + 4i$  и  $1 - i$

- (a) найти их положения на комплексной плоскости,
- (b) найти их сумму и произведение,
- (c) найти их модули и сопряженные.

Лежат они внутри или вне единичного круга?

**Упражнение 5.5.2.** Что можно сказать о

- (i) сумме комплексного числа и его сопряженного?
- (ii) сопряженном к числу, лежащему на единичной окружности?
- (iii) произведении двух чисел, лежащих на единичной окружности?
- (iv) сумме двух чисел, лежащих на единичной окружности?

**Упражнение 5.5.3.** Проверить, что при умножении двух чисел их модули перемножаются:

$$|(a + ib)(c + id)| = |a + ib| |c + id|, \quad |(r^{i\theta})(R e^{i\varphi})| = rR.$$

### ДЛИНЫ И ТРАНСПОНИРОВАНИЕ В КОМПЛЕКСНОМ СЛУЧАЕ

Мы возвращаемся к линейной алгебре и переходим от вещественных чисел к комплексным. Первый шаг состоит в допущении комплексных векторов и не вызывает затруднений: по определению *пространство  $C^n$  содержит все векторы  $x$  с  $n$  комплексными компонентами*

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad x_j = a_j + ib_j.$$

Векторы  $x$  и  $y$  по-прежнему складываются покомпонентно, но скалярное умножение теперь производится на комплексные числа. Как и ранее, векторы  $v_1, \dots, v_k$  линейно зависимы, если некоторая нетривиальная комбинация  $c_1 v_1 + \dots + c_k v_k$  дает нулевой вектор. Но коэффициенты  $c_j$  теперь могут быть комплексными. Единичные координатные векторы по-прежнему принадлежат  $\mathbb{C}^n$ , и они по-прежнему линейно независимы и образуют базис. Поэтому  $\mathbb{C}^n$  также является векторным пространством размерности  $n$ .

Мы уже указывали, что определение длины должно быть изменено, поскольку квадрат комплексного числа не обязательно положителен и формула  $\|x\|^2 = x_1^2 + \dots + x_n^2$  поэтому бесполезна. Новое определение совершенно естественно:  $x_j^2$  заменяется на модуль  $|x_j|^2$  и длина записывается так:

$$\|x\|^2 = |x_1|^2 + \dots + |x_n|^2. \quad (37)$$

В двумерном случае

$$\begin{aligned} x &= \begin{bmatrix} 1 \\ i \end{bmatrix} & \text{и} & \|x\|^2 = 2; \\ y &= \begin{bmatrix} 2+i \\ 2-4i \end{bmatrix} & \text{и} & \|y\|^2 = 25. \end{aligned}$$

Для вещественных векторов имела тесная связь между длиной и скалярным произведением:  $\|x\|^2 = x^T x$ . Мы хотим сохранить эту связь. Поэтому скалярное произведение должно быть модифицировано в соответствии с новым определением длины и стандартная модификация состоит в сопряжении первого вектора в скалярном произведении. Это означает, что вектор  $x$  заменяется на  $\bar{x}$  и скалярное произведение векторов  $x$  и  $y$  принимает вид

$$\bar{x}^T y = \bar{x}_1 y_1 + \dots + \bar{x}_n y_n. \quad (38)$$

Типичный пример в  $\mathbb{C}^2$ :

$$x = \begin{bmatrix} 1+i \\ 3i \end{bmatrix}, \quad y = \begin{bmatrix} 4 \\ 2-i \end{bmatrix},$$

$$\bar{x}^T y = (1-i)4 + (-3i)(2-i) = 1 - 10i.$$

И если рассмотреть скалярное произведение вектора  $x$  на себя, то мы вновь приходим к квадрату длины:

$$\bar{x}^T x = (\overline{1+i})(1+i) + (\overline{3i})(3i) = 2 + 9 = \|x\|^2.$$

Заметим, что  $\bar{y}^T x$  отлично от  $\bar{x}^T y$ ; в дальнейшем мы должны следить за порядком векторов в скалярном произведении. Имеется еще одно новшество: если вектор  $x$  заменяется на  $cx$ , то скалярное произведение векторов  $x$  и  $y$  умножается не на  $c$ , а на  $\bar{c}$ .

Остается сделать еще одно изменение. Это лишь изменение в обозначениях, которое сводит два символа в один: вместо черты для сопряжения и буквы  $T$  для транспонирования мы вводим сопряженное транспонирование и обозначаем его верхним индексом  $H$ . Поэтому  $\bar{x}^T = x^H$  и то же обозначение применяется к матрицам: **сопряженно транспонированная к  $A$**  есть

$$\bar{A}^T = A^H \quad \text{с элементами} \quad (A^H)_{ij} = \bar{A}_{ji}. \quad (39)$$

Если матрица  $A$  имеет размер  $m \times n$ , то  $A^H$  имеет размер  $n \times m$ . Например,

$$\begin{bmatrix} 2+i & 3i \\ 4-i & 5 \\ 0 & 0 \end{bmatrix}^H = \begin{bmatrix} 2-i & 4+i & 0 \\ -3i & 5 & 0 \end{bmatrix}.$$

Этот символ  $A^H$  служит официальным признанием того факта, что когда рассматриваются матрицы  $A$  с комплексными элементами, нам очень редко бывает нужна просто транспонированная к  $A$  матрица. Практически во всех случаях нам нужна сопряженно транспонированная (или *эрмитово транспонированная*) матрица<sup>1)</sup>. Суммируем все модификации, связанные с введением комплексных чисел:

- 5L. (i) Скалярное произведение векторов  $x$  и  $y$  равно  $x^H y$ , и эти векторы ортогональны, если  $x^H y = 0$ .  
 (ii) Длина вектора  $x$  есть  $\|x\| = (x^H x)^{1/2}$ .  
 (iii) Правило  $(AB)^T = B^T A^T$  после сопряжения всех элементов переходит в  $(AB)^H = B^H A^H$ .

**Упражнение 5.5.4.** Найти длины и скалярное произведение векторов

$$x = \begin{bmatrix} 2-4i \\ 4i \end{bmatrix} \quad \text{и} \quad y = \begin{bmatrix} 2+4i \\ 4 \end{bmatrix}.$$

**Упражнение 5.5.5.** Выписать матрицу  $A^H$  и вычислить  $C = A^H A$ , если

$$A = \begin{bmatrix} 1 & i & 0 \\ i & 0 & 1 \end{bmatrix}.$$

Каково соотношение между  $C$  и  $C^H$ ? Выполняется ли оно для любых векторов  $C$  вида  $A^H A$ ?

**Упражнение 5.5.6.** (i) Решить методом исключения уравнение  $Ax = 0$  с матрицей  $A$ , указанной выше.

<sup>1)</sup> Матрицу  $A^H$  иногда называют «эрмитовой к  $A$ ». К сожалению, здесь нужно быть внимательным, чтобы отличать это название от употребляемого в таком, например, контексте: « $A$  является эрмитовой», что означает выполнение равенства  $A = A^H$ .

(ii) Показать, что вычисленное нуль-пространство ортогонально к  $\mathfrak{R}(A^H)$ , но не к обычному пространству строк  $\mathfrak{R}(A^T)$ . Четырьмя фундаментальными пространствами в комплексном случае являются  $\mathfrak{R}(A)$  и  $\mathfrak{R}(A^H)$ , как и раньше, а затем  $\mathfrak{R}(A^H)$  и  $\mathfrak{R}(A^T)$ .

### ЭРМИТОВЫ МАТРИЦЫ

В предыдущих главах мы говорили о симметрических матрицах  $A$ ,  $A = A^T$ . Теперь, при наличии матриц с комплексными элементами, эта идея симметричности должна быть расширена. Правильным обобщением будут не матрицы, совпадающие со своими транспонированными, а **матрицы, совпадающие со своими сопряженно транспонированными**. Это эрмитовы матрицы, и типичным примером является

$$A = \begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix} = A^H. \quad (40)$$

Заметьте, что диагональные элементы должны быть вещественными, поскольку они не изменяются в процессе сопряжения. Каждый внедиагональный элемент соответствует своему зеркальному отражению относительно главной диагонали, и они являются комплексно-сопряженными друг с другом. В каждом случае  $a_{ij} = \overline{a_{ji}}$ , и на этом примере очень хорошо можно проиллюстрировать четыре основных свойства эрмитовых матриц.

Наша главная цель состоит в получении этих четырех свойств, и нужно вновь подчеркнуть, что они равным образом применимы и к вещественным симметрическим матрицам. Последние представляют собой частный и наиболее важный случай эрмитовых матриц.

|| **Свойство 1.** Если  $A = A^H$ , то для всех комплексных векторов  $x$  число  $x^H A x$  вещественно.

Любой элемент матрицы  $A$  вносит свой вклад в  $x^H A x$ :

$$\begin{aligned} x^H A x &= [\bar{u} \ \bar{v}] \begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \\ &= 2\bar{u}u + 5\bar{v}v + (3-3i)\bar{u}v + (3+3i)u\bar{v}. \end{aligned}$$

Каждый из «диагональных членов» является вещественным, поскольку  $2\bar{u}u = 2|u|^2$  и  $5\bar{v}v = 5|v|^2$ . Внедиагональные члены являются комплексно-сопряженными, так что они объединяются, давая удвоенную вещественную часть от  $(3-3i)\bar{u}v$ . Следовательно, все выражение  $x^H A x$  вещественно.

Для доказательства общего случая мы можем вычислить  $(x^H A x)^H$ . Мы должны получить сопряженную к матрице  $x^H A x$  размера  $1 \times 1$ , но фактически мы вновь получаем то же самое

число:  $(x^H Ax)^H = x^H A^H x^{HH} = x^H Ax$ , так что это число должно быть вещественным.

|| **Свойство 2.** Каждое собственное значение эрмитовой матрицы вещественно.

Доказательство. Пусть  $\lambda$  есть собственное значение, а  $x$  — соответствующий собственный вектор (ненулевой):  $Ax = \lambda x$ . Умножим это равенство на  $x^H$ :  $x^H Ax = \lambda x^H x$ . Левая часть вещественна по свойству 1, а в правой части  $x^H x = \|x\|^2$  вещественно и положительно, поскольку  $x \neq 0$ . Следовательно,  $\lambda$  должно быть вещественным. В нашем примере

$$\begin{aligned} |A - \lambda I| &= \begin{vmatrix} 2 - \lambda & 3 - 3i \\ 3 + 3i & 5 - \lambda \end{vmatrix} = \lambda^2 - 7\lambda + 10 - |3 - 3i|^2 = \\ &= \lambda^2 - 7\lambda - 8 = (\lambda - 8)(\lambda + 1). \end{aligned} \quad (41)$$

|| **Свойство 3.** Собственные векторы эрмитовой матрицы, соответствующие различным собственным значениям, взаимно ортогональны.

Доказательство вновь начинается с заданной информации  $Ax = \lambda x$ ,  $Ay = \mu y$  и  $\lambda \neq \mu$  и использует небольшой трюк. Сопряженно транспонированная к  $Ax = \lambda x$  обычно есть  $x^H A^H = \bar{\lambda} x^H$ , но, поскольку  $A = A^H$  и  $\lambda$  вещественно по свойству 2, это равняется  $x^H A = \lambda x^H$ . Умножая это равенство справа на  $y$ , а равенство  $Ay = \mu y$  слева на  $x^H$ , находим

$$x^H Ay = \lambda x^H y \quad \text{и} \quad x^H Ay = \mu x^H y.$$

Следовательно,  $\lambda x^H y = \mu x^H y$ , и, поскольку  $\lambda \neq \mu$ , заключаем, что  $x^H y = 0$ , т. е. вектор  $x$  ортогонален к  $y$ . В нашем примере с  $Ax = 8x$  и  $Ay = -y$  собственные векторы вычисляются обычным способом из

$$\begin{aligned} (A - 8I)x &= \begin{bmatrix} -6 & 3 - 3i \\ 3 + 3i & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, & x &= \begin{bmatrix} 1 \\ 1 + i \end{bmatrix}, \\ (A + I)y &= \begin{bmatrix} 3 & 3 - 3i \\ 3 + 3i & 6 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, & y &= \begin{bmatrix} 1 - i \\ -1 \end{bmatrix}. \end{aligned}$$

Эти два собственных вектора взаимно ортогональны:

$$x^H y = [1 \quad 1 - i] \begin{bmatrix} 1 - i \\ -1 \end{bmatrix} = 0.$$

Разумеется, любые кратные этим векторам,  $x/\alpha$  и  $y/\beta$ , равным образом будут собственными векторами. Пусть мы выбираем  $\alpha = \|x\|$  и  $\beta = \|y\|$ , так что  $x/\alpha$  и  $y/\beta$  являются единичными векто-

рами, т. е. собственные векторы нормализованы к единичной длине. Поскольку они уже были взаимно ортогональны, теперь они стали *ортонормированными*. Если записать их в виде столбцов матрицы  $S$ , то (как обычно, когда собственные векторы служат столбцами) мы получим равенство  $S^{-1}AS = \Lambda$ . *Столбцы диагонализующей матрицы ортонормированы*. Если исходная матрица  $A$  вещественная и симметрическая, то ее собственные значения и собственные векторы вещественны. Таким образом, нормализация собственных векторов превращает  $S$  в ортогональную матрицу. Вещественная симметрическая матрица может быть диагонализирована ортогональной матрицей  $Q$ .

В комплексном случае нам нужно новое название и новый символ для матрицы с ортонормированными столбцами: она называется *унитарной матрицей* и обозначается через  $U$ . По аналогии с вещественными ортогональными матрицами, для которых  $Q^T Q = I$  и потому  $Q^T = Q^{-1}$ , свойство ортонормированности столбцов переходит в

$$U^H U = \begin{bmatrix} \overline{x_1} \\ \overline{x_2} \\ \vdots \\ \overline{x_n} \end{bmatrix} \begin{bmatrix} | & | & | \\ x_1 & x_2 & x_n \\ | & | & | \end{bmatrix} = I, \quad \text{или} \quad U^H = U^{-1}. \quad (42)$$

Это приводит к последнему специальному свойству эрмитовых матриц:

**Свойство 4.** Если  $A = A^H$ , то существует диагонализующая матрица, которая является также и унитарной,  $S = U$ : ее столбцы ортонормированы и

$$U^{-1}AU = U^H AU = \Lambda. \quad (43)$$

Следовательно, любая эрмитова матрица может быть представлена в виде

$$A = U \Lambda U^H = \lambda_1 x_1 x_1^H + \lambda_2 x_2 x_2^H + \dots + \lambda_n x_n x_n^H. \quad (44)$$

Это разложение известно под названием *спектральной теоремы*. Оно выражает матрицу  $A$  в виде комбинации одномерных проекций  $x_i x_i^H$ , подобных проекциям  $aa^T$  из гл. 3. Они разбивают любой вектор  $b$  на его компоненты  $p = x_i (x_i^H b)$  по направлениям единичных собственных векторов, которые образуют множество взаимно перпендикулярных осей. Затем эти отдельные проекции  $p$  берутся с весами  $\lambda_i$  и суммируются, давая

$$Ab = \lambda_1 x_1 (x_1^H b) + \dots + \lambda_n x_n (x_n^H b). \quad (45)$$

Если все  $\lambda_i = 1$ , то мы получили сам вектор  $b$ ; матрица  $A = UIU^H$  является единичной. В любом случае мы можем проверить (44) и (45) непосредственно с помощью перемножения матриц:

$$\begin{aligned}
 Ab = U\Lambda U^H b &= \begin{bmatrix} | & | & | & | \\ x_1 & & & x_n \\ | & | & | & | \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \cdot \\ \cdot \\ \lambda_n \end{bmatrix} \begin{bmatrix} \text{---} \bar{x}_1 \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \bar{x}_n \text{---} \end{bmatrix} \begin{bmatrix} | \\ b \\ | \end{bmatrix} = \\
 &= \begin{bmatrix} | & | & | & | \\ x_1 & & & x_n \\ | & | & | & | \end{bmatrix} \begin{bmatrix} \lambda_1 x_1^H b \\ \cdot \\ \cdot \\ \lambda_n x_n^H b \end{bmatrix} = \lambda_1 x_1 x_1^H b + \dots + \lambda_n x_n x_n^H b.
 \end{aligned}$$

В нашем примере оба собственных вектора  $x = \begin{bmatrix} 1 \\ 1+i \end{bmatrix}$  и  $y = \begin{bmatrix} 1-i \\ -1 \end{bmatrix}$  имеют длину  $\sqrt{3}$ , и нормализация дает унитарную матрицу, которая диагонализует  $A$ :

$$U = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1-i \\ 1+i & -1 \end{bmatrix}, \quad U^{-1}AU = \begin{bmatrix} 8 & 0 \\ 0 & -1 \end{bmatrix}.$$

Тогда спектральное разложение  $U\Lambda U^H = \lambda_1 x_1 x_1^H + \lambda_2 x_2 x_2^H$  принимает вид

$$\frac{8}{3} \begin{bmatrix} 1 \\ 1+i \end{bmatrix} \begin{bmatrix} 1 & 1-i \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1-i \\ -1 \end{bmatrix} \begin{bmatrix} 1+i & -1 \end{bmatrix} = \begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix} = A.$$

**Упражнение 5.5.7.** Выписать любую неэрмитову матрицу, даже вещественную, и найти вектор  $x$ , такой, чтобы число  $x^H A x$  не было вещественным. (Только эрмитовы матрицы обладают свойством 1.)

**Упражнение 5.5.8.** Найти собственные значения и собственные векторы единичной длины для матрицы  $A = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}$ . Вычислить матрицы  $\lambda_1 x_1 x_1^H$  и  $\lambda_2 x_2 x_2^H$  и проверить, что в сумме они дают матрицу  $A$ , а их произведение есть нулевая матрица. (Почему  $(\lambda_1 x_1 x_1^H)(\lambda_2 x_2 x_2^H) = 0$ ?)

**Упражнение 5.5.9.** Используя собственные значения и собственные векторы, вычисленные на стр. 218, найти унитарную матрицу  $U$ , диагонализующую  $A$ , и выполнить спектральное разложение (в три матрицы  $\lambda_i x_i x_i^H$ ) для

$$A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

**Упражнение 5.5.10.** Доказать, что определитель любой эрмитовой матрицы является вещественным.

Спектральная теорема и остальное в свойстве 4 были на самом деле доказаны лишь в случае, когда собственные значения матрицы  $A$  различны. (Тогда, очевидно, имеется  $n$  независимых собственных векторов, что обеспечивает диагонализуемость матрицы  $A$ .) Тем не менее справедливо (см. § 5.6), что *даже при наличии кратных собственных значений эрмитова матрица по-прежнему имеет полный набор ортонормированных собственных векторов*. Следовательно, в любом случае матрица  $A$  может быть диагонализирована с помощью унитарной матрицы  $U$ .

Только эрмитовы матрицы обладают одновременно и вещественными собственными значениями, и ортонормированными собственными векторами. Если  $U^{-1}AU$  равняется вещественной диагональной матрице  $\Lambda$  и матрица  $U$  унитарна, то матрица  $A$  обязательно является эрмитовой:

$$A^H = (U\Lambda U^{-1})^H = U\Lambda U^H = A.$$

Для математика это разложение  $A = U\Lambda U^H$  отнюдь не выглядит неожиданным; это одно из основных соотношений, и никто не назвал бы его спорным. Однако психологи, например, не соглашались с этим (или друг с другом). Они ухитряются порождать множество дискуссий о спектральной теореме, и мы попытаемся это кратко объяснить. ***Но следующие две страницы читатель может пропустить, и это не скажется на цельности восприятия.***

#### ФАКТОРНЫЙ АНАЛИЗ И АНАЛИЗ ОСНОВНЫХ КОМПОНЕНТ

Наша цель состоит в отыскании некоторой естественной структуры в множестве экспериментальных данных — например, в наборе результатов тестов по физике, математике, истории и английскому языку. Эта задача относится к области так называемого многомерного анализа. Некоторые из его подразделов используются при обработке результатов экспериментов для «внешних целей», т. е. для прогнозирования чего-либо за пределами данных. Другие подразделы относятся полностью к их внутренней структуре.

(1) Проблема прогнозирования является типичной для *регрессионного анализа*: мы находим  $IQ^1$ ) каждого студента и пытаемся решить, можно ли было использовать результаты упомянутых выше тестов (или можно ли будет их использовать в будущем) для предсказания  $IQ$ . Другими словами, мы ищем линейную

<sup>1)</sup>  $IQ$  (intelligence quotient) — коэффициент умственного развития.—  
Прим. перев.

комбинацию результатов четырех тестов, дающую IQ. Такая комбинация содержит четыре неизвестных коэффициента, и каждому студенту соответствует одно уравнение. Это классическая задача, решаемая по методу наименьших квадратов, когда число уравнений больше числа неизвестных и собственные значения не участвуют в рассмотрении<sup>1)</sup>.

(2) Для иллюстрации вопросов о внутренней структуре предположим, что данные размещены в *корреляционной матрице*  $R$  размера  $4 \times 4$ . Диагональные элементы равняются 1 (каждая случайная переменная  $x_i$  идеально соотносится с собой), а внедиагональные элементы  $r_{ij}$  дают корреляцию между результатами  $i$ -го и  $j$ -го тестов, например, между физикой и математикой. Затем при помощи одной из разновидностей анализа, называемой *анализом основных компонент*, мы ищем некоррелирующие комбинации этих четырех тестов. Говоря языком линейной алгебры, мы ищем взаимно ортогональные векторы. Прекрасным средством для решения этой задачи является спектральное разложение вещественной симметрической матрицы  $R$ : если собственные векторы матрицы  $R$  суть  $q_1, \dots, q_4$ , то они образуют ортогональную матрицу  $Q$ , диагонализующую  $R$ :  $Q^{-1}RQ = \Lambda$ , или

$$R = Q\Lambda Q^1 = \begin{bmatrix} & & & \\ & & & \\ q_1 & \dots & q_4 & \\ & & & \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \lambda_4 \end{bmatrix} \begin{bmatrix} q_1^T \\ \vdots \\ q_4^T \end{bmatrix} = \\ = \begin{bmatrix} & & & \\ & & & \\ \sqrt{\lambda_1}q_1 & \dots & \sqrt{\lambda_4}q_4 & \\ & & & \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1}q_1^T \\ \vdots \\ \sqrt{\lambda_4}q_4^T \end{bmatrix}.$$

Матрица  $Q$  представляет собой унитарную матрицу  $U$ , которая в нашем случае вещественна. Каждый вектор  $\sqrt{\lambda_i}q_i$  дает комбинацию четырех тестов, и эти четыре специальные комбинации некоррелированы. Больших противоречий пока что не возникает.

(3) Теперь мы переходим к *факторному анализу*, который также должен объяснить смысл внедиагональных элементов корреляционной матрицы. Желательно, чтобы при этом использова-

<sup>1)</sup> Наш пример описывает в некотором роде «обратную» ситуацию, поскольку более естественно использовать IQ для каких-либо других предсказаний. Но, быть может, обратимость этих предсказаний, почти равноценная обратимости прошлого и будущего, имеет некий глубокий смысл.

лось сравнительно небольшое число факторов  $f_i$ , вновь являющихся линейными комбинациями исходных тестов. Но эти факторы не объясняют полностью смысл диагональных элементов  $r_{ii}=1$ . Если один тест дает очень узкий разброс оценок вследствие излишней трудности или излишней простоты вопросов, а другой дает широкий разброс, то факторный анализ игнорирует эту разницу. Он связан с получением правильной взаимозависимости, и идеальной ситуацией будет та, в которой корреляционная матрица разлагается так:

$$R = \begin{bmatrix} 1 & 0.74 & 0.24 & 0.24 \\ 0.74 & 1 & 0.24 & 0.24 \\ 0.24 & 0.24 & 1 & 0.74 \\ 0.24 & 0.24 & 0.74 & 1 \end{bmatrix} = \\ = \begin{bmatrix} 0.7 & 0.5 \\ 0.7 & 0.5 \\ 0.7 & -0.5 \\ 0.7 & -0.5 \end{bmatrix} \begin{bmatrix} 0.7 & 0.7 & 0.7 & 0.7 \\ 0.5 & 0.5 & -0.5 & -0.5 \end{bmatrix} + 0.26I.$$

В этом случае два фактора отвечают за все корреляции. Первый из них есть столбец с элементами, равными 0.7. Этот столбец следует интерпретировать как фактор общего умственного развития: хорошие результаты в математике соседствуют с хорошими результатами в родном языке. Второй фактор отличает сильную корреляцию между математикой и физикой, а также между английским и историей от более слабой корреляции между этими двумя группами. Это фактор типа «физика против лирики». Именно такие переменные являются правильными для интерпретации рассматриваемых данных, и их компоненты 0.7, 0.5 и  $-0.5$  называются *нагрузками* факторов на индивидуальных тестах. Во всех четырех тестах эти факторы отвечают за 0.74 единичной дисперсии главной диагонали; эти величины суть «общности», равные диагональным элементам матрицы  $FF^T$  в разложении  $R = FF^T + D$ .

Споры возникают по двум причинам. Во-первых, матрица факторных коэффициентов  $F$  и диагональная матрица  $D$  определяются отнюдь не единственным образом. Мы можем разложить ту же самую матрицу  $R$  так:

$$R = \begin{bmatrix} 0.6 & \sqrt{0.38} & 0 \\ 0.6 & \sqrt{0.38} & 0 \\ 0.4 & 0 & \sqrt{0.58} \\ 0.4 & 0 & \sqrt{0.58} \end{bmatrix} \begin{bmatrix} 0.6 & 0.6 & 0.4 & 0.4 \\ \sqrt{0.38} & \sqrt{0.38} & 0 & 0 \\ 0 & 0 & \sqrt{0.58} & \sqrt{0.58} \end{bmatrix} + D.$$

Здесь мы изменили число факторов, которое заранее никому не известно. На самом деле и «общности» никому не известны, так что мы можем изменять  $D$  и получить хоть миллион совершенно различных разложений. Абсолютно неясно также, какой вес должен быть назначен уровню общего умственного развития и как разделить «физику» от «лирики».

Другой источник споров появляется после выбора  $D$  и числа  $p$  искоемых факторов и состоит в возможности «вращения факторов». Для любой ортогональной матрицы  $Q$  порядка  $p$  матрица факторных коэффициентов  $\tilde{F} = FQ$  отвечает в точности за те же корреляции  $\tilde{F}\tilde{F}^T = FQQ^T F^T = FF^T$ , что и сама  $F$ . Следовательно, матрицы  $F$  и  $\tilde{F}$  полностью взаимозаменяемы. В типичной задаче исходные факторы могут иметь существенные нагрузки на десятки переменных и интерпретировать такой фактор практически невозможно. Он имеет математический смысл как вектор  $f_i$ , но не имеет никакого полезного смысла для социолога. Поэтому мы пытаемся подобрать вращение, которое давало бы простую структуру, с большими нагрузками на малое число компонент и пренебрежимо малыми нагрузками на остальные. Для получения положительных нагрузок используются даже наклонные факторы: в нашем втором разложении для  $R$  ортогональность  $f_i^T f_j = 0$  была потеряна. Поэтому два специалиста, оценивая  $D$ ,  $p$  и  $Q$ , очень легко могут получить совершенно различные интерпретации одних и тех же данных. Но техника эта настолько необходима, что, начинаясь как «паршивая овца» многомерного анализа, она распространилась от психологии в биологию, экономику и социологию.

**Упражнение 5.5.11.** Найти другое разложение  $FF^T + D$  той же самой матрицы  $R$ . Имеет ли оно какой-нибудь смысл?

### УНИТАРНЫЕ И КОСОЭРМИТОВЫ МАТРИЦЫ

Для начала рассмотрим некоторую аналогию. Эрмитову матрицу можно сравнить с вещественным числом, косоэрмитову матрицу — с чисто мнимым числом, а унитарную матрицу — с числом на единичной окружности, т. е. с комплексным числом  $e^{i\theta}$ , модуль которого равен единице. Для собственных чисел этих матриц предлагаемое сравнение является более чем аналогией: числа  $\lambda$  сами по себе являются вещественными, если  $A^H = A$ , чисто мнимыми, если  $K^H = -K$ , и находятся на единичной окружности при  $U^H = U^{-1}$ . Для всех этих матриц их собственные векторы ортогональны и могут быть выбраны ортонормированными<sup>1)</sup>.

<sup>1)</sup> В следующем параграфе мы определим более широкий класс матриц, соответствующий множеству всех комплексных чисел. Матрица, не имеющая ортогональных собственных векторов, не принадлежит ни к одному из этих классов и выпадает из всей аналогии.

Поскольку свойства эрмитовой матрицы уже установлены, остается рассмотреть матрицы  $K$  и  $U$ . Мы ищем аналогичный набор четырех фундаментальных свойств и получаем его без труда благодаря тесной связи косоэрмитовой матрицы  $K$  с эрмитовой матрицей  $A$ :

**5М.** Если матрица  $K$  косоэрмитова, так что  $K^H = -K$ , то матрица  $A = iK$  эрмитова:

$$A^H = (i)^H (K)^H = (-i)(-K) = A.$$

Аналогично, если матрица  $A$  эрмитова, то  $K = iA$  косоэрмитова.

Пример эрмитовой матрицы, приведенный на предыдущих страницах, дает

$$K = iA = \begin{bmatrix} 2i & 3+3i \\ -3+3i & 5i \end{bmatrix} = -K^H.$$

Диагональные элементы всегда являются кратными  $i$  (ноль допускается).

Это простое соотношение между матрицами  $K$  и  $A$  превращает свойства 1—4 в следующие:

(1') Для любого  $x$  комбинация  $x^H K x$  является чисто мнимой.

(2') Каждое собственное значение матрицы  $K$  является чисто мнимым.

(3') Собственные векторы, соответствующие различным собственным значениям, ортогональны.

(4') Существует унитарная матрица  $U$ , такая, что  $U^{-1} K U = \Lambda$ .

**Упражнение 5.5.12.** Как показать невырожденность матрицы  $K - I$ , если  $K$  косоэрмитова?

**Упражнение 5.5.13.** Взять произвольную вещественную косоэрмитову матрицу  $K$  и вещественный вектор  $x$  и проверить, что  $x^H K x = 0$ . Объяснить, как это следует из свойства 1'.

**Упражнение 5.5.14.** Любая матрица  $Z$  может быть разложена в сумму эрмитовой и косоэрмитовой компонент,  $Z = A + K$ , аналогично разбиению комплексного числа  $z$  в сумму  $a + ib$ . Вещественная часть числа  $z$  равняется половине  $z + \bar{z}$ , а «вещественная часть» матрицы  $Z$  есть половина  $Z + Z^H$ . Найти подобную формулу для «мнимой части»  $K$  и разложить

$$Z = \begin{bmatrix} 3+i & 4+2i \\ 0 & 5 \end{bmatrix}$$

в  $A + K$ .

Теперь мы переходим к унитарным матрицам. Они уже определены как *квадратные матрицы с ортонормированными столбцами*:

$$U^H U = I, \text{ или } U U^H = I, \text{ или } U^H = U^{-1}. \quad (46)$$

Это приводит прямо к первому из их свойств: умножение на  $U$  не влияет на скалярные произведения, углы или длины. Мы уже знаем это свойство для ортогональных матриц (которые являются вещественными унитарными матрицами,  $U = Q$ ). Теперь мы переходим к комплексному случаю, и доказательство вновь очевидно:

|| **Свойство 1"**  $(Ux)^H(Uy) = x^H U^H U y = x^H y$  и  $\|Ux\|^2 = \|x\|^2$  <sup>1)</sup>.

Следующее свойство указывает расположение собственных значений матрицы  $U$  аналогично свойствам 2 и 2': для унитарной матрицы каждое собственное значение лежит на единичной окружности.

|| **Свойство 2"**. Каждое собственное значение  $\lambda$  матрицы  $U$  имеет модуль  $|\lambda|$ , равный 1.

Это следует прямо из  $Ux = \lambda x$ , если мы сравним длины в обеих частях:  $\|Ux\| = \|x\|$  по свойству 1" и  $\|\lambda x\| = |\lambda| \|x\|$ . Следовательно,  $|\lambda| = 1$ .

|| **Свойство 3"**. Собственные векторы, соответствующие различным собственным значениям, ортогональны.

Для доказательства выпишем соотношения  $Ux = \lambda x$  и  $Uy = \sigma y$  и их скалярное произведение

$$(Ux)^H U y = (\lambda x)^H \sigma y, \text{ или } x^H y = (\bar{\lambda} \sigma) x^H y. \quad (47)$$

Если бы  $\sigma$  равнялось  $\lambda$ , то мы имели бы  $\bar{\lambda} \sigma = \bar{\lambda} \lambda = |\lambda|^2 = 1$ . Но, поскольку  $\sigma$  отлично от  $\lambda$ , мы знаем, что  $\bar{\lambda} \sigma \neq 1$ , и потому  $x^H y = 0$ .

Нормализуя эти собственные векторы к единичной длине и располагая их по столбцам матрицы  $S$ , мы тем самым превращаем диагонализующую матрицу в унитарную матрицу, которую не следует путать с диагонализуемой матрицей  $U$ !

|| **Свойство 4"**. Если  $U$  — унитарная матрица, то она имеет унитарную диагонализующую матрицу  $S$ :  $S^{-1} U S = S^H U S = \Lambda$ .

Все четыре свойства иллюстрируются матрицами вращений

$$U(t) = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}.$$

Их собственные значения суть  $e^{it}$  и  $e^{-it}$ , и они имеют модуль 1. Собственными векторами являются  $\begin{bmatrix} 1 \\ -i \end{bmatrix}$  и  $\begin{bmatrix} 1 \\ i \end{bmatrix}$ , которые взаимно

<sup>1)</sup> При сохранении скалярных произведений длины сохраняются: для доказательства достаточно положить  $y = x$ .

ортогональны. После нормализации они образуют унитарную матрицу  $S$ :

$$U = SAS^{-1} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -i/\sqrt{2} & i/\sqrt{2} \end{bmatrix} \begin{bmatrix} e^{it} & \\ & e^{-it} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & i/\sqrt{2} \\ 1/\sqrt{2} & -i/\sqrt{2} \end{bmatrix}.$$

Очевидно, что вращения сохраняют длины,  $\|Ux\| = \|x\|$ , и в примере на стр. 245 эта матрица  $U$  определяется как экспонента от кососимметрической матрицы  $K$ :

$$\text{если } K = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \text{то } e^{Kt} = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}.$$

Мы обещали тогда связать каждую матрицу  $K$  с унитарной матрицей  $U$ :

5N. Если матрица  $K$  косоэрмитова, то  $e^{Kt}$  унитарна. Поэтому решение уравнения  $du/dt = Ku$  имеет такую же длину, как и начальное значение  $u_0$ :  $\|u(t)\| = \|e^{Kt}u_0\| = \|u_0\|$ .

Простейшее доказательство основано на диагонализации матрицы  $K$ :  $K = SAS^{-1}$  для некоторой унитарной матрицы  $S$  и чисто мнимой матрицы  $\Lambda$ . Тогда  $e^{\Lambda t}$  является диагональной матрицей с элементами  $e^{\lambda_1 t}, \dots, e^{\lambda_n t}$ , модули которых равны единице. Такая матрица является унитарной. Поэтому  $e^{Kt} = Se^{\Lambda t}S^{-1}$  есть произведение трех унитарных матриц. Следующее упражнение показывает, что это произведение должно быть унитарной матрицей.

**Упражнение 5.5.15.** Показать, что матрица  $UV$  является унитарной, если унитарны  $U$  и  $V$ . Использовать условие  $U^H U = I$ .

**Упражнение 5.5.16.** Показать, что модуль определителя унитарной матрицы равен единице,  $|\det U| = 1$ , но сам определитель не обязательно равен 1. Описать все матрицы размера  $2 \times 2$ , которые являются одновременно диагональными и унитарными.

**Упражнение 5.5.17.** Найти такой третий столбец, чтобы матрица

$$U = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} \\ 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & -1/\sqrt{2} \end{bmatrix}$$

была унитарной. Насколько произвольным является этот выбор?

**Упражнение 5.5.18.** Диагонализировать матрицу  $K = \begin{bmatrix} i & i \\ i & i \end{bmatrix}$ , вычислить  $e^{Kt} = Se^{\Lambda t}S^{-1}$  и проверить унитарность этой экспоненты. Чему равняется ее производная при  $t=0$ ?

**Упражнение 5.5.19.** Описать все матрицы размера  $3 \times 3$ , которые являются одновременно эрмитовыми, унитарными и диагональными. Сколько их?

## ВЕЩЕСТВЕННОСТЬ И КОМПЛЕКСНОСТЬ

$\mathbb{R}^n$  — пространство векторов с  $n$  вещественными компонентами  $\leftrightarrow$   $\mathbb{C}^n$  — пространство векторов с  $n$  комплексными компонентами

$$\text{длина: } \|x\|^2 = x_1^2 + \dots + x_n^2 \leftrightarrow \text{длина: } \|x\|^2 = |x_1|^2 + \dots + |x_n|^2$$

транспонированная:  $A_{ij}^T = A_{ji} \leftrightarrow$  сопряженно транспонированная:

$$A_{ij}^H = \overline{A_{ji}}$$

$$(AB)^T = B^T A^T \leftrightarrow (AB)^H = B^H A^H$$

скалярное произведение:

скалярное произведение:

$$x^T y = x_1 y_1 + \dots + x_n y_n \leftrightarrow x^H y = \bar{x}_1 y_1 + \dots + \bar{x}_n y_n$$

$$(Ax)^T y = x^T (A^T y) \leftrightarrow (Ax)^H y = x^H (A^H y)$$

ортогональность:  $x^T y = 0 \leftrightarrow$  ортогональность:  $x^H y = 0$

симметрические матрицы:  $A^T = A \leftrightarrow$  эрмитовы матрицы:  $A^H = A$

$x^H A x$  вещественно, каждое собственное значение вещественно и

$$A = U \Lambda U^{-1} = U \Lambda U^H$$

кососимметрические матрицы:  $K^T = -K \leftrightarrow$  косоэрмитовы матрицы:  $K^H = -K$

$x^H K x$  мнимо, каждое собственное значение мнимо и  $K = iA$

ортогональные матрицы:

$\leftrightarrow$  унитарные матрицы:

$$Q^T Q = I, \text{ или } Q^T = Q^{-1} \quad U^H U = I, \text{ или } U^H = U^{-1}$$

$$(Qx)^T (Qy) = x^T y \text{ и } \|Qx\| = \|x\| \leftrightarrow (Ux)^H (Uy) = x^H y \text{ и } \|Ux\| = \|x\|$$

Столбцы, строки и собственные векторы являются ортонормированными

и каждое  $|\lambda| = 1$ .

## § 5.6. ПРЕОБРАЗОВАНИЯ ПОДОБИЯ И ТРЕУГОЛЬНЫЕ ФОРМЫ

Практически каждый шаг в этой главе имел отношение к комбинации  $S^{-1}AS$ , где столбцы матрицы  $S$  являлись собственными векторами матрицы  $A$ . Когда  $A$  была эрмитовой или косоэрмитовой, мы писали  $U$  вместо  $S$ , чтобы подчеркнуть ортонормированность собственных векторов. Но эта матрица по-прежнему оставалась матрицей из собственных векторов. Теперь в последнем параграфе мы будем рассматривать другие комбинации  $M^{-1}AM$ , образованные по типу  $S^{-1}AS$ , но с произвольной невырожденной матрицей  $M$ . Матрицы собственных векторов  $S$  может не существовать или мы можем не знать ее, или не хотим ее использовать.

Стоит вспомнить, каким образом возникают эти комбинации. Рассмотрим дифференциальное или разностное уравнение относи-

тельно неизвестной  $u$  и введем новую неизвестную  $v$  путем «замены переменных»  $u = Mv$ . Тогда

$$\frac{du}{dt} = Au \quad \text{переходит в} \quad M \frac{dv}{dt} = AMv, \quad \text{или} \quad \frac{dv}{dt} = M^{-1}AMv,$$

$$u_{n+1} = Au_n \quad \text{переходит в} \quad Mv_{n+1} = AMv_n, \quad \text{или} \quad v_{n+1} = M^{-1}AMv_n.$$

Новая матрица в рассматриваемом уравнении есть  $M^{-1}AM$ . В частном случае  $M = S$  система распадается и гармоники развиваются независимо. На языке приложения  $A$  это означает, что в качестве нового базиса пространства были выбраны собственные векторы и соответствующее преобразование представляется диагональной матрицей  $S^{-1}AS = \Lambda$ . Это максимально возможное упрощение, но и другие менее радикальные упрощения также полезны: даже при  $M \neq S$  мы можем надеяться, что работать с матрицей  $M^{-1}AM$  будет легче, чем с самой  $A$ .

Предположим для начала, что мы рассматриваем все возможные матрицы  $M$ . Это дает целое семейство матриц  $M^{-1}AM$ , которые называются **подобными** друг другу (и матрице  $A$ ). Переход от  $A$  к  $M^{-1}AM$  называется **преобразованием подобия**. Любую из этих матриц можно ввести в рассматриваемое дифференциальное или разностное уравнение простой заменой  $u = Mv$ , так что они должны иметь между собой что-то общее, а именно: **подобные матрицы имеют одинаковые собственные значения**.

|| **50.** Если  $B = M^{-1}AM$ , то  $A$  и  $B$  имеют одни и те же собственные значения с одинаковыми кратностями.

Доказательство этого факта очень простое. Собственные значения матрицы  $B$  являются корнями многочлена

$$\det(B - \lambda I) = \det(M^{-1}AM - \lambda I) = \det(M^{-1}(A - \lambda I)M) = \\ = \det M^{-1} \det(A - \lambda I) \det M = \det(A - \lambda I). \quad (48)$$

Поскольку  $A$  и  $B$  имеют один и тот же характеристический многочлен, они имеют одинаковые собственные значения. Если  $x$  есть собственный вектор, соответствующий собственному значению  $\lambda$ , то

$$Ax = \lambda x, \quad \text{или} \quad MBM^{-1}x = \lambda x, \quad \text{или} \quad B(M^{-1}x) = \lambda(M^{-1}x).$$

Это вновь доказывает, что  $\lambda$  является также и собственным значением матрицы  $B$  с собственным вектором  $M^{-1}x$ .

При решении уравнения  $Ax = b$  комбинации  $M^{-1}AM$  не возникают; там главная операция заключалась в умножении матрицы  $A$  (только слева!) на матрицу, которая осуществляет вычитание кратного одной строки из другой. Такое преобразование сохраняет нуль-пространство и пространство строк матрицы  $A$ , но оно совсем не заботится о собственных значениях. В отличие от этого преобразования подобия не изменяют собственных значений, и

фактически эти собственные значения вычисляются с помощью последовательности простых преобразований подобия. Матрица постепенно приобретает треугольную форму, и собственные значения постепенно появляются на главной диагонали. (Такая последовательность описывается в гл. 7, а один шаг ее иллюстрируется в упр. 5.6.2 ниже.) Это значительно лучше, чем пытаться вычислить многочлен  $\det(A - \lambda I)$ , корни которого должны быть собственными значениями. На практике оказывается невозможным сконцентрировать всю эту информацию в многочлене, а затем извлечь ее оттуда.

**Упражнение 5.6.1.** Показать, что, если матрица  $B$  подобна  $A$  с матрицей перехода  $M$ , а матрица  $C$  подобна  $B$  с матрицей  $M'$ , то  $C$  подобна  $A$ . Какие матрицы подобны единичной?

**Упражнение 5.6.2.** Рассмотреть произвольную матрицу  $A$  и специальную матрицу  $M$ :

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}, \quad M = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Выбрать такой угол поворота  $\theta$ , чтобы элемент на месте (3, 1) в матрице  $M^{-1}AM$  был равен нулю.

*Замечание.* Продолжить такое «уничтожение» не так просто, поскольку вращения, дающие нули на местах  $d$  и  $h$ , «испортят» полученный нами нуль в углу. Для матриц произвольного размера мы должны оставлять одну диагональ под главной и завершать вычисление собственных значений другим путем. В противном случае корни любого многочлена можно было бы находить с использованием только квадратных корней, определяющих  $\theta$ , а это невозможно.

**Упражнение 5.6.3.** (а) Для обратимой матрицы  $A$  найти такую матрицу  $M$ , которая позволит доказать подобие матриц  $AB$  и  $BA$ , откуда будет следовать равенство их собственных значений.

(б) Вывести отсюда, что матрицы  $AB$  и  $BA$  имеют одинаковый след, а затем доказать этот факт непосредственно путем сложения элементов их главных диагоналей, если

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad B = \begin{bmatrix} q & r \\ s & t \end{bmatrix}.$$

(с) Из рассмотрения следов матриц вывести, что равенство  $AB - BA = I$  не может выполняться для произвольных матриц  $A$  и  $B$ . В гильбертовом пространстве основные уравнения квантовой механики имеют вид

$$Au = \frac{du}{dx}, \quad Bu = xu, \quad (AB - BA)u = \frac{d}{dx}(xu) - x \frac{du}{dx} = u^1,$$

и они должны иметь решения.

<sup>1)</sup> Бесконечномерная линейная алгебра была использована фон Нейманом при создании математического фундамента для квантовой механики; момент  $A$  и положение  $B$  представляют собой одну из пар в принципе неопределенности Гейзенберга.

ТРЕУГОЛЬНЫЕ ФОРМЫ С УНИТАРНОЙ МАТРИЦЕЙ  $M$ 

Наше первое отступление от обычного случая  $M=S$  может показаться не слишком разумным: вместо того чтобы допустить более общую матрицу  $M$ , мы поступаем наоборот и *ограничиваемся унитарной*. Задача состоит в отыскании некоторой достаточно простой формы, которую может принять матрица  $M^{-1}AM$  при этом ограничении. Если собственные векторы матрицы не ортогональны, то получить диагональную матрицу  $\Lambda$  невозможно. Однако следующее утверждение, так называемая лемма Шура, дает очень полезную (по крайней мере теоретически) форму матрицы<sup>1)</sup>.

**5Р.** Для произвольной квадратной матрицы  $A$  существует унитарная матрица  $M=U$ , такая, что  $U^{-1}AU=T$  есть верхняя треугольная матрица. Собственные значения матрицы  $T$  совпадают с собственными значениями матрицы  $A$  и расположены на ее главной диагонали.

**Доказательство.** Любая матрица, например размера  $4 \times 4$ , имеет по крайней мере одно собственное значение  $\lambda_1$ ; в наихудшем случае оно будет повторяться четыре раза. Поэтому матрица  $A$  имеет по крайней мере один собственный вектор  $x$ . Мы нормализуем  $x$  к единичному вектору  $x_1$  и помещаем его в первый столбец матрицы  $U$ . На этом этапе определить остальные три столбца невозможно, и мы заполняем их произвольным образом, но так, чтобы получившаяся матрица  $U_1$  была унитарной. (Процесс Грама—Шмидта гарантирует осуществимость этого.) Произведение  $U_1^{-1}AU_1$  имеет по крайней мере один столбец нужного вида:  $Ax_1 = \lambda_1 x_1$  означает, что

$$AU_1 = U_1 \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{bmatrix}, \quad \text{или} \quad U_1^{-1}AU_1 = \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{bmatrix}.$$

На следующем шаге мы работаем с матрицей размера  $3 \times 3$ , расположенной в правом нижнем углу. Эта матрица имеет собственное значение  $\lambda_2$  и единичный собственный вектор  $x_2$ , который может быть помещен в первый столбец унитарной матрицы  $M_2$  размера  $3 \times 3$ . Тогда

$$U_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & M_2 & \\ 0 & & & \end{bmatrix} \quad \text{и} \quad U_2^{-1}(U_1^{-1}AU_1)U_2 = \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & \lambda_2 & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{bmatrix}.$$

<sup>1)</sup> Остаток этой главы посвящается главным образом теории, а не приложениям.

Наконец, на последнем шаге собственный вектор матрицы размера  $2 \times 2$  в правом нижнем углу помещается в унитарную матрицу  $M_3$ , находящуюся в углу матрицы  $U_3$  и

$$U_3^{-1} (U_2^{-1} U_1^{-1} A U_1 U_2) U_3 = \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & \lambda_2 & * & * \\ 0 & 0 & \lambda_3 & * \\ 0 & 0 & 0 & * \end{bmatrix} = T.$$

Произведение  $U = U_1 U_2 U_3$  вновь является унитарной матрицей (упр. 5.5.15), так что мы получаем требуемое представление  $U^{-1} A U = T^{-1}$ .

Поскольку эта лемма справедлива для всех матриц, она часто позволяет нам обойтись без предположения о диагонализуемости матрицы  $A$ . Например, мы можем использовать ее для доказательства того, что *степени  $A^k$  стремятся к нулю, когда все  $|\lambda_i| < 1$ , и что экспоненты  $e^{A t}$  стремятся к нулю, когда все  $\operatorname{Re} \lambda_i < 0$* , даже без полного набора собственных векторов, предполагавшегося теорией устойчивости в § 5.3 и 5.4.

**Упражнение 5.6.4.** Повторить шаги доказательства для нахождения  $U^{-1} A U = T$ , если  $A = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix}$ .

**Упражнение 5.6.5.** Найти треугольную форму  $U^{-1} A U = T$  для  $A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$ .

**Упражнение 5.6.6.** Доказать теорему Гамильтона—Кэли о том, что любая матрица удовлетворяет своему характеристическому уравнению: если  $p(\lambda) = \det(A - \lambda I)$ , то  $p(A) = 0$ . Указание: доказать, что  $p(T) = 0$ , например, для матрицы размера  $3 \times 3$  путем перемножения треугольных матриц  $p(T) = (T - \lambda_1)(T - \lambda_2)(T - \lambda_3)$ . Затем подставить матрицу  $A = U T U^{-1}$  в  $p(A) = (A - \lambda_1)(A - \lambda_2)(A - \lambda_3)$  и получить равенство  $p(A) = 0$ .

### ДИАГОНАЛИЗАЦИЯ ЭРМИТОВЫХ МАТРИЦ И СПЕКТРАЛЬНАЯ ТЕОРЕМА

В качестве применения этой треугольной формы мы хотим показать, что любая эрмитова матрица имеет полный набор ортонормированных собственных векторов безотносительно к тому, являются ли ее собственные значения различными или нет. Доказательство состоит из двух шагов:

(i) Если матрица  $A$  эрмитова, а  $U$  унитарна, то  $U^{-1} A U$  автоматически является эрмитовой:

$$(U^{-1} A U)^H = U^H A^H (U^{-1})^H = U^{-1} A U.$$

<sup>1)</sup> Для матриц большего размера доказательство проводится по индукции.

(ii) Матрица  $T = U^{-1}AU$  из теоремы 5Р оказывается поэтому одновременно треугольной и эрмитовой, так что она должна быть диагональной. Матрица  $T$  должна совпадать с  $\Lambda$ , когда  $A$  эрмитова.

|| 5Q. Любая эрмитова матрица  $A$  может быть диагонализирована при помощи подходящей унитарной матрицы  $U$ .

*Замечание 1.* То же самое справедливо для любой косоэрмитовой матрицы  $K$ , поскольку  $K = iA$ .

*Замечание 2.* Если матрица  $A$  вещественная и симметрическая, то такой же будет и матрица собственных векторов (или по крайней мере ее собственные векторы могут быть выбраны вещественными). Поэтому матрица  $U$  оказывается вещественной и унитарной, другими словами, она является ортогональной матрицей  $Q$ .

*Замечание 3.* Для эрмитовых матриц вполне обоснованно существование полного набора ортонормированных собственных векторов даже при наличии кратных собственных значений. Мы можем трактовать матрицу  $A$  как предел эрмитовых матриц, имеющих различные собственные значения, и при подходе к пределу собственные векторы остаются ортогональными. В отличие от этого неэрмитовы матрицы

$$A(\theta) = \begin{bmatrix} 0 & \cos \theta \\ 0 & \sin \theta \end{bmatrix}$$

имеют собственные векторы  $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$  и  $\begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}$ . При  $\theta \rightarrow 0$  второй собственный вектор стремится к первому, который является *единственным* собственным вектором недиагонализуемой матрицы  $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ .

*Замечание 4.* Спектральная теорема доказана теперь даже для матрицы

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = A^H,$$

которая имеет повторяющиеся собственные значения  $\lambda_1 = \lambda_2 = 1$ ,  $\lambda_3 = -1$ . Одним из возможных наборов ортонормированных собственных векторов является

$$x_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad x_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

Они помещаются в столбцы унитарной матрицы  $U$ , и матрица  $A = U\Lambda U^H$  принимает вид

$$A = \sum \lambda_i x_i x_i^H = \lambda_1 \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \lambda_3 \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Но, поскольку  $\lambda_1 = \lambda_2$ , первые две проекции (каждая ранга 1) дают проекцию  $P_1$  ранга 2 и матрица  $A$  принимает вид

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \lambda_1 P_1 + \lambda_3 P_3 = (+1) \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} + (-1) \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Здесь имеется целая плоскость собственных векторов, соответствующих  $\lambda = 1$ , и векторы  $x_1$  и  $x_2$  можно было выбирать до некоторой степени произвольно. Поэтому  $x_1 x_1^H$  и  $x_2 x_2^H$  были столь же произвольны и только их сумма, т. е. проекция  $P_1$  на всю плоскость, определена единственным образом. Каждая эрмитова матрица с  $k$  различными собственными значениями имеет свое собственное «спектральное разложение» вида  $A = \lambda_1 P_1 + \dots + \lambda_k P_k$ , где  $P_i$  есть проекция на собственное подпространство, соответствующее числу  $\lambda_i$ . Поскольку здесь имеется полный набор собственных векторов, эти проекции дают в сумме единичную матрицу. И, поскольку эти собственные пространства ортогональны, применение последовательно двух разных проекций должно давать нуль:  $P_j P_i = 0$ .

**Упражнение 5.6.7.** (а) Найти такую унитарную матрицу  $U$ , что  $U^{-1}AU = \Lambda$  при

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{и} \quad \Lambda = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Затем найти другую пару ортонормированных собственных векторов  $x_1, x_2$  для  $\lambda = 0$ .

(б) Проверить, что матрица  $P = x_1 x_1^H + x_2 x_2^H$  одна и та же для обеих пар.

**Упражнение 5.6.8.** Доказать в два этапа, что каждая унитарная матрица  $A$  диагонализуема:

(i) Если матрицы  $A$  и  $U$  унитарны, то и матрица  $T = U^{-1}AU$  унитарна.

(ii) Верхняя треугольная матрица  $T$ , являющаяся одновременно унитарной, должна быть диагональной.

Отсюда следует, что треугольная матрица  $T$  в теореме есть  $\Lambda$  и что любая унитарная матрица (с различными или кратными собственными значениями) имеет полный набор ортонормированных собственных векторов:  $U^{-1}AU = \Lambda$ . Все собственные значения удовлетворяют соотношению  $|\lambda| = 1$ .

Мы подошли очень близко к ответу на естественный и важный вопрос, для каких матриц треугольная матрица  $T$  совпадает с диагональной матрицей  $\Lambda$ , и легко можем пройти оставшуюся часть пути. Эрмитовы, косоэрмитовы и унитарные матрицы уже попали в этот класс: они соответствуют числам на вещественной оси, на чисто мнимой оси и на единичной окружности. Теперь мы хотим найти весь класс, соответствующий всем комплексным числам.

**5R.** Матрица  $N$  называется **нормальной**, если она коммутирует с  $N^H$ :  $NN^H = N^HN$ . Для таких и только таких матриц треугольная матрица  $T = U^{-1}NU$  является диагональной матрицей  $\Lambda$ . Нормальные матрицы — это в точности те, которые обладают полным набором ортонормированных собственных векторов.

Отметим, что эрмитовы (или косоэрмитовы) матрицы, безусловно, являются нормальными: если  $A = A^H$ , то  $AA^H$  и  $A^HA$  равняются  $A^2$ . Унитарные матрицы также являются нормальными, поскольку  $UU^H$  и  $U^HU$  равняются единичной матрице. В этих частных случаях мы доказывали, что  $T = \Lambda$ , в два этапа, и эти же два этапа будут участвовать в доказательстве в случае произвольной нормальной матрицы:

(i) Если матрица  $N$  нормальная, то такой же оказывается и матрица  $T = U^{-1}NU$ :

$$\begin{aligned} TT^H &= U^{-1}NUU^HN^HU = U^{-1}NN^HU = U^{-1}N^HNU = \\ &= U^HN^HUU^{-1}NU = T^HT. \end{aligned}$$

(ii) Верхняя треугольная матрица  $T$ , являющаяся нормальной, автоматически должна быть диагональной (упр. 5.6.10). Поэтому если матрица  $N$  нормальная, то треугольная матрица  $U^{-1}NU$  должна быть диагональной и, поскольку ее собственные значения такие же как и у матрицы  $N$ , она должна совпадать с  $\Lambda$ . Собственные векторы матрицы  $N$  суть столбцы матрицы  $U$ , и они ортонормированы. Фактически для матрицы  $N$  справедлива та же самая спектральная теорема, что и для эрмитовой матрицы,  $N = U\Lambda U^{-1} = \sum \lambda_i x_i x_i^H$ . Единственная разница состоит в том, что собственные значения  $\lambda$  не обязательно являются вещественными.

**Упражнение 5.6.9.** Найти нормальную матрицу, которая не является эрмитовой, косоэрмитовой, унитарной или диагональной.

**Упражнение 5.6.10.** Пусть  $T$  есть треугольная матрица размера  $3 \times 3$  с элементами  $t_{ij}$ . Сравнить элементы матриц  $TT^H$  и  $T^HT$  и показать, что если они равны между собой, то матрица  $T$  должна быть диагональной.

**Упражнение 5.6.11.** Доказать, что матрица с ортонормированными собственными векторами должна быть нормальной, как утверждалось в 5R: если  $U^{-1}NU = \Lambda$ , то  $NN^H = N^HN$ .

## ЖОРДАНОВА ФОРМА

До сих пор в этом параграфе мы старались сделать все что можно с помощью унитарных матриц: требуя, чтобы  $M$  была унитарной матрицей  $U$ , мы получали матрицу  $M^{-1}AM$  треугольной формы  $T$ . Теперь мы снимаем это ограничение на матрицу  $M$ . Будет допускаться любая матрица, и наша цель состоит в том, чтобы сделать матрицу  $M^{-1}AM$  как можно более близкой к диагональной.

Результатом этого наивысшего усилия в области диагонализации является жорданова форма  $J$ . Когда матрица  $A$  имеет полный набор независимых собственных векторов, мы полагаем  $M = S$  и получаем  $J = S^{-1}AS = \Lambda$ , жорданова форма в этом случае совпадает с диагональной матрицей  $\Lambda$ . Для «дефектной» матрицы это невозможно, и за каждый потерянный собственный вектор в жордановой форме будет появляться единица непосредственно над главной диагональю. Собственные значения должны при этом находиться на самой диагонали, поскольку матрица  $J$  является треугольной. И различные собственные значения могут быть «разъединены» точно так же, как и в диагональном случае; только при кратном  $\lambda$  может потребоваться (а может и не потребоваться) внедиагональный элемент в матрице  $J$ .

5S. Если матрица  $A$  имеет  $s$  независимых собственных векторов, то она подобна матрице вида

$$J = M^{-1}AM = \begin{bmatrix} J_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ & & & & J_s \end{bmatrix}.$$

Каждая жорданова клетка  $J_i$  является треугольной матрицей с единственным собственным значением и единственным собственным вектором:

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \cdot & \cdot & \\ & & \cdot & 1 \\ & & & & \lambda_i \end{bmatrix}.$$

Одно и то же собственное значение  $\lambda_i$  может появляться в нескольких клетках, если оно соответствует нескольким независимым собственным векторам. Две матрицы являются подобными тогда и только тогда, когда они имеют одну и ту же жорданову форму  $J$ .

Многие авторы делают эту теорему кульминационным пунктом своих курсов линейной алгебры. Честно говоря, я считаю это ошибкой. Совершенно справедливо, что не все матрицы можно привести к диагональному виду и что жорданова форма является наиболее общим случаем. Но именно поэтому ее построение оказывается одновременно и технически сложным, и исключительно неустойчивым. (Ничтожное изменение в матрице  $A$  может вернуть все потерянные собственные векторы и убрать внедиагональные единицы.) Поэтому надлежащим местом для обсуждения деталей жордановой формы служит приложение<sup>1)</sup>, а знакомство с ней лучше всего начать с некоторых характерных и не слишком сложных примеров.

Мы приведем два таких примера:

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{и} \quad B = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Поскольку нуль является собственным значением кратности три для обеих матриц, он будет присутствовать в каждой жордановой клетке, т. е. имеется либо одна клетка размера  $3 \times 3$ , либо клетки размера  $2 \times 2$  и  $1 \times 1$ , либо три клетки размера  $1 \times 1$ . Поэтому возможными жордановыми формами являются

$$J_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad J_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{и} \quad J_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

В случае матрицы  $A$  единственным собственным вектором будет  $(1, 0, 0)^T$ . Поэтому ее жорданова форма имеет только одну клетку и в соответствии с основной теоремой 5S матрица  $A$  должна быть подобна  $J_1$ . Матрица  $B$  имеет дополнительный собственный вектор  $(0, 1, 0)^T$ , и поэтому ее жорданова форма есть  $J_2$ : на диагонали должны быть две клетки. Что же касается  $J_3$ , то она представляет собой целый класс; единственная матрица, подобная нулевой, есть  $M^{-1}0M = 0$ .

В этих примерах для определения матрицы  $J$  было достаточно простого подсчета собственных векторов, и это всегда осуществимо, если мы рассматриваем такие простые примеры. Но в качестве общей теории эта техника не подходит, как показывает последнее из приведенных ниже упражнений.

<sup>1)</sup> Каждый из авторов пытается эти детали упростить, и мне кажется, что наилучшим является новое доказательство, полученное Филипповым. Оно почти настолько простое, чтобы изменить наше решение и вернуть построение матрицы  $J$  в основной текст.

**Упражнение 5.6.12.** Найти матрицу  $M$ , для которой  $M^{-1}BM = J$ .

**Упражнение 5.6.13.** Показать, что никакие две из этих жордановых форм не являются подобными:  $J_1 \neq M^{-1}J_2M$ ,  $J_1 \neq M^{-1}J_3M$  и  $J_2 \neq M^{-1}J_3M$ . (Использовать отсутствие подходящих  $M$ .)

**Упражнение 5.6.14.** Выписать все возможные жордановы формы для матрицы размера  $4 \times 4$  с собственным значением кратности четыре. (Общепринято записывать клетки меньшего размера в матрице  $J$  ниже клеток большего размера.) В случае двух независимых собственных векторов показать, что для выбора  $J$  есть две различные возможности.

#### ТАБЛИЦА ПРЕОБРАЗОВАНИЙ ПОДОБИЯ

1.  $A$  **диагонализуема**: столбцы матрицы  $S$  являются собственными векторами матрицы  $A$ , и матрица  $S^{-1}AS = \Lambda$  **диагональная**.
2.  $A$  **произвольна**: столбцы матрицы  $M$  являются собственными векторами и обобщенными собственными векторами матрицы  $A$ , и жорданова форма  $M^{-1}AM = J$  является **клеточно диагональной**.
3.  $A$  **произвольна** и  $U$  унитарна: матрица  $U$  может быть выбрана таким образом, что  $U^{-1}AU = T$  будет верхней треугольной.
4.  $A$  **нормальна**, т. е.  $AA^H = A^HA$ : можно выбрать такую матрицу  $U$ , что  $U^{-1}AU = \Lambda$ .

*Частные случаи при ортонормированных собственных векторах:*

- а. Если матрица  $A$  эрмитова, то  $\Lambda$  вещественна.
- а'. Если матрица  $A$  вещественная и симметрическая, то матрица  $\Lambda$  вещественная и  $U = Q$  ортогональная.
- б. Если матрица  $A$  косоэрмитова, то  $\Lambda$  мнима.
- с. Если матрица  $A$  унитарна, то все  $|\lambda_i| = 1$ .

#### ОБЗОРНЫЕ УПРАЖНЕНИЯ

5.1. Найти собственные векторы и собственные значения для

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}$$

и подобрать такую матрицу  $S$ , чтобы  $S^{-1}AS$  была диагональной.

5.2. Найти характеристический многочлен  $\det(A - \lambda I)$  и собственные значения для матрицы

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

5.3. Если матрица  $A$  имеет собственные значения 0 и 1, соответствующие собственным векторам

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{и} \quad \begin{bmatrix} 2 \\ -1 \end{bmatrix},$$

то можно ли заранее предсказать, что матрица  $A$  является симметрической? Чему равна сама  $A$ ? Чему равны ее след и определитель?

5.4. Каковы будут собственные значения и собственные векторы матрицы  $A^2$  из предыдущего упражнения? Как связаны между собой матрицы  $A^2$  и  $A$ ?

5.5. Существует ли такая матрица  $A$ , что все семейство матриц  $A + cI$  обратимо для любых чисел  $c$ ?

5.6. Найти матрицу  $A^n$ , если

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}.$$

5.7. Предпочитаете ли вы, чтобы при ежедневном начислении процентов вам платили 5% в первые три года и 6% в последующие три, или наоборот?

5.8. Решить систему

$$du/dt = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} u, \quad u_0 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

5.9. Найти  $e^{At}$ , если

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}.$$

5.10. Найти все квадратные корни из предыдущей матрицы  $A$ .

5.11. Верны или неверны следующие утверждения (привести контрпример в последнем случае):

- (i) Если матрица  $B$  образована из  $A$  путем перестановки двух строк, то  $B$  подобна  $A$ .
- (ii) Если верхняя треугольная матрица подобна диагональной матрице, то она уже диагональна.
- (iii) Любые два из следующих утверждений влекут за собой третье:  $A$  эрмитова,  $A$  унитарна,  $A^2 = I$ .
- (iv) Если матрицы  $A$  и  $B$  диагонализуются, то и  $A + B$  диагонализуема.

5.12. Найти ортогональную матрицу  $Q$  и диагональную матрицу  $\Lambda$ , такие, что  $Q^T A Q = \Lambda$ , если

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

5.13. Разложить матрицу

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

на  $A + K$ , где  $A = A^H$  и  $K = -K^H$ .

5.14. Выписать матрицу размера  $3 \times 3$ , у которой суммы элементов в строках равны 1, и показать, что число  $\lambda = 1$  является ее собственным значением. Чему равен соответствующий собственный вектор?

5.15. Что произойдет с последовательностью чисел Фибоначчи, если изменить направление времени на противоположное, и как число  $F_{-k}$  связано с  $F_k$ ? Правило  $F_{k+2} = F_{k+1} + F_k$  остается в силе, так что  $F_{-1} = 1$ .

# ПОЛОЖИТЕЛЬНО ОПРЕДЕЛЕННЫЕ МАТРИЦЫ

## § 6.1. МАКСИМУМЫ, МИНИМУМЫ И СЕДЛОВЫЕ ТОЧКИ

До сих пор нас не интересовали *знаки собственных значений*. Действительно, было бы преждевременно спрашивать, положительно  $\lambda$  или отрицательно, пока не станет известно, вещественно ли оно. Но в гл. 5 было отмечено, что наиболее важные матрицы — симметрические матрицы в вещественном случае и эрмитовы матрицы в комплексном случае — имеют вещественные собственные значения. Следовательно, имеет смысл выяснить, положительны ли они, и одна из наших последующих целей — найти критерий, с помощью которого по самой симметрической матрице  $A$ , не вычисляя ее собственных значений, можно заранее установить, что все ее собственные значения положительны.

Прежде чем заняться поисками такого критерия, рассмотрим задачу, в которой знаки собственных значений важны. Здесь совсем другое положение, чем при исследовании устойчивости в теории дифференциальных уравнений, где были нужны отрицательные собственные значения, а не положительные. (Нам не следовало бы так мимоходом говорить об этом, но все же: если для  $-A$  выполняется критерий, который мы ищем, то уравнение  $du/dt = Au$  имеет затухающее решение  $e^{\lambda t}x$  для каждого собственного значения  $\lambda < 0$ , а уравнение  $d^2u/dt^2 = Au$  имеет чисто осциллирующее решение  $e^{i\omega t}x$  с  $\omega = \sqrt{-\lambda}$ .) Эта новая задача возникает во многих приложениях, и мы надеемся, что читатель имеет о ней представление. Поэтому мы начинаем непосредственно с ее математической постановки. Это задача отыскания минимума, и мы формулируем ее с помощью двух примеров:

$$F(x, y) = 7 + 2(x + y)^2 - y \sin y - x^3 + y^4$$

и

$$f(x, y) = 2x^2 + 4xy + y^2.$$

*Имеет ли хотя бы одна из этих функций минимум в точке  $x = y = 0$ ?*

*Замечание 1.* Члены нулевого порядка  $F(0, 0) = 7$  и  $f(0, 0) = 0$  не влияют на ответ. Их изменение соответствует простому сдвигу вверх или вниз графиков функций  $F$  и  $f$ .

*Замечание 2.* Линейные члены дают необходимое условие: чтобы данная точка могла быть точкой минимума, она должна быть стационарной. Для этого первые производные должны обращаться в нуль при  $x = y = 0$ , что и происходит:

$$\frac{\partial F}{\partial x} = 4(x + y) - 3x^2 = 0,$$

$$\frac{\partial F}{\partial y} = 4(x + y) - y \cos y - \sin y + 4y^3 = 0,$$

$$\frac{\partial f}{\partial x} = 4x + 4y = 0 \quad \text{и} \quad \frac{\partial f}{\partial y} = 4x + 2y = 0.$$

Таким образом, выбранная точка является стационарной для обеих функций  $F$  и  $f$ . Геометрическая поверхность  $z = F(x, y)$  касается горизонтальной плоскости  $z = 7$ , а поверхность  $z = f(x, y)$  касается плоскости  $z = 0$ . Вопрос о том, лежат ли графики функций  $F$  и  $f$  над этими плоскостями, когда мы удаляемся от точки касания  $x = y = 0$ .

*Замечание 3.* Квадратичные члены или, другими словами, вторые производные являются решающими:

$$\frac{\partial^2 F}{\partial x^2} = 4 - 6x = 4,$$

$$\frac{\partial^2 f}{\partial x^2} = 4,$$

$$\frac{\partial^2 F}{\partial x \partial y} = \frac{\partial^2 F}{\partial y \partial x} = 4,$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 4,$$

$$\frac{\partial^2 F}{\partial y^2} = 4 + y \sin y - 2 \cos y + 12y^2 = 2, \quad \frac{\partial^2 f}{\partial y^2} = 2.$$

Эти производные уже содержат ответ на интересующий нас вопрос, и поскольку они одинаковы у обеих функций, то ответ должен быть один и тот же. Таким образом, эти две функции ведут себя вблизи начала координат одинаково, поэтому функция  $F$  имеет в начале координат минимум тогда и только тогда, когда в этой точке имеет минимум функция  $f$ .

*Замечание 4.* Члены более высокого порядка в  $F$  не оказывают влияния на вопрос о локальном минимуме, но они могут помешать существованию в этой точке глобального минимума. Наш пример — это именно такой случай: из-за члена  $-x^3$  функция  $F$  с какого-то момента устремится к  $-\infty$  независимо от того, что происходит вблизи точки  $x = y = 0$ . Подобная ситуация невозможна для функции  $f$  или для любой другой квадратичной формы, которая не имеет членов высшего порядка. Каждая квадратичная

форма  $f = ax^2 + 2bxy + cy^2$  имеет стационарную точку в начале координат, где  $\partial f/\partial x = \partial f/\partial y = 0$ , и если она имеет локальный минимум при  $x = y = 0$ , то эта точка является также точкой глобального минимума. Поверхность  $z = f(x, y)$  будет по форме похожа на чашу, покоящуюся на одной точке в начале координат.

Подведем итог. Вопрос локального минимума для  $F$  эквивалентен такому же вопросу для  $f$ . Если бы стационарной точкой функции  $F$  оказалась точка  $x = \alpha$ ,  $y = \beta$ , а не  $x = y = 0$ , то единственным отличием было бы использование вторых производных в точке  $\alpha$ ,  $\beta$ :

$$f(x, y) = \frac{x^2}{2} \frac{\partial^2 F}{\partial x^2}(\alpha, \beta) + xy \frac{\partial^2 F}{\partial x \partial y}(\alpha, \beta) + \frac{y^2}{2} \frac{\partial^2 F}{\partial y^2}(\alpha, \beta). \quad (1)$$

Тогда  $f$  ведет себя вблизи точки  $(0, 0)$  так же, как  $F$  ведет себя вблизи точки  $(\alpha, \beta)$ .

Один случай пока надо исключить из рассмотрения. Он соответствует возможности  $F'' = 0$ , которая создает большие затруднения даже для функции одного переменного. Тогда в рассмотрение вводятся третьи производные, так как вторые производные не дают определенного решения. Чтобы избежать этих трудностей, обычно требуют, чтобы квадратичная часть была невырожденной, т. е. для настоящего минимума допускается обращение  $f$  в нуль *только* при  $x = y = 0$ . Квадратичная форма, которая строго положительна во всех других точках, называется **положительно определенной**.

Проблема теперь сводится к следующему: что является для функции от двух переменных  $x$  и  $y$  корректной заменой условия  $F'' > 0$ ? В случае одной переменной знака второй производной достаточно, чтобы сделать выбор между минимумом или максимумом. Но сейчас мы имеем три производные:  $F_{xx}$ ,  $F_{xy} = F_{yx}$  и  $F_{yy}$ . Эти три члена точно определяют функцию  $f$ , и они должны определять, имеет ли  $F$  (так же как  $f$ ) минимум. *Каковы же условия на  $a$ ,  $b$  и  $c$ , которые будут обеспечивать положительную определенность  $f = ax^2 + 2bxy + cy^2$ ?*

Легко найти одно необходимое условие:

(i) Если  $f$  положительно определена, то обязательно  $a > 0$ .

Мы просто рассматриваем точку  $x = 1$ ,  $y = 0$ , где  $ax^2 + 2bxy + cy^2$  равно  $a$ ; значение  $f$  должно быть положительно, если  $f$  положительно определена. Применительно к функции  $F$  это означает, что  $\partial^2 F/\partial x^2 > 0$ , поскольку, фиксируя значение  $y = 0$  и изменяя только  $x$ , для минимума мы должны иметь  $F'' > 0$ . Аналогично, если мы фиксируем  $x = 0$  и изменяем только  $y$ , возникает условие на коэффициент  $c$ :

(ii) Если  $f$  положительно определена, то обязательно  $c > 0$ .

Не так легко решить, являются ли условия (i) и (ii), взятые вместе, достаточными для положительной определенности формы  $f$ ? Есть ли необходимость рассмотрения смешанной производной, равной коэффициенту  $b$ ?

**Пример.**  $f = x^2 - 10xy + y^2$ . В этом случае величины  $a = 1$  и  $c = 1$  одновременно положительны. Выберем точку  $x = y = 1$ . Так как  $f(1, 1) = -8$ , то форма  $f$  не положительно определена. Условия  $a > 0$  и  $c > 0$  обеспечивают здесь возрастание  $f$  по направлениям  $x$  и  $y$ , но она может еще убывать вдоль какой-нибудь другой прямой. В данном случае это прямая  $x = y$ . Коэффициент  $b$  может подавлять  $a$  и  $c$ . В действительности невозможно установить положительную определенность, просматривая  $f$  вдоль какого-либо конечного числа фиксированных прямых.

Очевидно, что решение вопроса зависит от  $b$ . Для первоначальной функции  $f$  коэффициент  $b$  был положителен. Достаточно ли этого, чтобы обеспечить минимум? Ответ опять отрицателен: знак  $b$  не играет роли. Несмотря на положительность всех коэффициентов наша исходная форма  $2x^2 + 4xy + y^2$  не является положительно определенной, и ни  $F$ , ни  $f$  не имеют минимума, так как вдоль прямой  $x = -y$  мы имеем  $f(1, -1) = 2 - 4 + 1 = -1$ .

Таким образом, для того чтобы форма  $f$  была положительно определенной, нужно контролировать величину  $b$  (в зависимости от  $a$  и  $c$ ). Сейчас мы хотим найти точный критерий, дающий необходимое и достаточное условие для положительной определенности. Простейший способ — привести  $f$  к «полному квадрату»:

$$f = ax^2 + 2bxy + cy^2 = a \left( x + \frac{b}{a} y \right)^2 + \left( c - \frac{b^2}{a} \right) y^2. \quad (2)$$

Очевидно, что первый член справа положителен, так как он является полным квадратом, умноженным на положительный коэффициент  $a$  — необходимое условие (i) все еще в силе. Первый квадрат, однако, когда  $x = -b/a$  и  $y = -a/c$ , обращается в нуль, и в этой точке мы имеем  $f(-b/a, -a/c) = a(ac - b^2)$ . Следовательно, требуется новое необходимое условие.

(iii) Если  $f$  положительно определена, то обязательно  $ac > b^2$ .

Заметьте, что условия (i) и (iii), взятые вместе, автоматически приводят к условию (ii): если  $a > 0$  и  $ac > b^2 \geq 0$ , то обязательно  $c > 0$ . При этом правая часть (2), безусловно, будет положительной, и мы получаем, наконец, ответ на наш вопрос.

**6А.** Квадратичная форма  $f = ax^2 + 2bxy + cy^2$  положительно определена тогда и только тогда, когда  $a > 0$  и  $ac - b^2 > 0$ . Соответственно  $F$  имеет (строгий) минимум при  $x = y = 0$  тогда и только тогда, когда

$$\frac{\partial^2 F}{\partial x^2}(0, 0) > 0, \quad \left[ \frac{\partial^2 F}{\partial x^2}(0, 0) \right] \left[ \frac{\partial^2 F}{\partial y^2}(0, 0) \right] > \left[ \frac{\partial^2 F}{\partial x \partial y}(0, 0) \right]^2.$$

Отсюда легко получаются условия для максимума, так как  $f$  имеет максимум всякий раз, когда  $-f$  имеет минимум. Этот способ обращает знаки  $a$ ,  $b$  и  $c$  на противоположные и оставляет второе условие  $ac - b^2 > 0$  неизменным: квадратичная форма **отрицательно определена** тогда и только тогда, когда  $a < 0$  и  $ac - b^2 > 0$ . Те же самые изменения применимы для  $F$ .

Квадратичная форма вырождена, когда  $ac - b^2 = 0$ . Этот случай мы пока не рассматривали. Второй член в (2) должен исчезнуть, оставив только один квадрат  $f = a(x + (b/a)y)^2$ , который либо **положительно полуопределен**, когда  $a > 0$ , либо **отрицательно полуопределен**, когда  $a < 0$ . Приставка «полу» допускает возможность обращения  $f$  в нуль, как это происходит, например, в точке  $x = b$ ,  $y = -a$ . Геометрически это означает, что поверхность  $z = f(x, y)$  вырождается из «чаши» в бесконечно длинный «желоб». (Если говорить о поверхности  $z = x^2$  в трехмерном пространстве, то этот желоб тянется вдоль оси  $y$ , и каждое поперечное сечение будет той же самой параболой  $z = x^2$ .) Еще более вырожденную квадратичную форму дают условия  $a = b = c = 0$ , она является как положительно полуопределенной, так и отрицательно полуопределенной. Чаша становится совершенно плоской.

В случае одного измерения все возможности для функции  $F(x)$  сейчас были бы исчерпаны: либо  $F$  имеет минимум, либо она имеет максимум, либо  $F'' = 0$ . При двух переменных, однако, остается еще одна очень важная возможность, так как *величина  $ac - b^2$  может быть отрицательной*. Это имело место в нашем примере, когда величина  $b$  доминировала над  $a$  и  $c$  и соответственно  $f$  на одних направлениях была положительна, а на других отрицательна. Такая же ситуация возникает даже независимо от  $b$ , если  $a$  и  $c$  имеют противоположные знаки. Направления осей  $x$  и  $y$  дают противоположные результаты — на одной оси  $f$  возрастает, на другой убывает. Здесь полезно рассмотреть два частных случая:

$$f_1 = 2xy \quad \text{и} \quad f_2 = x^2 - y^2.$$

В первом из них  $b$  доминирует (при  $a = c = 0$ ); во втором величины  $a$  и  $c$  имеют противоположные знаки. В обоих случаях  $ac - b^2 = -1$ .

Это **неопределенные** квадратичные формы, потому что они могут принимать значения любого знака, т. е. в зависимости от  $x$  и  $y$  возможно как  $f > 0$ , так и  $f < 0$ . Таким образом, мы имеем стационарную точку, которая не является ни точкой максимума, ни точкой минимума. Она носит название **седловой точки**. (По названию вы уже можете предположить, что поверхность  $z = f(x, y)$ , скажем  $z = x^2 - y^2$ , имеет форму, подобную седлу: она спускается вниз вдоль оси  $y$  и поднимается вверх вдоль оси  $x$ .)

Вы можете представить себе дорогу, идущую в горах, где вершина перевала является минимумом, если смотреть вдоль цепи гор, и в то же время является максимумом, если смотреть на нее вдоль дороги (рис. 6.1).

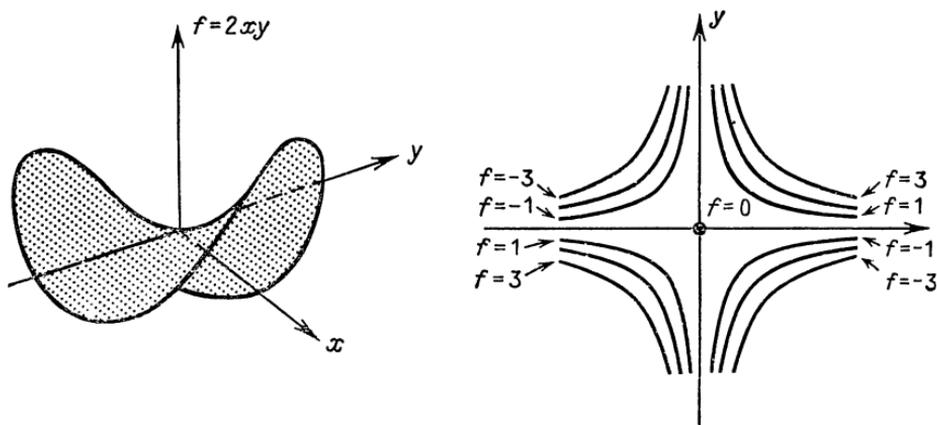


Рис. 6.1. Седло  $f=2xy$  и его линии уровня.

Седла  $2xy$  и  $x^2 - y^2$  практически одинаковы. Если мы повернем одно из них на угол  $45^\circ$ , то мы получим другое. Заметим также, что их почти невозможно нарисовать.

**Упражнение 6.1.1.** Показать, что исходная квадратичная форма  $f = 2x^2 + 4xy + y^2$  имеет седловую точку в начале координат, несмотря на положительность ее коэффициентов. Показать, как можно записать  $f$  в виде разности двух полных квадратов.

Чтобы установить условия минимума  $F_{xx} > 0$  и  $F_{xx}F_{yy} > F_{xy}^2$ , достаточно вычислений. Однако линейная алгебра позволит сделать больше, если мы поймем, как можно из коэффициентов формы  $f$  составить симметрическую матрицу  $A$ . Если коэффициенты членов  $ax^2$  и  $cy^2$  поставить в матрице на диагональ, а коэффициент члена  $2bxy$  расщепить на наддиагональный и поддиагональный элементы, то форма  $f$  будет тождественно равна такому произведению матриц:

$$ax^2 + 2bxy + cy^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \quad (3)$$

Это равенство является ключом ко всей главе. Его можно переписать в виде  $f = x^T A x$  и сразу обобщить на общий  $n$ -мерный случай. Такая сокращенная форма записи очень удобна для изучения проблемы максимума и минимума. В случае  $n$  независимых переменных  $x_1, \dots, x_n$  (вместо переменных  $x$  и  $y$  в двумерном случае) они составляют просто вектор-столбец  $x$ . Тогда для **ЛЮ-**

**бой симметрической матрицы  $A$  произведение  $f = x^T A x$  определяет чистую квадратичную форму.** Она имеет стационарную точку в начале координат и не содержит членов более высокого порядка:

$$\begin{aligned} x^T A x &= [x_1 \ x_2 \ \dots \ x_n] \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \\ &= a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + \dots + a_{nn}x_n^2 = \\ &= \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_ix_j. \end{aligned} \quad (4)$$

Заданную функцию  $F(x_1, \dots, x_n)$ , имеющую стационарную точку в начале координат (все первые ее производные равны нулю), мы будем исследовать на минимум, максимум или седловую точку, строя «матрицу Гессе»  $A$ , элементами которой являются вторые производные, т. е.  $a_{ij} = \partial^2 F / \partial x_i \partial x_j$ . Она автоматически будет симметрической матрицей, поскольку члены  $a_{12}x_1x_2$  и  $a_{21}x_2x_1$  равны между собой (это соответствует расщеплению  $2bxy$  в двумерном случае). При этом функция имеет минимум в начале координат, когда  $f$  положительно определена, или, иначе говоря, когда матрица  $A$  положительно определена, т. е.

$$x^T A x > 0 \quad \text{для всех ненулевых векторов } x. \quad (5)$$

**Упражнение 6.1.2.** Исследовать следующие матрицы на положительную определенность и выписать соответствующие функции  $f$ :

$$\begin{aligned} \text{(a)} \quad A &= \begin{bmatrix} 1 & 3 \\ 3 & 5 \end{bmatrix}, & \text{(b)} \quad A &= \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \\ \text{(c)} \quad A &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, & \text{(d)} \quad A &= \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}. \end{aligned}$$

В случае (d) записать  $x^T A x$  в виде суммы трех квадратов. В случае (b) показать, что вырожденная матрица соответствует вырожденной квадратичной форме (оба факта являются следствием обращения в нуль величины  $ac - b^2$ , равной определителю  $A$ ). Установить, вдоль какой прямой  $f$  тождественно равна нулю.

**Упражнение 6.1.3.** Для положительно определенной функции  $f$  кривая  $f(x, y) = 1$  является эллипсом (эта кривая — сечение чаши  $z = f(x, y)$  плоскостью  $z = 1$ ). Изобразить этот эллипс для случая  $a = c = 2$ ,  $b = -1$ .

**Упражнение 6.1.4.** Исследовать на минимум, максимум или седловую точку функции

$$\begin{aligned} \text{(a)} \quad F &= -1 + 4(e^x - x) - 5x \sin y + 6y^2 \quad \text{в точке } x = y = 0, \\ \text{(b)} \quad F &= (x^2 - 2x) \cos y \quad \text{в стационарной точке } x = 1, y = \pi. \end{aligned}$$

## § 6.2. КРИТЕРИИ ПОЛОЖИТЕЛЬНОЙ ОПРЕДЕЛЕННОСТИ

Какие симметрические матрицы обладают свойством  $x^T A x > 0$  для всех ненулевых векторов  $x$ ? Существует четыре или пять различных способов ответить на этот вопрос, и мы надеемся найти все их. Предыдущий параграф был начат с некоторых соображений относительно знаков собственных значений, но дальнейшая дискуссия была отложена. Вопрос о собственных значениях уступил место двум условиям на элементы матрицы: если  $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ , то для положительной определенности необходимы и достаточны условия  $a > 0$ ,  $ac - b^2 > 0$ . Наша цель сейчас — обобщить эти условия на матрицы порядка  $n$  и найти их связь со знаками собственных значений. По крайней мере в случае матрицы порядка 2 сформулированные условия означают, что оба ее собственных значения положительны. Действительно, их произведение равно определителю  $ac - b^2 > 0$ , и, следовательно, оба собственных значения либо положительны, либо отрицательны. Но так как их сумма равна следу  $a + c > 0$  матрицы  $A$ , то оба собственных значения будут обязательно положительными.

Интересно, как эти два подхода, один из которых связан с непосредственными вычислениями, а другой использует внутренние свойства матриц (их собственные значения), отражают две части настоящей книги, о которых уже говорилось выше. Если более внимательно посмотреть на первый критерий, то нетрудно заметить, что в нем фигурируют ведущие элементы матрицы. Действительно, если представить  $f$  в виде суммы квадратов

$$ax^2 + 2bxy + cy^2 = a \left( x + \frac{b}{a} y \right)^2 + \frac{ac - b^2}{a} y^2,$$

то коэффициенты  $a$  и  $(ac - b^2)/a$  в точности являются ведущими элементами матрицы  $A$  порядка 2. Если эта зависимость сохраняется для матриц большего порядка, то мы получим очень простой критерий положительной определенности матрицы, основанный на определении знаков ведущих элементов. Этот критерий имеет очень простую интерпретацию: форма  $x^T A x$  положительно определена тогда и только тогда, когда она представима в виде суммы  $n$  независимых квадратов.

Еще одно предварительное замечание. Как уже отмечалось, обе части книги объединяет теория определителей, и, следовательно, возникает вопрос, какую роль в положительной определенности матриц играют определители. Очевидно, что условия  $\det A > 0$  недостаточно для положительной определенности матрицы  $A$ , так как при  $a = c = -1$  и  $b = 0$  мы приходим к матрице  $A = -I$ , которая является отрицательно определенной. Оказывается, что выполнение рассматриваемого критерия (через опре-

делители) нужно требовать не только от матрицы  $A$  (в данном случае  $ac - b^2 > 0$ ), но также от ее подматрицы  $a$  порядка один, стоящей в левом верхнем углу. Естественным обобщением сказанного на  $n$ -мерный случай будет выполнение этих условий для всех  $n$  левых верхних подматриц

$$A_1 = [a_{11}], \quad A_2 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad A_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \dots, \quad A_n = A.$$

Сформулируем и дадим детальное доказательство следующей основной теоремы.

**6В.** *Каждый из перечисленных ниже критериев является необходимым и достаточным условием положительной определенности матрицы  $A$ :*

(I)  $x^T A x > 0$  для всех ненулевых векторов  $x$ .

(II) Все собственные значения  $\lambda_i$  матрицы  $A$  положительны.

(III) Определители всех подматриц  $A_k$  положительны.

(IV) Все ведущие элементы  $d_i$  (без перестановок строк) положительны.

**Доказательство.** Критерий I является определением положительно определенной матрицы, и в качестве первого шага мы покажем его эквивалентность критерию II. Для этого сначала мы предположим, что выполняется критерий I, и установим положительность каждого собственного значения  $\lambda_i$  матрицы  $A$ . Пусть  $x_i$  нормированный, т. е.  $x_i^T x_i = 1$ , собственный вектор  $A$ , соответствующий ее собственному значению  $\lambda_i$ . Тогда

$$A x_i = \lambda_i x_i \quad \text{и} \quad x_i^T A x_i = x_i^T \lambda_i x_i = \lambda_i,$$

поскольку  $x_i^T x_i = 1$ . Так как, согласно критерию I,  $x^T A x$  положительно для всех  $x$  и в том числе для  $x = x_i$ , то величина  $x_i^T A x_i = \lambda_i$  также должна быть положительной и, следовательно, **собственные значения положительно определенной матрицы положительны.**

Теперь, предполагая, что все  $\lambda_i > 0$ , покажем, что  $x^T A x > 0$  (это будет показано для произвольного ненулевого вектора  $x$ , а не только для собственного вектора). Поскольку симметрическая матрица обладает полным набором ортонормированных собственных векторов (см. стр. 272), то любой вектор  $x$  может быть представлен в виде линейной комбинации  $c_1 x_1 + \dots + c_n x_n$ . Отсюда имеем

$$A x = c_1 A x_1 + \dots + c_n A x_n = c_1 \lambda_1 x_1 + \dots + c_n \lambda_n x_n$$

и в силу условия ортонормированности системы  $x_i$

$$\begin{aligned} x^T A x &= (c_1 x_1^T + \dots + c_n x_n^T) (c_1 \lambda_1 x_1 + \dots + c_n \lambda_n x_n) = \\ &= c_1^2 \lambda_1 + \dots + c_n^2 \lambda_n. \end{aligned} \quad (6)$$

Если каждое  $\lambda_i > 0$  и не все  $c_i$  равны нулю, то  $x^T A x > 0$ . Таким образом, выполнение критерия II обеспечивает выполнение критерия I.

Эквивалентность критериев III и IV критерию I докажем в три этапа.

Сначала докажем, что из выполнения критерия I следует выполнение критерия III. С этой целью отметим, что определитель любой матрицы равен произведению ее собственных значений. Поэтому если выполнен критерий I и, следовательно, известно, что все собственные значения  $\lambda_i$  положительны, то

$$\det A = \lambda_1 \lambda_2 \dots \lambda_n > 0.$$

Чтобы установить этот же результат для всех подматриц  $A_k$ , покажем, что из положительной определенности  $A$  следует положительная определенность всех  $A_k$ . Для этого рассмотрим все векторы  $x$ , последние  $n - k$  компонент которых равны нулю,

т. е. векторы вида  $x = \begin{bmatrix} x_k \\ 0 \end{bmatrix}$ . Тогда

$$x^T A x = \begin{bmatrix} x_k^T & 0 \end{bmatrix} \begin{bmatrix} A_k & * \\ * & * \end{bmatrix} \begin{bmatrix} x_k \\ 0 \end{bmatrix} = x_k^T A_k x_k,$$

и, следовательно, если  $x^T A x > 0$  для всех ненулевых  $x$ , то, в частности,  $x_k^T A_k x_k > 0$  для всех ненулевых  $x_k$ . Итак, критерий I выполнен для  $A_k$ , так что к матрицам  $A_k$  применимы те же соображения, что мы применяли к самой матрице  $A$ . В частности, их собственные значения (которые не совпадают с  $\lambda_i$ !) и определители, которые являются произведениями соответствующих собственных значений, должны быть положительны.

Теперь покажем, что из критерия III следует критерий IV. Это устанавливается очень легко, поскольку между определителями подматриц  $A_k$  и ведущими элементами имеется прямая связь. Согласно стр. 205,  $k$ -й ведущий элемент  $d_k$  является в точности отношением  $\det A_k$  к  $\det A_{k-1}$ . Таким образом, если все определители положительны, то все ведущие элементы также положительны и, следовательно, положительно определенные матрицы ни в каких перестановках строк не нуждаются.

В заключение докажем, что из критерия IV следует критерий I. Иначе говоря, мы должны показать, что из положительности ведущих элементов следует неравенство  $x^T A x > 0$ . Для матриц порядка два мы делали это путем представления формы

$ax^2 + 2bxy + cy^2$  в виде суммы квадратов. Чтобы установить подобный факт для матриц любого порядка, необходим следующий важный факт: в методе исключения Гаусса для симметрической матрицы верхняя треугольная матрица  $U$  равна транспонированной к нижней треугольной матрице  $L$  (стр. 55) и, следовательно, разложение  $A = LDU$  переходит в разложение  $A = LDL^T$

**Пример**

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 \\ \frac{3}{2} \\ \frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 1 \end{bmatrix} = \\ = LDL^T.$$

Умножая эти выражения слева на  $x^T$ , а справа на  $x$ , мы приходим к сумме квадратов, в которой коэффициентами являются ведущие элементы:

$$x^T A x = [u \ v \ w] \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 \\ \frac{3}{2} \\ \frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \\ = 2 \left( u - \frac{1}{2} v \right)^2 + \frac{3}{2} \left( v - \frac{2}{3} w \right)^2 + \frac{4}{3} (w)^2.$$

Положительные ведущие элементы обеспечивают положительность выражения  $x^T A x$ . Таким образом, критерий I следует из критерия IV, что завершает последний этап доказательства теоремы<sup>1)</sup>. Аналогичная теорема справедлива и в комплексном случае для эрмитовых матриц  $A = A^H$ .

Было бы ошибкой, если бы от критерия III осталось впечатление, что здесь чрезвычайно важно рассматривать именно верхние подматрицы  $A_k$ . В равной степени мы могли бы проверить определители нижних правых подматриц. Или мы могли бы использовать любую цепочку главных подматриц, начиная с произвольного диагонального элемента  $a_{ii}$  в качестве первой подматрицы, и добавляя каждый раз необходимые элементы из новой строки и нового столбца с одинаковыми номерами. В частности,

<sup>1)</sup> Доказательство последнего этапа было проведено лишь для конкретного примера, полное доказательство будет дано чуть дальше посредством цепочки  $IV \Rightarrow V \Rightarrow I$ .

положительность элементов  $a_{ii}$ , которые являются коэффициентами при  $x_i^2$ , является *необходимым условием* положительной определенности матрицы  $A$ . Однако, как это видно из примеров, это условие далеко не является достаточным.

Ведущие элементы  $d_i$  ни в коем случае нельзя путать с собственными значениями  $\lambda_i$ . Для типичной положительно определенной матрицы это два совершенно различных множества положительных чисел. В последнем примере матрицы порядка 3 наиболее простым, по-видимому, является критерий, использующий определители (критерий III):

$$\det A_1 = 2, \quad \det A_2 = 3, \quad \det A_3 = \det A = 4.$$

Для больших матриц гораздо проще следить за ведущими элементами в методе исключения, который в рассматриваемом примере дает  $d_1 = 2$ ,  $d_2 = 3/2$ ,  $d_3 = 4/3$ . Обычно критерий, использующий собственные значения, наиболее трудоемкий, но для нашего примера мы знаем, что они положительны:

$$\lambda_1 = 2 - \sqrt{2}, \quad \lambda_2 = 2, \quad \lambda_3 = 2 + \sqrt{2}.$$

Заметим, что, хотя последний критерий является самым тяжелым в применении к конкретной матрице  $A$ , он является наиболее удобным для теоретических целей.

**Упражнение 6.2.1.** Исследовать на положительную определенность матрицы

$$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

Используя разложение  $B = LDL^T$ , записать выражение  $x^T B x = u^2 + v^2 + w^2 + 2uw$  в виде суммы квадратов.

**Упражнение 6.2.2.** Построить матрицу, наибольшие элементы которой лежат на главной диагонали (например, вида

$$A = \begin{bmatrix} 1 & \alpha & \beta \\ \alpha & 1 & \alpha \\ \beta & \alpha & 1 \end{bmatrix}$$

с  $|\alpha| < 1$ ,  $|\beta| < 1$ ) и которая не является даже положительно полуопределенной (Вычислить  $\det A$ .)

**Упражнение 6.2.3.** Показать, что из положительной определенности матрицы  $A$  следует положительная определенность матриц  $A^2$  и  $A^{-1}$ .

**Упражнение 6.2.4.** Показать, что из положительной определенности матриц  $A$  и  $B$  следует положительная определенность матрицы  $A + B$ . Какой из критериев I—IV здесь наиболее удобен?

Я надеюсь, вы мне позволите привести еще один критерий положительной определенности. Мы уже подошли к нему очень близко, он позволяет даже лучше понять приведенные ранее критерии:

**6С.** Матрица  $A$  положительно определена тогда и только тогда, когда

(V) существует такая невырожденная матрица  $W$ , что  $A = W^T W$ .

Из этого нового условия следует, что

$$x^T A x = x^T W^T W x = \|Wx\|^2. \quad (7)$$

Эта величина либо положительна, либо по крайней мере неотрицательна в зависимости от того, возможно или нет равенство  $Wx=0$  для ненулевого вектора  $x$ . Но поскольку матрица  $W$  невырожденная, то оно возможно только для  $x=0$  и, следовательно, из критерия V вытекает выполнение критерия I из 6B. Все это очень похоже на метод наименьших квадратов с матрицей весов на стр. 178; наиболее общая мера длины — это  $\|Wx\|$  и квадрат этой длины вводит положительно определенную матрицу  $A = W^T W$ .

Остается показать, что если матрица  $A$  удовлетворяет условиям I—IV, то она также будет удовлетворять условию V. Для этого нам нужно построить матрицу  $W$ , что в действительности дважды мы уже почти сделали.

(i) На завершающем этапе основной теоремы для матрицы  $A$  было построено разложение  $LDL^T$ . Так как ведущие элементы матрицы  $A$  положительны, то выбирая в качестве  $\sqrt{D}$  диагональную матрицу, диагональными элементами которой являются квадратные корни из соответствующих ведущих элементов, мы приходим к разложению  $A = L\sqrt{D}\sqrt{D}L^T$ <sup>1)</sup>. Таким образом, одним из возможных выборов  $W$  является верхняя треугольная матрица  $\sqrt{D}L^T$ . Это доказывает, что условие I является следствием условия IV.

(ii) Рассмотрим еще один способ, который дает другую матрицу  $W$ , используя вместо IV условие II. Так как собственные значения матрицы  $A$  положительны и ее собственные векторы

<sup>1)</sup> Это так называемое *разложение Холецкого*, когда ведущие элементы равномерно распределяются между нижним треугольным и верхним треугольным сомножителями. Оно почти не отличается от разложения  $A = LDL^T$ , и я обычно избегаю вычисления этих квадратных корней.

образуют ортогональную матрицу  $Q$ , которая диагонализует  $A$ , то справедливы разложения

$$A = QLQ^T = Q\sqrt{\Lambda}\sqrt{\Lambda}Q^T = (\sqrt{\Lambda}Q^T)^T(\sqrt{\Lambda}Q^T).$$

Теперь для  $W$  имеем другое представление  $\sqrt{\Lambda}Q^T$ , которое психологи используют в анализе основных компонент.

**Упражнение 6.2.5.** Построить двумя указанными способами матрицы  $W$  для матрицы

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

**Упражнение 6.2.6.** Вычислить квадратный корень  $R = Q\sqrt{\Lambda}Q^T$  для матрицы  $A$  из предыдущего упражнения. Проверить непосредственно и из ее формулы, что  $R$  симметрическая и  $R^2 = A$ . Собственные значения матрицы  $R$  являются положительными квадратными корнями  $\sqrt{\lambda_i}$ , поэтому нетрудно показать, что  $R$  является единственным симметрическим и положительно определенным квадратным корнем из  $A$ . Заметим также, что  $R$  является еще одной из возможных матриц  $W$ .

**Упражнение 6.2.7.** Доказать, что если матрица  $A$  положительно определена и матрица  $C$  невырожденная, то матрица  $B = C^T A C$  положительно определена.

**Упражнение 6.2.8.** Мы можем назвать несимметрическую матрицу  $A$  «положительно определенной», если ее эрмитова часть  $(A^H + A)/2$  является положительно определенной матрицей. Для такой матрицы, несмотря на то что величина  $x^H A x$  может быть комплексной, ее вещественная часть  $\frac{1}{2}(x^H A x + x^H A^H x)$  будет положительной.

(а) Показать, что для такой матрицы вещественная часть каждого ее собственного значения положительна. (Указание: рассмотреть вещественную часть скалярного произведения векторов  $Ax = (\alpha + i\beta)x$  и  $x^H$ .)

(б) На примере треугольной матрицы  $A$  порядка два показать, что ее собственные числа могут быть положительны, в то время как матрица  $A^H + A$  не является положительно определенной. (В теореме устойчивости БК для уравнения  $u' = Au$  из отрицательной определенности  $A^H + A$  следовало убывание решения в каждый момент времени, но уравнение могло быть устойчивым и без этого условия.)

**Упражнение 6.2.9.** Выписать условия отрицательной определенности матрицы  $A$ , обратив особое внимание на условие III, где нужно установить связь между  $\det(-A)$  и  $\det A$ .

### ***n*-МЕРНЫЕ ЭЛЛИПСОИДЫ**

Уравнение  $x^T A x = 1$  является ключевым для понимания геометрической интерпретации свойства положительной определенности. Вектор  $x$ , удовлетворяющий этому уравнению, лежит на поверхности в  $n$ -мерном пространстве, которая, как мы покажем, является эллипсом, или, более точно,  $n$ -мерным эллипсоидом с центром в начале координат.

Пример. Пусть  $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$  и  $x = \begin{bmatrix} u \\ v \end{bmatrix}$ . Тогда для анализа

уравнения  $x^T A x = 2u^2 - 2uv + 2v^2 = 1$  лучше всего переписать его в виде суммы квадратов с собственными значениями  $\lambda_2 = 3$  и  $\lambda_1 = 1$  матрицы  $A$  в качестве коэффициентов:

$$2u^2 - 2uv + 2v^2 = 1 \left( \frac{u}{\sqrt{2}} + \frac{v}{\sqrt{2}} \right)^2 + 3 \left( \frac{u}{\sqrt{2}} - \frac{v}{\sqrt{2}} \right)^2 = 1. \quad (8)$$

Обозначим выражения в скобках через  $w$  и  $z$  соответственно. Тогда графиком уравнения  $w^2 + 3z^2 = 1$  будет обычный эллипс. Концом его длинной оси будет точка  $w = 1, z = 0$ , а концом короткой — точка  $w = 0, z = 1/\sqrt{3}$ . Иначе говоря, половина длины большей из осей равна  $1/\sqrt{\lambda_1}$ , а половина длины меньшей из осей равна  $1/\sqrt{\lambda_2}$  (рис. 6.2). Кроме того, направления этих осей сов-

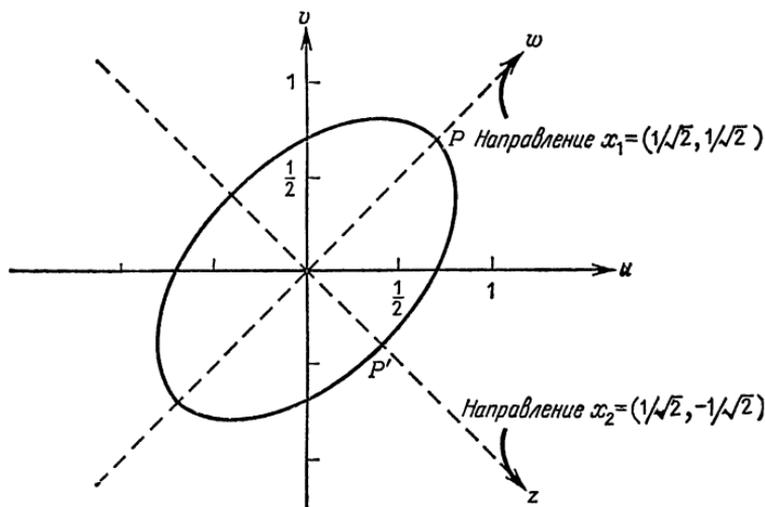


Рис. 6.2. Эллипс  $2u^2 - 2uv + 2v^2 = 1$ .

падают с направлениями соответствующих собственных векторов  $x_1 = (1/\sqrt{2}, 1/\sqrt{2})$  и  $x_2 = (1/\sqrt{2}, -1/\sqrt{2})$ . Таким образом, геометрия положительной определенности полностью связана с собственными числами и векторами матрицы  $A$ .

Если понятно, что было проделано с уравнением (8), то то же самое мы можем осуществить для любой поверхности  $x^T A x = 1$ . Основным шагом всегда является диагонализация матрицы  $A$  с помощью ортогональной матрицы  $Q$ , столбцами которой являются единичные собственные векторы  $x_j$  матрицы  $A$ :  $Q^T A Q = \Lambda$

или  $A = Q\Lambda Q^T$ . Это приводит к записи  $x^T A x$  в виде суммы квадратов:

$$\begin{aligned} x^T A x &= \begin{bmatrix} x^T \end{bmatrix} \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \\ &= \lambda_1 (x_1^T x)^2 + \dots + \lambda_n (x_n^T x)^2. \end{aligned}$$

В новых переменных  $y_1 = x_1^T x, \dots, y_n = x_n^T x$ , что соответствует замене переменных по формуле  $y = Q^T x$ , уравнение рассматриваемой поверхности принимает вид

$$x^T A x = \lambda_1 y_1^2 + \dots + \lambda_n y_n^2 = 1. \quad (9)$$

Это общий случай, частным примером которого является уравнение (8). Но даже в этом общем случае легко видеть, что точка с координатами  $y_1 = 1/\sqrt{\lambda_1}, y_2 = \dots = y_n = 0$  лежит на поверхности. Она лежит на конце наибольшей из осей и, следовательно, расположена дальше всех от начала координат. Более того, соответствующий вектор  $x$  является кратным первого собственного вектора, или, точнее,  $x = x_1/\sqrt{\lambda_1}$ . На рис. 6.2 эта точка обозначена буквой  $P$ . Ближайшая к началу координат точка лежит на конце наименьшей из осей и имеет координаты  $y_1 = \dots = y_{n-1} = 0, y_n = 1/\sqrt{\lambda_n}$ . Эта точка, обозначенная на рис. 6.2 буквой  $P'$ , лежит на прямой, задаваемой последним собственным вектором  $x_n$ . Остальные промежуточные оси задаются соответствующими промежуточными собственными векторами матрицы  $A$ .

**6D.** Предположим, что матрица  $A$  положительно определена и ее единичные собственные векторы являются столбцами ортогональной матрицы  $Q$ . Тогда  $A = Q\Lambda Q^T$ , и преобразование  $y = Q^T x$  позволяет записать выражение  $x^T A x$  в виде суммы квадратов:

$$x^T A x = x^T Q\Lambda Q^T x = y^T \Lambda y = \lambda_1 y_1^2 + \dots + \lambda_n y_n^2. \quad (10)$$

При этом уравнение  $x^T A x = 1$  описывает эллипсоид, конец  $j$ -й оси которого находится в точке с координатами  $y_1, \dots, y_n$ , где  $y_j$  определяется условием  $\lambda_j y_j^2 = 1$ , а остальные  $y_i$  равны нулю ( $j = 1, \dots, n$ ). Эти точки лежат на прямых, задаваемых собственными векторами матрицы  $A$ , и половина длины соответствующей оси равна  $1/\sqrt{\lambda_j}$ .

**Упражнение 6.2.10.** Эллипс  $u^2 + 4v^2 = 1$  соответствует матрице  $A = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$ .

Выписать собственные значения и векторы матрицы  $A$  и нарисовать этот эллипс.

**Упражнение 6.2.11.** Преобразовать уравнение  $3u^2 - 2\sqrt{2}uv + 2v^2 = 1$  к сумме квадратов, вычислив собственные значения и векторы соответствующей матрицы  $A$ . Нарисовать эллипс, задаваемый полученным уравнением.

**Упражнение 6.2.12.** Уравнение  $\lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2 = 1$  с тремя переменными задает эллипсоид, когда все  $\lambda_i > 0$ . Описать все возможные виды поверхностей в положительно полуопределенном случае, т. е. когда одно или несколько собственных значений равны нулю.

### § 6.3. ПОЛУОПРЕДЕЛЕННЫЕ И НЕОПРЕДЕЛЕННЫЕ МАТРИЦЫ. ОБОБЩЕННАЯ ЗАДАЧА НА СОБСТВЕННЫЕ ЗНАЧЕНИЯ $Ax = \lambda Bx$

Основной задачей этого раздела (после того как критерии положительной определенности полностью установлены) является более широкое рассмотрение проблемы определенности матриц. Под этим мы понимаем три основные задачи:

- (1) критерии положительной полуопределенности матриц,
- (2) связь между собственными значениями и ведущими элементами любой симметрической матрицы независимо от того, является она определенной или не является,
- (3) обобщенная задача на собственные значения  $Ax = \lambda Bx$ .

На первый вопрос мы можем дать быстрый и четкий ответ, поскольку вся необходимая для этого работа уже проделана. Второй и третий вопросы связаны с так называемым «законом инерции» для симметрических матриц, который с математической точки зрения является основным результатом этого параграфа. Одним из следствий этого закона является утверждение, что **знаки собственных значений определяют знаки ведущих элементов**. Другое же следствие утверждает, что если матрица  $B$  положительно определена, то собственные значения задачи  $Ax = \lambda Bx$  имеют те же знаки, что и собственные значения обычной задачи  $Ax = \lambda x$ . Собственные векторы задачи  $Ax = \lambda Bx$  «взаимно ортогональны», но в другом смысле, чем это было раньше. Обобщенная задача на собственные значения очень важна для приложений. Поэтому один пример будет рассмотрен в настоящем параграфе и другой (связанный с методом конечных элементов) — в параграфе 6.5.

Сформулируем для случая положительно полуопределенных матриц аналог утверждений 6B и 6C, доказанных ранее для положительно определенных матриц.

**6Е.** Каждый из приводимых ниже критериев является необходимым и достаточным условием **положительной полуопределенности** матрицы  $A$ .

(I')  $x^T A x \geq 0$  для всех векторов  $x$ .

(II') Все собственные значения  $\lambda_i$  матрицы  $A$  неотрицательны, т. е.  $\lambda_i \geq 0$ .

(III') Все главные подматрицы  $A_k$  имеют неотрицательные определители.

(IV') Все ведущие элементы  $d_i$  неотрицательны, т. е.  $d_i \geq 0$ .

(V') Существует такая матрица  $W$  (возможно, вырожденная), что  $A = W^T W$ .

Если ранг  $A$  равен  $r$ , то  $f = x^T A x$  представима в виде суммы  $r$  полных квадратов.

Связь между неравенствами  $x^T A x \geq 0$  и  $\lambda_i \geq 0$ , которая является для нас наиболее важной, устанавливается следующим образом.

Диагонализация  $A = Q \Lambda Q^T$  приводит к равенствам

$$x^T A x = x^T Q \Lambda Q^T x = y^T \Lambda y = \lambda_1 y_1^2 + \dots + \lambda_n y_n^2, \quad (11)$$

и это выражение неотрицательно для любого  $x$  только в том случае, когда все  $\lambda_i$  неотрицательны. Если ранг матрицы  $A$  равен  $r$ , то она имеет только  $r$  ненулевых собственных значений и, следовательно, в правой части (11) имеется только  $r$  полных квадратов.

Что же касается определителя матрицы  $A$ , то он, будучи равным произведению всех собственных значений  $A$ , неотрицателен. Далее, поскольку все главные подматрицы исходной матрицы  $A$  тоже будут положительно полуопределенными матрицами, то их собственные значения и определители также неотрицательны. Тем самым эквивалентность условия III' условиям I' и II' установлена. (Главные подматрицы матрицы  $A$  образуются отбрасыванием ее строк и столбцов с одинаковыми номерами, например строк и столбцов с номерами один и четыре, что сохраняет для подматриц свойство симметричности. Для главных подматриц сохраняется также свойство положительной полуопределенности: если  $x^T A x \geq 0$  для любого  $x$ , то это неравенство выполняется для всех векторов, у которых, например, первая и четвертая компоненты равны нулю.) Новым здесь является то, что III' прилагается ко всем главным подматрицам, а не только к расположенным в левом верхнем углу. Иначе мы не могли бы различать две матрицы, у которых все левые верхние определители равны нулю. Например, матрица  $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$  положительно полуопре-

делена, а матрица  $\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$  отрицательно полуопределена. Ведущие элементы для положительно полуопределенной матрицы опять будут отношениями определителей и, следовательно, будут неотрицательны. Однако при реализации исключений здесь может возникнуть необходимость в перестановках строк, и, чтобы не нарушать симметричность матрицы, мы должны одновременно осуществлять перестановку столбцов с теми же самыми номерами. Это означает, что на самом деле «симметрическое исключение» осуществляется не для матрицы  $PA$ , а для матрицы  $PAP^T$ . Далее, так как такие перестановки только перемещают главные подматрицы, оставляя их определители неизменными, то из III' обязательно вытекает условие IV': после  $r$  этапов исключения, где  $r$  — ранг матрицы  $A$ , оставшаяся подматрица в правом нижнем углу является нулевой матрицей и, следовательно,  $d_{r+1} = \dots = d_n = 0$ .

Наконец, выполнение условия IV' означает, что существует разложение  $PAP^T = LDL^T$  (или  $A = P^T LDL^T P$ ) с диагональной матрицей  $D$ , на диагонали которой стоят неотрицательные ведущие элементы. Теперь, выбирая  $W = \sqrt{D} L^T P$ , получаем  $A = W^T W$ , что означает выполнение V'. В свою очередь условие I' легко следует из условия V', так как  $x^T A x = x^T W^T W x = \|Wx\|^2 \geq 0$ , что завершает полный цикл доказательства утверждения 6Е.

**Пример.** Матрица

$$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

положительно полуопределена в силу следующих критериев:

- (II') Собственные значения  $\lambda_1 = 0$ ,  $\lambda_2 = \lambda_3 = 3$  неотрицательны.
- (III') Определители главных подматриц порядка один равны двум, порядка два равны трем и  $\det A = 0$ .
- (IV') Так как

$$\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & -\frac{3}{2} \\ 0 & -\frac{3}{2} & \frac{3}{2} \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

то величины  $d_1 = 2$ ,  $d_2 = 3/2$  и  $d_3 = 0$  являются ведущими элементами.

**Упражнение 6.3.1.** Для матрицы  $A$  из последнего примера найти матрицу  $W$  и записать выражение  $x^T A x = \|Wx\|^2$  в виде суммы двух квадратов. Описать в трехмерном пространстве поверхность  $x^T A x = 1$ .

Упражнение 6.3.2. Применить любые три критерия из 6Е к матрице

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

*Замечание.* Условия положительной полуопределенности можно также вывести из первоначальных условий I—V следующим образом. Добавим к исходной матрице  $A$  некоторое малое кратное  $\varepsilon I$  единичной матрицы, чтобы получить положительно определенную матрицу  $A + \varepsilon I$ , и устремим  $\varepsilon$  к нулю. Тогда, так как собственные значения и определители непрерывно зависят от  $\varepsilon$ , они должны быть положительны до самого последнего момента и, следовательно, должны быть неотрицательны при  $\varepsilon = 0$ .

### ПРЕОБРАЗОВАНИЯ КОНГРУЭНТНОСТИ И ЗАКОН ИНЕРЦИИ

Ранее в этой книге (в процессе исключения и задачах на собственные значения) основное внимание уделялось элементарным операциям, которые делают исходную матрицу проще. В каждом из этих случаев самым важным для нас было знать, какие свойства матрицы остаются неизменными. Когда кратное одной строки вычиталось из другой, список таких инвариантов был достаточно большим: нуль-пространство, пространство строк, ранг и определитель матрицы оставались теми же самими. В случае задачи на собственные значения, когда основной операцией было преобразование подобия  $A \rightarrow S^{-1}AS$  (или  $A \rightarrow M^{-1}AM$ ), неизменными оставались ее собственные значения (а также, согласно приложению В, ее жорданова форма). Теперь мы хотим задать аналогичный вопрос для симметрических матриц. Что представляют собой «элементарные операции» и их инварианты для  $x^T A x$ ?

Основной операцией над квадратичными формами является преобразование переменных. Новый вектор  $y$  выражается через вектор  $x$  с помощью некоторой невырожденной матрицы  $C$  по формуле  $x = Cy$ , в результате чего квадратичная форма принимает вид  $y^T C^T A C y$ . Таким образом, основной матричной операцией в теории квадратичных форм является **преобразование конгруэнтности**

$$A \rightarrow C^T A C \quad \text{с некоторой невырожденной матрицей } C. \quad (12)$$

Свойство симметричности при этом преобразовании сохраняется, поскольку матрица  $C^T A C$  симметрическая, если симметрической является матрица  $A$ . Но нас интересует вопрос, какие еще свойства матрицы  $A$  остаются неизменными при преобразовании конгруэнтности? Ответ на него дает **закон инерции** Сильвестра.

**6F.** Матрица  $C^T A C$  имеет столько же положительных, отрицательных и равных нулю собственных значений, что и матрица  $A$ .

Другими словами, знаки собственных значений (но не сами собственные значения) сохраняются при преобразовании конгруэнтности. При доказательстве этого утверждения мы для удобства предположим, что матрица  $A$  невырожденная. Тогда матрица  $C^T A C$  тоже невырожденная и, следовательно, не надо беспокоиться из-за нулевых собственных значений. (В противном случае, мы можем работать с невырожденными матрицами  $A + \varepsilon I$  или  $A - \varepsilon I$  и в конце устремить  $\varepsilon$  к нулю.)

Воспользуемся для доказательства 6F одним приемом из топологии, или, более точно, из теории гомотопий. Предположим, что матрица  $C$  связана с некоторой ортогональной матрицей  $Q$  непрерывной цепочкой матриц  $C(t)$ ,  $0 \leq t \leq 1$  (т. е.  $C(0) = C$  и  $C(1) = Q$ ), ни одна из которых не является вырожденной. Тогда собственные значения матрицы  $C^T(t) A C(t)$  непрерывно изменяются (при изменении  $t$  от 0 до 1) от собственных значений матрицы  $C^T A C$  до собственных значений  $Q^T A Q$ . Так как ни одна из матриц  $C(t)$  не является вырожденной, то ни одно из собственных значений матриц  $C^T(t) A C(t)$  не может стать равным нулю (и тем более перескочить через нуль). Следовательно, число собственных значений справа от нуля и число собственных значений слева от нуля одни и те же для матриц  $C^T A C$  и  $Q^T A Q$ . Отсюда, поскольку собственные значения матрицы  $A$  совпадают с собственными значениями подобной ей матрицы  $Q^{-1} A Q = Q^T A Q$ , вытекает, что столько же положительных и отрицательных собственных значений и у матрицы  $A$ <sup>1)</sup>. Утверждение 6F доказано.

**Пример 1.** Если  $A = I$ , то матрица  $C^T A C = C^T C$  положительно определена (возьмите, например,  $C = W$  из условия V). Все собственные значения единичной матрицы и матрицы  $C^T C$  положительны, что подтверждает закон инерции.

**Пример 2.** Если  $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ , то определитель матрицы  $C^T A C$  отрицателен:

$$\det C^T A C = (\det C^T) (\det A) (\det C) = - (\det C)^2 < 0.$$

<sup>1)</sup> В доказательстве нам необходима ортогональная матрица  $Q$ . Хороший способ выбора  $Q$  доставляет применение процесса Грама—Шмидта к столбцам матрицы  $C$ . Итак, пусть  $C = QR$ , где  $R$  — верхняя треугольная матрица с положительными диагональными элементами. Тогда в качестве непрерывной цепочки, связывающей  $C$  с  $Q$ , можно выбрать семейство матриц  $C(t) = tQ + (1-t)QR = Q[tI + (1-t)R]$ . Действительно, при любом  $t \in [0, 1]$  матрица  $C(t)$  невырожденная, так как  $Q$  — ортогональная матрица, а второй множитель  $tI + (1-t)R$  является верхней треугольной матрицей с положительными диагональными элементами.

Поскольку этот определитель является произведением собственных значений, то одно собственное значение матрицы  $C^T A C$  должно быть положительным, а другое отрицательным (как и у матрицы  $A$ ), что опять согласуется с законом инерции.

**Пример 3.** Следующее применение закона инерции является очень важным.

**6G.** У любой симметрической матрицы  $A$  знаки ведущих элементов согласованы со знаками собственных значений. Точнее число положительных, отрицательных и равных нулю собственных значений матрицы  $A$  равно числу ее положительных, отрицательных и равных нулю ведущих элементов соответственно.

Известно, что  $A = Q \Lambda Q^T$ , где  $Q$  — ортогональная матрица,  $\Lambda$  — диагональная матрица с собственными значениями матрицы  $A$  по диагонали, а формула Холецкого имеет вид

$$A = LDL^T, \text{ или } D = L^{-1} A (L^T)^{-1} = (L^{-1} Q) \Lambda (L^{-1} Q)^T. \quad (13)$$

Выберем  $C = (L^{-1} Q)^T$ . Тогда, согласно закону инерции, собственные значения матрицы  $D = C^T A C$  (одновременно являющиеся ведущими элементами для  $A$ ) имеют те же знаки, что и собственные значения матрицы  $\Lambda$ . Если же в процессе исключения (сохраняющем симметричность) требовались перестановки строк и столбцов, то те же рассуждения нужно применить к матрице  $P A P^T$ .

**Упражнение 6.3.3.** Используя матрицы  $C = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$  и  $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ , показать, что матрица  $C^T A C$  имеет собственные значения тех же знаков, что и  $A$ . Можете ли вы сконструировать цепочку невырожденных матриц  $C(t)$ , осуществляющих переход от  $C$  к ортогональной матрице  $Q$ ?

**Упражнение 6.3.4.** Используя знаки ведущих элементов, показать, что матрица

$$A = \begin{bmatrix} 1 & 3 & 0 \\ 3 & 8 & 7 \\ 0 & 7 & 6 \end{bmatrix}$$

имеет одно отрицательное и два положительных собственных значения. Затем найти ведущие элементы матрицы  $A + 2I$  и показать, что добавление двойки к собственным значениям матрицы  $A$  делает их все положительными. Следовательно, отрицательное собственное значение матрицы  $A$  должно принадлежать интервалу  $(-2, 0)$ .

Последнее упражнение указывает путь вычисления собственных значений, и это будет первый практический метод, который мы укажем. (Он был одним из основных методов около двадцати лет назад и до сих пор остается одним из самых эффективных.)

Идея метода состоит в том, чтобы последовательно разрезать интервал неопределенности; если мы знаем, что  $\lambda$  лежит в интервале  $(-2, 0)$ , то на следующем шаге надо выяснить, будет ли наименьшее собственное значение больше или меньше  $-1$ . Работая с матрицей  $A + I$ , собственные значения которой на единицу больше собственных значений  $A$ , мы устанавливаем, что один из ее ведущих элементов отрицателен и, следовательно, искомое собственное значение  $\lambda$  меньше чем  $-1$ . Следующий шаг заключается в исследовании матрицы  $A + 3I/2$ , у которой опять оказывается один отрицательный ведущий элемент. Поэтому  $\lambda$  должно принадлежать интервалу  $(-2, -3/2)$ . На каждом шаге граница уточняется на один двоичный разряд.

Для больших симметрических матриц этот метод является слишком дорогим, так как для плотных матриц каждый шаг требует  $n^3/6$  действий для вычисления ведущих элементов. Но если до начала вычислений матрица может быть преобразована к трехдиагональной, как это делается в § 7.3, то каждое деление интервала неопределенности пополам будет требовать только  $2n$  умножений. Исключения осуществляются очень быстро, и поиск собственных значений путем последовательного уточнения границ интервала неопределенности становится чрезвычайно простым. Это так называемый *метод Гивенса*.

### ОБОБЩЕННАЯ ЗАДАЧА НА СОБСТВЕННЫЕ ЗНАЧЕНИЯ

Я не уверен относительно экономики, но физика, техника и статистика обычно бывают столь «любезны», что в рассматриваемых ими задачах на собственные значения матрицы оказываются симметрическими. Однако в этих задачах может оказаться *две матрицы, а не одна*, т. е. вместо задачи  $Ax = \lambda x$  мы сталкиваемся с задачей  $Ax = \lambda Bx$ . В качестве примера такой ситуации рассмотрим задачу о движении двух неодинаковых масс.

Согласно закону Ньютона,  $F_1 = m_1 a_1$  для первой массы и  $F_2 = m_2 a_2$  для второй. Если мы выберем ту же самую систему пружин, что и на стр. 249, то силы  $F_1$  и  $F_2$  останутся теми же самыми. Физику процесса мы можем описать следующим образом: если массы смещены на расстояния  $v$  и  $w$ , как на рис. 6.3,

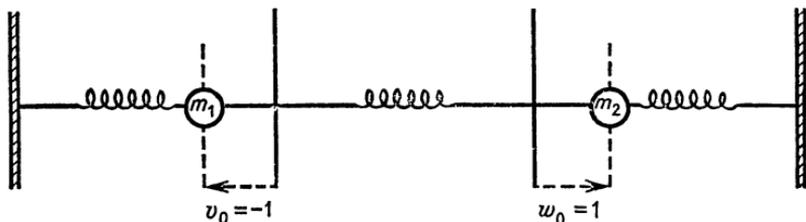


Рис. 6.3. Колебательная система с неодинаковыми массами.

то пружины тянут их назад соответственно с силами  $F_1 = -2v + w$  и  $F_2 = v - 2w$ . Новое и важное явление наблюдается в левой части уравнений, где возникают уже две массы:

$$m_1 \frac{d^2 v}{dt^2} = -2v + w, \quad \text{или} \quad \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \frac{d^2 u}{dt^2} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} u.$$

$$m_2 \frac{d^2 w}{dt^2} = v - 2w,$$

Когда массы равны, т. е.  $m_1 = m_2 = 1$ , мы возвращаемся к старой системе  $u'' = Au$ . Теперь же, поскольку имеется «матрица масс»  $B$ , мы приходим к системе  $Bu'' = Au$ . Когда ищется экспоненциальное решение  $e^{i\omega t}x$ , мы приходим к задаче на собственные значения вида

$$B(i\omega)^2 e^{i\omega t}x = Ae^{i\omega t}x.$$

Сокращая на  $e^{i\omega t}$  и заменяя  $(i\omega)^2$  на  $\lambda$ , получим

$$Ax = \lambda Bx, \quad \text{или} \quad \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} x = \lambda \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} x. \quad (14)$$

Решение этой системы существует только в том случае, когда матрица  $A - \lambda B$  вырожденная. Поэтому  $\lambda$  должно быть корнем уравнения  $\det(A - \lambda B) = 0$ . При этом специальный выбор  $B = I$  возвращает нас к обычному уравнению  $\det(A - \lambda I) = 0$ .

Рассмотрим пример, когда  $m_1 = 1$  и  $m_2 = 2$ . Тогда

$$\det(A - \lambda B) = \det \begin{bmatrix} -2 - \lambda & 1 \\ 1 & -2 - 2\lambda \end{bmatrix} = 2\lambda^2 + 6\lambda + 3$$

и, следовательно,

$$\lambda = \frac{-3 \pm \sqrt{3}}{2}.$$

Оба собственных значения отрицательны, и двумя естественными частотами являются величины  $\omega_i = \sqrt{-\lambda_i}$ . Соответствующие собственные векторы вычисляются обычным способом:

$$(A - \lambda_1 B)x_1 = \begin{bmatrix} \frac{-1 - \sqrt{3}}{2} & 1 \\ 1 & 1 - \sqrt{3} \end{bmatrix} x_1 = 0, \quad x_1 = \begin{bmatrix} \sqrt{3} - 1 \\ 1 \end{bmatrix},$$

$$(A - \lambda_2 B)x_2 = \begin{bmatrix} \frac{-1 + \sqrt{3}}{2} & 1 \\ 1 & 1 + \sqrt{3} \end{bmatrix} x_2 = 0, \quad x_2 = \begin{bmatrix} 1 + \sqrt{3} \\ -1 \end{bmatrix}.$$

Эти собственные векторы задают «гармоники». При меньшей частоте  $\omega_1$  обе массы осциллируют вместе (смещения в одинаковую сторону) с тем лишь отличием, что первая масса сдвигается

только на  $\sqrt{3}-1 \approx 0.73$ , а вторая (большая) в силу соответствующих причин имеет большую амплитуду колебаний. Для большей частоты  $\omega_2$  компоненты вектора  $x_2$  имеют противоположные знаки и, следовательно, массы двигаются в противоположных направлениях. При этом меньшая масса отклоняется от точки равновесия значительно сильнее.

Соответствующую теорию гораздо легче объяснить, если предварительно разложить  $B$  в произведение  $W^T W$  (как и в примере, здесь предполагается, что матрица  $B$  положительно определена). Тогда после замены переменных  $y = Wx$  уравнение  $Ax = \lambda Bx = \lambda W^T Wx$  приходит к виду  $AW^{-1}y = \lambda W^T y$ . Теперь, если обозначить матрицу  $W^{-1}$  через  $C$  и умножить последнее уравнение на матрицу  $(W^T)^{-1} = C^T$ , мы придем к стандартной задаче на собственные значения

$$C^T A C y = \lambda y \quad (15)$$

с одной матрицей  $C^T A C$ . Собственные значения этой задачи совпадают с собственными значениями исходной задачи  $Ax = \lambda Bx$ , а собственные векторы связаны соотношением  $y_j = Wx_j$ . Свойства симметрической матрицы  $C^T A C$  непосредственно определяют соответствующие свойства задачи  $Ax = \lambda Bx$ .

(1) Все ее собственные значения вещественны.

(2) Согласно закону инерции, ее собственные значения имеют те же самые знаки, что и собственные значения матрицы  $A$ .

(3) Собственные векторы  $y_j$  ортонормированы, что соответствует « $B$ -ортонормированности» собственных векторов  $x_j$  исходной задачи:

$$x_i^T B x_j = x_i^T W^T W x_j = y_i^T y_j = \begin{cases} 1, & \text{если } i = j, \\ 0, & \text{если } i \neq j. \end{cases} \quad (16)$$

Если векторы  $x_j$  являются столбцами матрицы  $S$ , то (16) означает, что  $S^T B S = I$ . Аналогично

$$x_i^T A x_j = x_i^T \lambda_j B x_j = \begin{cases} \lambda_j, & \text{если } i = j, \\ 0, & \text{если } i \neq j. \end{cases} \quad (17)$$

В матричных обозначениях это означает, что  $S^T A S = \Lambda$ , и выражает следующий замечательный факт: *матрицы  $A$  и  $B$  одновременно диагонализуются преобразованием конгруэнтности  $S$ .*

Геометрический смысл последнего факта не очень понятен. В случае положительной определенности обеих матриц две по-

1) Более быстрым способом получения задачи с одной матрицей был бы переход к уравнению  $B^{-1}Ax = \lambda x$ . Но матрица  $B^{-1}A$  не является симметрической, и мы предпочитаем воздействовать «половинками» матрицы  $B^{-1}$  с разных сторон на матрицу  $A$ , чтобы получить симметрическую матрицу  $C^T A C$ .

верхности  $x^T A x = 1$  и  $x^T B x = 1$  являются эллипсоидами. Очевидно, преобразование  $x = Sz$  задает такие новые координаты (так как  $S$  неортогональная матрица, то это преобразование не является чистым вращением), что эти два эллипсоида становятся «правильно расположенными»:

$$\begin{aligned} x^T A x &= z^T S^T A S z = \lambda_1 z_1^2 + \dots + \lambda_n z_n^2 = 1, \\ x^T B x &= z^T S^T B S z = z_1^2 + \dots + z_n^2 = 1. \end{aligned} \quad (18)$$

Второй эллипсоид в действительности является сферой! Конечно, основной момент теории постепенно проясняется и его легко сформулировать: в случае положительно определенной матрицы  $B$  обобщенная задача на собственные значения  $Ax = \lambda Bx$  ведет себя в точности как задача  $Ax = \lambda x$ .

**Упражнение 6.3.5.** Для рассмотренного примера с  $m_1 = 1$  и  $m_2 = 2$  проверить, что гармоники  $B$ -ортогональны, т. е.  $x_1^T B x_2 = 0$ . Какими будут коэффициенты  $a_i$  результирующего движения  $u = a_1 \cos \omega_1 t x_1 + a_2 \cos \omega_2 t x_2$ , если обе массы сместить на одинаковое расстояние, равное единице ( $v_0 = -1$  и  $\omega_0 = 1$ ), а затем отпустить? Какое максимальное расстояние от положения равновесия достигается первой массой (когда два косинуса соответственно равны  $+1$  и  $-1$ )?

**Упражнение 6.3.6.** Найти собственные значения и собственные векторы задачи

$$\begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix} x = \frac{\lambda}{18} \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix} x.$$

**Упражнение 6.3.7.** Если симметрические матрицы  $A$  и  $B$  порядка 2 не являются определенными, то задача  $Ax = \lambda Bx$  может иметь комплексные собственные значения. Построить соответствующий пример.

## § 6.4. ПРИНЦИПЫ МИНИМУМА И ОТНОШЕНИЕ РЕЛЕЯ

В этом параграфе, приближаясь уже к концу книги, мы впервые откажемся от линейных уравнений. Неизвестный вектор  $x$  не будет даваться как решение уравнений  $Ax = b$  или  $Ax = \lambda x$ . Вместо этого он будет определен с помощью принципа минимума.

Удивительно, как много физических законов может быть выражено с помощью принципов минимума. Такой простой факт, что тяжелые жидкости опускаются на дно, является следствием минимизации их потенциальной энергии. Или когда вы сидите на стуле, или лежите на кровати, пружины сами занимают такое положение, чтобы опять потенциальная энергия была минимальна. Конечно, имеются и более серьезные примеры: здание похоже на очень сложный стул, выдерживающий свой собственный вес, и основным принципом строительной техники является минимизация полной энергии. В физике имеются «лагранжианы»

и «интегралы действия»; соломинка в стакане с водой выглядит как бы сломанной, поскольку свет выбирает такой путь, чтобы достичь глаза как можно скорее<sup>1)</sup>.

Мы сразу же хотим сказать, что «энергии» всех этих процессов есть не что иное, как *положительно определенные квадратичные формы*. Очевидно, что производная квадратичной формы линейна. Поэтому минимизация приводит нас опять к привычным линейным уравнениям, когда мы приравниваем нулю первые производные. Нашей ближайшей целью в настоящем параграфе будет *нахождение задачи минимизации, соответствующей уравнению  $Ax=b$ , а также другой задачи минимизации, соответствующей уравнению  $Ax=\lambda x$* . Мы собираемся проделать в конечномерном случае то, что вариационное исчисление проделывает в непрерывном случае, где обращение в нуль первых производных приводит к дифференциальным уравнениям (уравнениям Эйлера). При этом в каждой конкретной ситуации мы вольны выбирать, искать решение линейного уравнения или задачи на минимум квадратичной формы, и, как это будет показано в следующем параграфе, для многих задач вторую возможность игнорировать не следует.

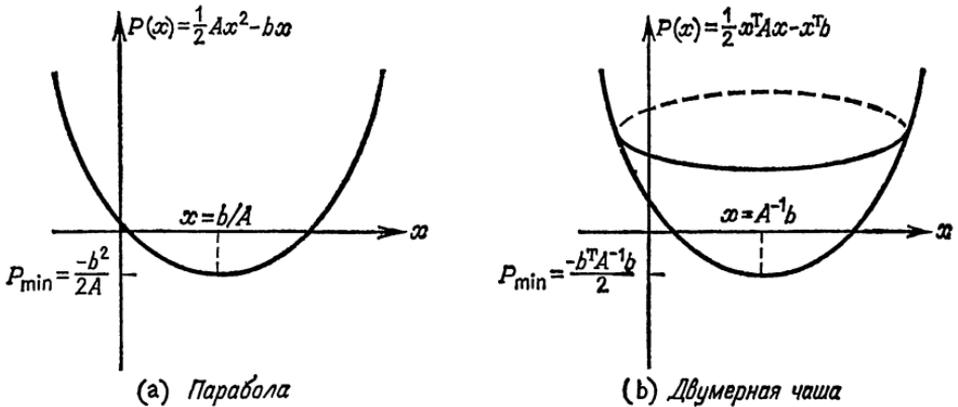
Первый шаг очень прост: требуется найти «параболу»  $P$ , минимум которой достигается в точке, являющейся решением уравнения  $Ax=b$ . Если  $A$  является скаляром, то это сделать очень легко:

$$P(x) = \frac{1}{2} Ax^2 - bx \quad \text{и} \quad \frac{dP}{dx} = Ax - b.$$

Эта парабола достигает минимума только в случае, когда  $A$  положительно и, значит, парабола открыта вверх; тогда обращение в нуль ее первой производной дает уравнение  $Ax=b$  (рис. 6.4). В многомерном случае эта парабола переходит в параболоид, но формула для  $P$  и условие положительности для существования минимума остаются неизменными.

**6Н.** Если  $A$  — симметрическая и положительно определенная матрица, то квадратичная форма  $P(x) = \frac{1}{2} x^T Ax - x^T b$  достигает своего минимума в точке, являющейся решением системы  $Ax=b$ .

<sup>1)</sup> Я уверен, что растения и люди развиваются в соответствии с некоторыми принципами минимума. Возможно, и сама цивилизация основана на законе наименьшего действия. Открытие таких законов является фундаментальным шагом в переходе от наблюдений к объяснению явлений. В общественных науках и биологии предстоит найти еще некоторые законы такого типа.

Рис. 6.4. Минимум квадратичной функции  $P(x)$ .

**Доказательство.** Предположим, что  $x$  является решением системы  $Ax=b$ . Тогда для любого вектора  $y$  имеем

$$\begin{aligned}
 P(y) - P(x) &= \frac{1}{2} y^T A y - y^T b - \frac{1}{2} x^T A x + x^T b = \\
 &= \frac{1}{2} y^T A y - y^T A x + \frac{1}{2} x^T A x = \\
 &= \frac{1}{2} (y-x)^T A (y-x).
 \end{aligned} \tag{19}$$

Так как матрица  $A$  положительно определена, эта величина нигде не принимает отрицательных значений и равна нулю, только когда  $y-x=0$ . Во всех других точках  $P(y)$  больше, чем  $P(x)$ , и, следовательно, минимум достигается в точке  $x$ .

**Упражнение 6.4.1.** Рассмотреть систему

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 4 \end{bmatrix}.$$

Построить соответствующую квадратичную форму  $P(x_1, x_2, x_3)$ , вычислить ее частные производные  $\partial P/\partial x_i$  и показать, что они обращаются в нуль только на решении исходной системы.

**Упражнение 6.4.2.** Найти минимум (если он существует) функций  $P_1 = x^2/2 + xy + y^2 - 3y$  и  $P_2 = x^2/2 - 3y$ . Какая матрица соответствует функции  $P_2$ ?

**Упражнение 6.4.3.** Другая квадратичная форма, которая достигает минимума на решении системы  $Ax=b$ , задается равенством

$$Q(x) = \frac{1}{2} \|Ax - b\|^2 = \frac{1}{2} x^T A^T A x - x^T A^T b + \frac{1}{2} b^T b.$$

Сравнивая  $Q$  с  $P$  и игнорируя константу  $b^T b/2$ , ответить на вопрос, какую систему уравнений для определения минимума мы в этом случае получаем. Как называются эти уравнения в методе наименьших квадратов?

Наша следующая цель — найти задачу минимизации, соответствующую уравнению  $Ax = \lambda x$ . Это не так легко. Функция, которую мы должны построить, не может быть просто квадратичной, так как тогда ее дифференцирование приведет к линейным уравнениям, а задача на собственные значения в сущности своей нелинейна. Выход из положения дает нам переход к *отношению квадратичных функций*, и нам необходимо лишь так называемое *отношение Релея*

$$R(x) = \frac{x^T Ax}{x^T x}.$$

Мы перейдем непосредственно к основной теореме.

**6.1. Принцип Релея.** *Отношение  $R(x)$  минимизируется первым собственным вектором  $x_1$ , и это минимальное значение равно наименьшему собственному значению  $\lambda_1$  матрицы  $A$ :*

$$R(x_1) = \frac{x_1^T A x_1}{x_1^T x_1} = \frac{x_1^T \lambda_1 x_1}{x_1^T x_1} = \lambda_1. \quad (20)$$

Проанализируем сначала сделанное утверждение с геометрической точки зрения. Представьте себе, что мы зафиксировали числитель равным единице и стараемся сделать знаменатель как можно больше. Уравнение  $x^T Ax = 1$  для числителя определяет, по крайней мере для положительно определенной матрицы  $A$ , эллипсоид. Тогда, так как знаменатель равен  $x^T x = \|x\|^2$ , мы должны искать точку на поверхности эллипсоида, которая расположена дальше всех от начала координат (иначе говоря, вектор  $x$  должен иметь наибольшую длину). Но из нашего предыдущего изучения эллипсоида вытекает, что направление его наибольшей оси (на которой находится искомая точка) совпадает с направлением первого собственного вектора.

Алгебраически в этом очень легко убедиться (даже без требования положительной определенности  $A$ ), если мы диагонализуем матрицу  $A$  с помощью ортогональной матрицы  $Q$ :  $Q^{-1} A Q = \Lambda$ , где  $Q^T = Q^{-1}$ . После преобразования  $x = Qy$  отношение Релея принимает более простой вид

$$R(x) = \frac{(Qy)^T A (Qy)}{(Qy)^T (Qy)} = \frac{y^T \Lambda y}{y^T y} = \frac{\lambda_1 y_1^2 + \dots + \lambda_n y_n^2}{y_1^2 + \dots + y_n^2}.$$

Минимум этого отношения, очевидно, достигается в точке с координатами  $y_1 = 1$  и  $y_2 = y_3 = \dots = y_n = 0$ . В этой точке отношение равно  $\lambda_1$ . В любой другой точке отношение будет больше, так как  $\lambda_1$  — наименьшее из всех  $\lambda$ ,

$$\lambda_1 (y_1^2 + y_2^2 + \dots + y_n^2) \leq (\lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2).$$

Поделив последнее неравенство на выражение, стоящее в первых скобках, получим  $\lambda_1 \leq R(x)$ . Таким образом, минимум достигается на первом собственном векторе или, точнее, на любом векторе  $x = cx_1$ , кратном ему, так как «растягивающий» множитель  $c$  не изменяет отношения, т. е.

$$R(cx) = \frac{(cx)^T A (cx)}{(cx)^T (cx)} = \frac{c^2 x^T A x}{c^2 x^T x} = R(x).$$

Принцип Релея означает, что для каждого вектора  $x$  отношение  $R(x)$  дает верхнюю границу для  $\lambda_1$ . Заметим, что для оценки  $\lambda_1$  сверху нам не нужны собственные векторы и переход к новому вектору  $y$ . Это было полезно при доказательстве, но в приложениях мы можем выбирать совершенно произвольный вектор  $x$ .

**Упражнение 6.4.4.** Каково отношение  $R(x)$  для произвольной матрицы  $A$  и конкретного вектора  $x = (1, 0, \dots, 0)$ . Показать, что  $\lambda_1$  не превосходит наименьшего из диагональных элементов  $a_{ii}$  матрицы  $A$ .

**Упражнение 6.4.5.** Для матрицы  $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$  найти вектор  $x$ , который дает меньшее значение отношения  $R(x)$ , чем оценка  $\lambda_1 \leq 2$ , полученная с помощью диагональных элементов матрицы  $A$ . Каково минимальное значение  $R(x)$ ?

**Упражнение 6.4.6.** Используя отношение Релея, показать, что для положительной определенной матрицы  $B$  наименьшее собственное значение матрицы  $A+B$  больше наименьшего собственного значения матрицы  $A$ .

**Упражнение 6.4.7.** Пусть  $\lambda_1$  — наименьшее собственное значение матрицы  $A$ , а  $\mu_1$  — наименьшее собственное значение матрицы  $B$ . Показать, что наименьшее собственное значение  $\theta_1$  матрицы  $A+B$  по крайней мере не меньше, чем  $\lambda_1 + \mu_1$ . (Указание: подставить соответствующий собственный вектор  $x$  в отношение Релея.)

Последние два упражнения — это, возможно, самые типичные и наиболее важные результаты, которые легко получить из отношения Релея, но очень трудно непосредственно из исходных уравнений.

Применение отношения Релея не ограничивается первым собственным значением  $\lambda_1$  и соответствующим ему собственным вектором  $x_1$ . В действительности легко видеть, что максимум отношения

$$R(x) = \frac{\lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2}{y_1^2 + y_2^2 + \dots + y_n^2} \quad (21)$$

достигается на другом конце последовательности единичных векторов  $y$ , а именно в точке с координатами  $y_n = 1$  и  $y_1 = y_2 = \dots = y_{n-1} = 0$ . Эта точка соответствует последнему собственному вектору  $x = x_n$ , а само максимальное значение отношения равно  $\lambda_n$ . Следовательно, каждое значение отношения  $R(x)$  является не

только верхней границей для  $\lambda_1$ , но и нижней границей для  $\lambda_n$ <sup>1)</sup>. В заключение заметим, что промежуточные собственные векторы  $x_2, \dots, x_{n-1}$  являются седловыми точками отношения Релея.

**Упражнение 6.4.8.** Вычислить частные производные отношения (21) и показать, что точка с координатами  $y_2 = 1$  и  $y_1 = y_3 = \dots = y_n = 0$  является его стационарной точкой. Найти максимум отношения

$$R(x) = (x_1^2 - x_1 x_2 + x_2^2) / (x_1^2 + x_2^2).$$

### ПРИНЦИП МИНИМАКСА ДЛЯ СОБСТВЕННЫХ ЗНАЧЕНИЙ

Трудность с седловыми точками заключается в том, что для произвольного вектора  $x$  мы не можем сказать, будет отношение Релея  $R(x)$  больше или меньше любого из промежуточных собственных значений  $\lambda_2, \dots, \lambda_{n-1}$ . Для приложений существует принцип экстремума (минимума или максимума), который действительно оказывается полезным. Поэтому мы рассмотрим такой принцип, подразумевая, что максимум или минимум будет достигаться на  $j$ -м собственном векторе  $x_j$ <sup>2)</sup>. Руководящую идею дает нам основное свойство симметрической матрицы: вектор  $x_j$  ортогонален к другим собственным векторам матрицы  $A$ . Потребуем, чтобы векторы  $x$ , реализующие принцип минимума, были ортогональны к первым собственным векторам  $x_1, \dots, x_{j-1}$ , т. е.  $y_1 = y_2 = \dots = y_{j-1} = 0$ . На остальные параметры  $y_j, \dots, y_n$  не накладывается ограничений и отношение Релея принимает вид

$$\frac{\lambda_j y_j^2 + \lambda_{j+1} y_{j+1}^2 + \dots + \lambda_n y_n^2}{y_j^2 + y_{j+1}^2 + \dots + y_n^2}.$$

Его минимум равен  $\lambda_j$  и достигается, когда  $y_j = 1$ , а  $y_{j+1} = \dots = y_n = 0$ .

**6 J.** В предположении, что вектор  $x$  ортогонален к собственным векторам  $x_1, \dots, x_{j-1}$ , отношение  $R(x)$  минимизируется следующим собственным вектором  $x_j$  и его минимальное значение равно  $\lambda_j$ :

$$\lambda_j = \min R(x) \text{ при условии, что } x^T x_1 = \dots = x^T x_{j-1} = 0. \quad (22)$$

Аналогично  $x_j$  максимизирует  $R(x)$  на множестве векторов  $x$ , перпендикулярных векторам  $x_{j+1}, \dots, x_n$ .

<sup>1)</sup> Используя рассуждения из упр. 6.4.4, легко показать, что ни один диагональный элемент не превосходит  $\lambda_n$ .

<sup>2)</sup> Эта тема несколько специальна, и метод Релея—Ритца конечных элементов основывается на тех принципах минимума, которые уже найдены. Поэтому ничто не мешает нам перейти прямо к § 6.5.

Пользуясь сформулированным принципом, теоретически возможно определить все собственные значения в возрастающем порядке. Первое собственное значение  $\lambda_1$  является абсолютным минимумом для  $R(x)$ , который достигается на векторе  $x_1$ . Следующее собственное значение  $\lambda_2$  является минимальным значением  $R(x)$  в предположении, что векторы  $x$  ортогональны  $x_1$ . При этом для любого  $x$ , ортогонального к  $x_1$  и не совпадающего с  $x_2$ , значение  $R(x)$  больше, чем  $\lambda_2$ .

**Пример.** Матрица

$$A = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 1 \end{bmatrix}$$

положительно полуопределена, и вектор  $x_1 = (1, 1, 1, 1)$  лежит в ее нуль-пространстве и является собственным вектором, соответствующим собственному значению  $\lambda_1 = 0$ . Для того чтобы найти некоторую верхнюю грань для  $\lambda_2$ , выберем некоторый пробный вектор  $x$ , ортогональный к  $x_1$ , и вычислим значение  $R(x)$ :

пусть  $x = (1, 1, -1, -1)$  и  $\frac{x^T A x}{x^T x} = 1$ , а следовательно,  $\lambda_2 \leq 1$ .

Здесь только одна проблема: если мы не знаем точно вектор  $x_1$  (а знаем мы его в очень редких случаях), то мы не знаем, будет ли вектор  $x$  ортогонален к нему. В задачах минимизации с ограничениями само ограничение должно быть известно!

Нас интересует, существует ли принцип экстремума для  $\lambda_2$  и  $x_2$ , который не требовал бы знания вектора  $x_1$ . Ответ здесь очень тонкий. Предположим, что вектор  $x_1$  неизвестен и в качестве пробных векторов  $x$  выбираются векторы, ортогональные к некоторому, вообще говоря, произвольному вектору  $z$ . Если  $z = x_1$ , то минимальное значение  $R(x)$  будет равно  $\lambda_2$ . В гораздо более правдоподобной ситуации, когда вектор  $z$  отличен от  $x_1$ , относительно искомого минимума можно сказать лишь следующее: **он не больше, чем  $\lambda_2$** . Иначе говоря, для любого  $z$

$$\min_{x^T z = 0} R(x) \leq \lambda_2. \quad (23)$$

**Доказательство.** Любая ненулевая комбинация первых двух собственных векторов  $x = \alpha x_1 + \beta x_2$  будет ортогональна к вектору  $z$ , если наложить определенные условия на коэффициенты  $\alpha$  и  $\beta$ . Тогда для любой такой комбинации

$$R(x) = \frac{(\alpha x_1 + \beta x_2)^T A (\alpha x_1 + \beta x_2)}{(\alpha x_1 + \beta x_2)^T (\alpha x_1 + \beta x_2)} = \frac{\lambda_1 \alpha^2 + \lambda_2 \beta^2}{\alpha^2 + \beta^2} \leq \lambda_2.$$

Отсюда, так как найден вектор  $x$ , для которого  $R(x) \leq \lambda_2$ , следует, что минимум  $R(x)$  обязательно будет меньше, чем  $\lambda_2$ . Этот факт непосредственно приводит нас к «принципу максимина».

**6 К.** Если мы минимизируем  $R(x)$  на множестве векторов, подчиненных одному условию  $x^T z = 0$ , и выбираем  $z$  так, чтобы максимизировать этот минимум, то получим  $\lambda_2$ , т. е.

$$\lambda_2 = \max_z \left[ \min_{x^T z = 0} R(x) \right]. \quad (24)$$

Величина в скобках, согласно (23), меньше или равна  $\lambda_2$  и при специальном выборе  $z = x_1$  точно равна  $\lambda_2$ .

Геометрически принцип максимина имеет естественную интерпретацию. Предположим, что эллипсоид разрезан плоскостью, проходящей через начало координат. Тогда сечение эллипсоида опять будет эллипсоидом меньшей на единицу размерности. При этом неравенство (23) просто означает, что наибольшая ось нового эллипсоида по крайней мере не меньше второй по величине оси исходного эллипсоида (на рис. 6.5 это соответствует неравенству

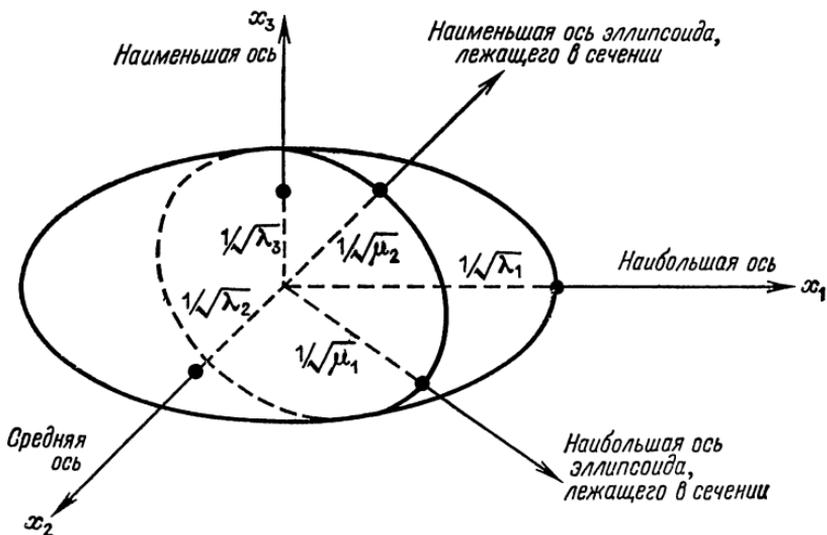


Рис. 6.5. Иллюстрация принципов максимина и минимакса:  $\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \mu_2 \leq \lambda_3$ .

$\mu_1 \leq \lambda_2$ ). При специальном выборе разрезающей плоскости наибольшая ось возникающего эллипсоида будет точно равна второй по величине оси исходного эллипсоида. Это достигается выбором плоскости, перпендикулярной направлению  $z = x_1$  наибольшей оси исходного эллипсоида. В этом случае  $\mu_1 = \lambda_2$ .

Геометрическая картина сразу может быть переведена на язык матричной алгебры.

**6 L.** Пусть задана симметрическая матрица  $A$  с собственными значениями  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  и подматрица  $A_{n-1}$  получена из матрицы  $A$  путем отбрасывания ее последних строки и столбца. Тогда если  $\mu_1$  — наименьшее собственное значение матрицы  $A_{n-1}$ , то  $\lambda_1 \leq \mu_1 \leq \lambda_2$ .

Сначала заметим, что  $\lambda_1$  является абсолютным минимумом отношения  $R(x)$ , а  $\mu_1$  — минимальным значением  $R(x)$  на множестве векторов, последняя компонента которых равна нулю. Этот минимум с ограничением не может быть меньше абсолютного минимума, т. е.  $\lambda_1 \leq \mu_1$ . С другой стороны, условие равенства нулю последней компоненты  $x$  эквивалентно условию  $x^T z = 0$  при специальном выборе  $z = (0, \dots, 0, 1)$ . Отсюда, согласно (23), имеем неравенство  $\mu_1 \leq \lambda_2$ .

**Пример 1.** Собственные значения матрицы  $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$  равны единице и трем, в то время как собственное значение подматрицы, получаемой отбрасыванием последних строки и столбца, равно двум.

**Пример 2.** Собственные значения матрицы

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

равны  $2 - \sqrt{2}$ ,  $2$ ,  $2 + \sqrt{2}$ . Если отбросить последнюю строку и столбец  $A$ , то наименьшее собственное значение полученной подматрицы возрастет до единицы, но не превысит собственное значение  $\lambda_2$  матрицы  $A$ :  $2 - \sqrt{2} < 1 < 2$ .

**Пример 3.** Если  $B$  положительно определена, то, согласно принципу Релея,

$$\lambda_1(A+B) = \min \frac{x^T(A+B)x}{x^T x} > \min \frac{x^T A x}{x^T x} = \lambda_1(A).$$

Эта задача ставилась в упр. 6.4.6. Теперь принцип максимина (24) позволяет получить такое неравенство для вторых собственных значений:  $\lambda_2(A+B) > \lambda_2(A)$ , так как  $R(x)$  опять возрастает при добавлении к  $A$  матрицы  $B$ .

**Пример 4.** Для заданной осциллирующей системы масс и пружин предположим, что одна из них вынужденно остается в

положении равновесия. Тогда наименьшая частота системы возрастает, но при этом не превосходит первоначального собственного значения  $\lambda_2$ .

**Упражнение 6.4.9.** Найти собственные значения матрицы

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

и показать, используя неравенство  $\lambda_1 \leq \mu_1 \leq \lambda_2$ , что независимо от того, какая пара строка-столбец вычеркнута, наименьшее собственное значение полученной таким образом матрицы будет равно единице. Что при этом происходит с эллипсоидом  $x^T A x = 1$ ?

**Упражнение 6.4.10.** Обобщить принцип максимина на случай  $j$  различных ограничений:

$$\max_{z_1, \dots, z_j} \left[ \min_{x^T z_1 = 0, \dots, x^T z_j = 0} R(x) \right] = \lambda_{j+1}. \quad (25)$$

Полагая  $j=2$ , вывести неравенство для наименьшего собственного значения подматрицы  $A$ , которая получается отбрасыванием последних двух строк и столбцов.

Наступило время обратить последние из сформулированных теорем и установить **принцип минимакса**. Это означает, что мы сначала должны максимизировать отношение Релея, а затем найти минимальное значение этого максимума.

Существует несколько подходов к решению этой задачи в зависимости от того, какие собственные значения нам нужны. Например, чтобы найти наибольшее собственное значение  $\lambda_n$ , мы просто должны максимизировать  $R(x)$ . Но предположим, что нам нужно найти  $\lambda_2$ . Так как отношение Релея имеет вид

$$\frac{\lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2}{y_1^2 + y_2^2 + \dots + y_n^2},$$

то очевидный способ нахождения  $\lambda_2$  как максимума — это наложить условие, что  $y_3 = y_4 = \dots = y_n = 0$ . Эти  $n-2$  ограничения приводят к двумерному подпространству, порожденному первым и вторым собственными векторами  $x_1$  и  $x_2$  матрицы  $A$ . Максимум  $R(x)$  в этом двумерном пространстве равен  $\lambda_2$ , но для этого нужны векторы  $x_1$  и  $x_2$ , а они неизвестны.

Когда подобная проблема возникла в принципе максимина, она была разрешена выбором некоторого, вообще говоря, произвольного ограничения  $x^T z = 0$ . Мы применим ту же идею, максимизируя  $R(x)$  в некотором произвольном двумерном подпространстве. Мы не можем узнать, будет ли это подпространство содержать какие-либо собственные векторы, но это маловероятно.

Тем не менее максимум отношения Релея позволяет получить оценку для  $\lambda_2$ . Если ранее минимум отношения  $R(x)$  при условии  $x^T z = 0$  давал оценку для  $\lambda_2$  снизу, то теперь его максимум по рассматриваемому подпространству дает для  $\lambda_2$  оценку сверху. Иначе говоря, для любого двумерного подпространства  $S_2$  из  $\mathbb{R}^n$

$$\max_{z \in S_2} R(x) \geq \lambda_2. \quad (26)$$

**Доказательство.** Любое из подпространств  $S_2$  всегда содержит вектор  $x$ , ортогональный к вектору  $x_1$ , и для него  $R(x) \geq \lambda_2$ . Отсюда сразу следует принцип минимакса.

**6М.** Если мы максимизируем  $R(x)$  по всем возможным двумерным подпространствам  $S_2$ , то минимально возможное значение этих максимумов равно  $\lambda_2$ :

$$\lambda_2 = \min_{S_2} \left[ \max_{x \in S_2} R(x) \right]. \quad (27)$$

Величина в скобках всегда больше или равна  $\lambda_2$ , а для частного выбора, когда  $S_2$  порождается первым и вторым собственными векторами матрицы  $A$ , равна  $\lambda_2$ .

**Упражнение 6.4.11.** Для матрицы  $A$  из примера 2 вычислить собственное значение  $\mu_2$  ее подматрицы порядка два, стоящей в левом верхнем углу, и сравнить его с  $\lambda_2$  и  $\lambda_3$ .

Мы закончим настоящий раздел двумя замечаниями. Я надеюсь, что вы убедитесь в их справедливости на основании интуитивных соображений, без детальных доказательств.

**Замечание 1.** Принцип минимакса может быть использован для нахождения произвольного  $\lambda_j$ , если воспользоваться  $j$ -мерными подпространствами  $S_j$ :

$$\lambda_j = \min_{S_j} \left[ \max_{x \in S_j} R(x) \right]. \quad (28)$$

**Замечание 2.** Все перечисленные принципы можно обобщить на случай обобщенной задачи на собственные значения, если в отношении Релея заменить величину  $x^T x$  на  $x^T B x$ . Используя замену  $x = S z$ , где в качестве столбцов матрицы  $S$  выбраны  $B$ -ортонормированные собственные векторы матрицы  $B^{-1}A$ , получим

$$R(x) = \frac{x^T A x}{x^T B x} = \frac{\lambda_1 z_1^2 + \dots + \lambda_n z_n^2}{z_1^2 + \dots + z_n^2}. \quad (29)$$

Для этого и проводилась одновременная диагонализация матриц  $A$  и  $B$  на стр. 303, когда обе квадратичные формы представлялись в виде суммы полных квадратов. При этом легко видеть,

что  $\lambda_1 = \min R(x)$ , а  $\lambda_n = \max R(x)$ . Для примера осциллирующей системы с *неодинаковыми* массами вынужденное закрепление одной массы в состоянии равновесия будет увеличивать наименьшую частоту и уменьшать наибольшую.

**Упражнение 6.4.12.** На каком  $j$ -мерном подпространстве  $S_j$  из соотношения (28) минимум достигается, или, иначе говоря, для какого  $S_j$  максимум  $R(x)$  равен  $\lambda_j$ ?

**Упражнение 6.4.13.** Принцип минимакса (28) можно записать с помощью ограничений, а не подпространств:

$$\lambda_j = \min_{z_1, \dots, z_{n-j}} \left[ \max_{z^T z_1 = \dots = z^T z_{n-j} = 0} R(x) \right].$$

Какова связь между векторами  $z$  и подпространствами  $S_j$ ?

**Упражнение 6.4.14.** Показать, что наименьшее собственное значение  $\lambda_1$  задачи  $Ax = \lambda Bx$  не больше отношения  $a_{11}/b_{11}$  диагональных элементов матриц  $A$  и  $B$ .

**Упражнение 6.4.15** (трудное). Показать, что если последние строку и столбец симметрической матрицы  $A$  выбросить, то собственные значения  $\mu_i$  оставшейся подматрицы удовлетворяют неравенствам

$$\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \mu_2 \leq \dots \leq \mu_{n-1} \leq \lambda_n. \quad (30)$$

Напомним, что отбрасывание последних строки и столбца соответствует простому ограничению: последняя компонента  $x_n$  вектора  $x$  равна нулю, т. е.  $x$  ортогонален к вектору  $z = (0, \dots, 0, 1)$ . Неравенство  $\mu_j \geq \lambda_j$  выводится из принципа минимакса, а неравенство  $\mu_j \leq \lambda_{j+1}$  — из принципа максимина.

## § 6.5. ПРИНЦИП РЕЛЕЯ — РИТЦА И МЕТОД КОНЕЧНЫХ ЭЛЕМЕНТОВ

Настоящий параграф содержит два основных результата:

(i) Решение системы  $Ax = b$  эквивалентно минимизации квадратичной формы  $P(x) = \frac{1}{2} x^T Ax - x^T b$ .

(ii) Решение задачи  $Ax = \lambda_1 x$  эквивалентно минимизации отношения Релея  $R(x) = x^T Ax / x^T x$ .

Попытаемся объяснить, как эти факты могут быть применены.

История вопроса очень длинна, поскольку эти факты были известны уже более века назад. В некоторых классических задачах техники и физики, таких, как колебание пластин и стационарные состояния (собственные значения и функции) атома, эти минимизации использовались для построения грубых аппроксимаций точных решений. Эти аппроксимации *просто оказывались* грубыми, так как единственным инструментом исследователя были

карандаш, бумага и арифмометр. Конечно, были и математические принципы, но они не могли быть использованы.

Очевидно, что вычислительные машины принесли с собой революцию в численные методы, но в первую очередь она отбросила принципы минимума, так как они были слишком стары и слишком медленны. Вперед выдвинулся метод конечных разностей, поскольку он позволял легко «дискретизировать» дифференциальные уравнения. В § 1.6 мы уже видели, как каждая производная заменяется соответствующим разностным выражением. Физическая область покрывается решеткой или «сеткой», в каждом узле которой строится уравнение, такое, например, как  $-u_{j+1} + 2u_j - u_{j-1} = h^2 f_j$ . Математически задача сводится к системе  $Au = f$ , на разработку быстрых методов решения которой для случая больших размерностей и сильной разреженности матриц в основном были направлены усилия математиков-вычислителей в пятидесятые годы.

Но мы тогда не осознавали полностью, что для реальных задач техники, таких, как определение напряжений в узлах самолета или собственных частот человеческого мозга, метод конечных разностей невероятно усложняется. *Основные трудности здесь возникают не при решении (разностных) уравнений, а при построении.* Для нерегулярных областей нужны нерегулярные сетки, одновременно содержащие треугольные, четырехугольные или четырехгранные ячейки, а также единообразные способы аппроксимации соответствующих физических законов. Иначе говоря, вычислительные машины должны помогать не только при решении дискретных задач, но и при их формулировке.

Можете себе представить, что произошло. Вернулись старые методы, но с новой идеей и новым названием. Новое название — **метод конечных элементов**, а новая идея сделала возможным использовать в большей степени мощности вычислительных машин (при построении дискретной аппроксимации, при решении полученных дискретных задач и графическом изображении результатов), чем любые другие способы научных вычислений<sup>1)</sup>. Важно сохранить простоту математических идей, тогда можно усложнять приложения. Центр тяжести в этих приложениях переносится сейчас с задач конструирования самолетов на проблемы надежности ядерных реакторов, а в период написания книги шли ожесточенные дискуссии о применимости этого метода к динамике жидкостей. Для задач такого масштаба не обсуждается, однако, вопрос их стоимости, причем миллиард долларов, боюсь, пока является очень умеренной оценкой. Я надеюсь, что некоторым читателям будет достаточно интересно ознакомиться с методом

---

<sup>1)</sup> Пожалуйста, простите мне этот энтузиазм. Я знаю, что этот метод может и не оказаться бессмертным.

конечных элементов и попытаться извлечь из него серьезную пользу.

Чтобы разъяснить сам метод, мы начнем с классического принципа Релея—Ритца, который потом дополним новой идеей конечных элементов. Мы будем иметь дело со старой задачей, т. е. с уравнением  $-u'' = f(x)$  и старыми краевыми условиями  $u(0) = u(1) = 0$ , которая ранее исследовалась с помощью конечных разностей. Разумеется, это «бесконечномерная» задача, где вектор  $b$  заменен функцией  $f$ , а матрица  $A$  — оператором  $-d^2/dx^2$ . Поэтому мы можем провести аналогию с матричными задачами, выписав квадратичную форму, которую требуется минимизировать, в виде (скалярные произведения заменяются интегралами!)

$$P(v) = \frac{1}{2} v^T A v - v^T b = \frac{1}{2} \int_0^1 v(x) (-v''(x)) dx - \int_0^1 v(x) f(x) dx. \quad (31)$$

Это выражение нужно минимизировать на множестве всех функций  $v$ , удовлетворяющих краевым условиям, и та функция, которая доставляет минимум, является решением и исходной задачи. Дифференциальное уравнение сведено к задаче минимизации, и нам только остается проинтегрировать один из членов  $P(v)$  по частям:

$$\int_0^1 v(-v'') dx = \int_0^1 (v')^2 dx - [v v']_{x=0}^{x=1}.$$

Поскольку  $[v v']_{x=0}^{x=1} = 0$ , получаем

$$P(v) = \frac{1}{2} \int_0^1 (v')^2 dx - \int_0^1 v f dx.$$

Квадратичный член  $\int (v')^2$  симметрический (аналогично форме  $x^T A x$ ) и положительный; поэтому существование минимума гарантировано.

Как найти этот минимум? Точное его нахождение эквивалентно точному решению дифференциального уравнения, которое является бесконечномерной задачей. Принцип Релея—Ритца заменяет эту задачу  $n$ -мерной введением  $n$  пробных функций  $v = V_1, v = V_2, \dots, v = V_n$ . В классе всевозможных линейных комбинаций  $V = y_1 V_1(x) + \dots + y_n V_n(x)$  вычисляется такая частная комбинация  $U$ , которая минимизирует  $P(V)$ . Повторим сказанное: идея заключается в замене минимизации по всем возможным  $v$  минимизацией по подпространству и нахождении вместо  $u$  функции  $U$ , которая доставляет минимум формы  $P(v)$  в этом подпространстве. Мы надеемся, что эти функции окажутся «близкими».

После замены  $v$  на  $V$  квадратичная форма принимает вид

$$P(V) = \frac{1}{2} \int_0^1 (y_1 V_1' + \dots + y_n V_n')^2 - \int_0^1 (y_1 V_1 + \dots + y_n V_n) f.$$

Напомним, что функции  $V_j$  выбраны заранее, а неизвестными являются веса  $y_j$ . Если из этих весов мы составим вектор  $y$ , то  $P(V)$  точно совпадает с одной из квадратичных форм, которая нам уже встречалась:

$$P(V) = \frac{1}{2} y^T A y - y^T b, \quad (32)$$

где  $A_{ij} = \int V_i V_j'$  — коэффициент при члене  $y_i y_j$ , а  $b_j = \int V_j f$  — коэффициент при  $y_j$ . Разумеется, можно найти минимум формы  $\frac{1}{2} y^T A y - y^T b$ ; это эквивалентно решению системы  $Ay = b$ . Таким образом, этапами метода Релея—Ритца являются: (i) выбор пробных функций  $V_j$ ; (ii) вычисление коэффициентов  $A_{ij}$  и  $b_j$ ; (iii) решение системы  $Ay = b$  и (iv) построение приближенного решения  $U = y_1 V_1 + \dots + y_n V_n$ .

Все зависит от этапа (i). Если функции  $V_j$  не будут чрезвычайно просты, то другие этапы в сущности окажутся неосуществимыми. Кроме того, если никакая комбинация  $V_j$  не будет достаточно хорошо приближать истинное решение  $u$ , то последующие этапы будут бесполезны. Таким образом, мы должны одновременно учитывать требования точности и вычислительной реализуемости метода. *Ключевой идеей, которая сделала метод конечных элементов очень эффективным, явилось использование в качестве  $V_j$  кусочно-полиномиальных функций.*

Самыми простыми и наиболее широко используемыми являются кусочно-линейные функции. Сначала узлы сетки размещаются в точках  $x_1 = h$ ,  $x_2 = 2h$ , ...,  $x_n = nh$  (аналогично методу конечных разностей) и краевые условия требуют, чтобы в точках  $x_0 = 0$  и  $x_{n+1} = 1$  каждая функция  $V_j$  обращалась в нуль. В качестве функций  $V_j$  выберем функцию в виде «крыши», которая равна единице в узле  $x_j$  и нулю в остальных узлах (рис. 6.6a).

Эти функции «концентрируются» в малых интервалах около соответствующих узлов и равны нулю на остальной части отрезка. Любая комбинация  $y_1 V_1 + \dots + y_n V_n$  имеет значение  $y_j$  в узле с таким же номером, поскольку другие функции  $V_k$  в этом узле равны нулю, в силу чего ее легко изобразить графически (рис. 6.6b). Это завершает этап (i).

На следующем этапе вычисляются коэффициенты  $A_{ij} = \int V_i V_j'$ , являющиеся элементами «матрицы жесткости»  $A$ . Наклон  $V_j'$  функции  $V_j$  равен  $1/h$  в малом интервале слева от  $x_j$  и  $-1/h$  справа,

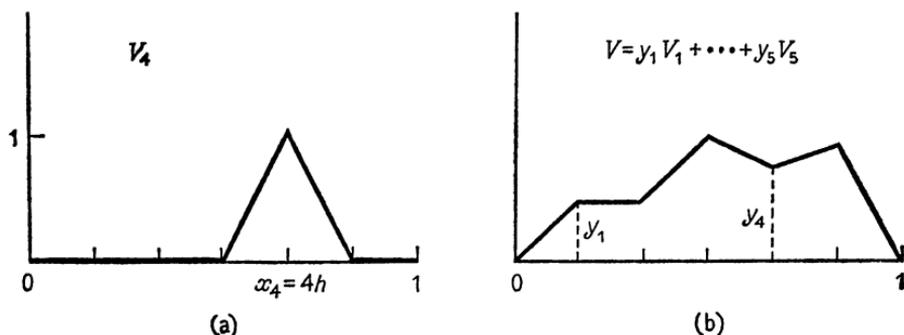


Рис. 6.6. Функции в виде «крыши» и их линейные комбинации.

и если эти «сдвоенные интервалы» для функций  $V_i$  и  $V_j$  не перекрываются, то произведение  $V_i V_j$  тождественно равно нулю. Но они перекрываются, только когда либо  $i = j$ , и тогда

$$\int V_i V_j = \int \left(\frac{1}{h}\right)^2 + \int \left(-\frac{1}{h}\right)^2 = \frac{2}{h},$$

либо когда  $i = j \pm 1$ , и тогда

$$\int V_i V_j = \int \left(\frac{1}{h}\right) \left(-\frac{1}{h}\right) = -\frac{1}{h}.$$

Следовательно, матрица жесткости оказывается трехдиагональной:

$$A = \frac{1}{h} \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix},$$

и в точности совпадает с матрицей  $A$  метода конечных разностей, что приводит к бесконечным дискуссиям о связях между этими двумя методами. Более сложные конечные элементы — полиномы более высоких степеней. Они определяются на интервалах для обыкновенных дифференциальных уравнений или на треугольниках и четырехугольниках для дифференциальных уравнений с частными производными и также приводят к разреженным матрицам  $A$ . Вы можете представлять себе метод конечных элементов как единообразный способ построения точных разностных уравнений на нерегулярных сетках, так что он лежит на «пересечении» методов Релея — Ритца и метода конечных разностей. Основным достоинством этого метода является простота кусочно-полиномиальных функций, поскольку на каждом подынтервале их наклоны (производные) легко находятся и интегрируются.

Компоненты  $b_j$  вектора правой части оказываются совершенно новыми. Вместо значения  $f$  в узле  $x_j$ , как в методе конечных разностей, берется специальное осреднение  $f$  в окрестности этого узла:  $b_j = \int V_j f$ . На этапе (iii) мы решаем систему  $Ay = b$ , решение которой дает коэффициенты минимизирующей пробной функции  $U = y_1 V_1 + \dots + y_n V_n$ . Наконец, соединяя вершины  $y_j$  ломаной линией, мы получаем изображение приближенного решения  $U$ .

Отметим одно специальное свойство рассматриваемой частной задачи:  $U$  не просто близка к точному решению, но совпадает с ним в узлах сетки. Иначе говоря,  $U$  является линейной интерполяцией  $u$ . Нельзя ожидать, чтобы точное решение имело такой вид для достаточно сложных уравнений. Но ошибка  $U - u$  оказывается очень мала даже на грубой сетке. Соответствующая теория сходимости освещается в книге автора и Г. Фикса (М.: Мир, 1977). В ней также обсуждаются дифференциальные уравнения с частными производными, для которых рассматриваемый метод сохраняет свои достоинства.

**Упражнение 6.5.1.** Для  $f \equiv 1$  решение уравнения  $-u'' = f$  является параболой  $u = (x - x^2)/2$ . Вычислить коэффициенты  $b_j = \int V_j f$  и показать, что точные значения  $y_j = u(x_j)$  удовлетворяют системе метода конечных элементов  $Ay = b$ .

**Упражнение 6.5.2.** В случае  $A = I$  квадратичная форма имеет вид  $P(y) = \frac{1}{2} y^T y - y^T b$  и достигает минимума на векторе  $y = b$ . Доказать, что  $P(y) - P(b) = \frac{1}{2} \|y - b\|^2$ , и используя это равенство, объяснить, почему вектор, минимизирующий  $P(y)$  на подпространстве пробных функций, является ближайшим к  $b$  Принцип Релея—Ритца автоматически задает проекцию истинного решения на подпространство пробных функций.

## ЗАДАЧИ НА СОБСТВЕННЫЕ ЗНАЧЕНИЯ

Идея Релея—Ритца осуществлять минимизацию на конечномерном семействе функций  $V$  вместо минимизации на всем пространстве так же полезна для задач на собственные значения, как и для стационарных уравнений. Теперь минимизируемым будет отношение Релея. Его точным минимумом является основная частота  $\lambda_1$ , и его приближенный минимум  $\Lambda_1$  увеличится при переходе от всего пространства функций  $v$  к более узкому классу пробных функций  $V_1$ . И снова построить дискретную задачу (применить практически сам принцип) можно, лишь если функции  $V_j$  легко вычислимы. Таким образом, шаг, который был проделан за последние двадцать лет, был совершенно естествен и неминуем. Заключается он в применении идей нового метода

конечных элементов к вариационной формулировке задачи на собственные значения.

Наиболее распространенным и в то же время наиболее простым примером является задача

$$-u'' = \lambda u, \quad u(0) = u(1) = 0. \quad (33)$$

Ее первое собственное значение  $\lambda_1 = \pi^2$ , и ему соответствует собственная функция  $u = \sin \pi x$ . Эти значения и функция удовлетворяют уравнениям (33) и дают минимум отношения Релея

$$R(v) = \frac{v^T [-d^2/dx^2] v}{v^T v} = \frac{\int_0^1 v (-v'')}{\int_0^1 v^2} = \frac{\int_0^1 (v')^2}{\int_0^1 v^2}.$$

Физически это отношение потенциальной и кинетической энергий, которые сбалансированы на собственной функции. Эта собственная функция обычно неизвестна, и, чтобы вычислить ее приближенно, выберем в качестве допустимых только функции  $V = y_1 V_1 + \dots + y_n V_n$ . В результате имеем

$$R(V) = \frac{\int_0^1 (y_1 V_1' + \dots + y_n V_n')^2}{\int_0^1 (y_1 V_1 + \dots + y_n V_n)^2} = \frac{y^T A y}{y^T B y}.$$

Теперь перед нами встала задача минимизации отношения  $y^T A y / y^T B y$ . Если матрица  $B$  единичная, то это приводит к стандартной задаче на собственные значения  $Ay = \lambda y$ . Но наша матрица  $B$  будет трехдиагональной, и, следовательно, мы приходим к *обобщенной задаче на собственные значения* (см. стр. 314). Минимальное значение  $\Lambda_1$  отношения  $R(V)$  будет наименьшим собственным значением задачи  $Ay = \lambda By$ , а соответствующий собственный вектор  $y$  дает возможность построить приближение  $U = y_1 V_1 + \dots + y_n V_n$  к первой собственной функции исходной задачи.

Как и для стационарных задач, наш метод содержит четыре этапа: (i) выбираются функции  $V_j$ ; (ii) вычисляются матрицы  $A$  и  $B$ ; (iii) решается задача  $Ay = \lambda By$  и (iv) выбирается  $\Lambda_1$  и строится соответствующая функция  $U$ . Я не знаю, почему это стоит миллиард долларов.

**Упражнение 6.5.3.** Для кусочно-линейных функций  $V_1$  и  $V_2$ , соответствующих узлам  $x = h = 1/3$  и  $x = 2h = 2/3$ , вычислить матрицу  $B$  порядка два. Составить построенную дискретную задачу на собственные значения с задачей из упр. 6.3.6 и сравнить  $\Lambda_1$  с фактическим собственным значением  $\lambda_1 = \pi^2$ .

**Упражнение 6.5.4.** Объяснить, почему из принципа минимакса следует неравенство  $\Lambda_2 \geq \lambda_2$ .

## ВЫЧИСЛЕНИЯ С МАТРИЦАМИ

## § 7.1. ВВЕДЕНИЕ

Целью этой книги является рассмотрение некоторых прикладных разделов теории матриц. Излагаемая здесь теория принципиально ничем не отличается от стандартных курсов абстрактной линейной алгебры, но сейчас для нас самое главное, что рассматриваемая теория действительно важна для приложений. Разница состоит в том, что *переместился центр тяжести* в соответствии с новой точкой зрения. Метод исключения Гаусса — это теперь не только способ нахождения базиса пространства строк, а процесс Грама — Шмидта — не просто способ доказательства, что каждое подпространство имеет ортонормированный базис. Напротив, нам действительно *нужны* эти алгоритмы и нужно удобное описание ( $A = LU$  или  $A = QR$ ) того, что они делают.

Настоящая глава осуществляет еще несколько шагов в этом направлении. Я полагаю, что они диктуются скорее вычислительной необходимостью, чем соображениями элегантности, но не знаю, следует ли мне извиняться за это. Они могут показаться очень поверхностными, но это неверно. Речь будет идти о старейших и важнейших задачах  $Ax = b$  и  $Ax = \lambda x$ , но каждая из них была по-настоящему понята только современным поколением математиков. В численном анализе существует своего рода естественный отбор идей и методов, и мы хотим описать некоторые из тех, которые выжили. Они делятся на три группы:

1. *Методы решения системы  $Ax = b$ .* Метод исключения Гаусса является идеальным алгоритмом, кроме, быть может, частных задач, имеющих особые свойства — такие свойства имеет почти каждая задача. Наше внимание будет сконцентрировано на свойстве *разреженности*, когда большинство элементов матрицы  $A$  равно нулю, и на развитии *итерационных, а не прямых, методов решения системы  $Ax = b$ .* Итерационный метод является «самокорректирующимся», и этот процесс коррекции повторяется многократно. При применении итерационного метода никогда не

достигается точный ответ, но цель состоит в другом — получить приближенное решение быстрее, чем при исключении. В некоторых случаях это можно сделать. Во многих других случаях исключение оказывается лучше и быстрее, если использовать разреженность матрицы. Эта «конкуренция» далека от разрешения, и наша первая цель — установить условия, которые при итерационном методе гарантируют сходимость к истинному решению  $A^{-1}b$  и определяют ее скорость. В дальнейшем, в § 7.4, мы применим эти условия к методу верхней релаксации и другим итерационным методам.

2. *Методы решения задачи  $Ax = \lambda x$ .* Задача на собственные значения является одним из примеров замечательных успехов численного анализа. Она четко поставлена, ее значение очевидно, но до недавнего времени никто не знал, как ее решать. Были предложены десятки алгоритмов, но их, конечно, нельзя расположить в строгом порядке, от наилучшего до самого наихудшего. Все зависит от размеров и свойств матрицы  $A$  и от числа собственных значений, которые мы хотим вычислить. Другими словами, опасно даже запрашивать в вычислительном центре подпрограммы вычисления собственных значений, не зная того, какой метод в них используется. (Конечно, я надеюсь, вам не надо проверять каждое фортрановское предложение в этих программах.) Прекрасная книга Уилкинсона и Райнша<sup>1)</sup> разъясняет целый ряд алгоритмов, и мы выбрали из нее две или три идеи, которые вытесняют почти все предшествующие:  $QR$ -алгоритм, класс «степенных методов» и преобразования, приводящие симметрическую матрицу к трехдиагональному виду.

Первые два метода итерационные, а последний — прямой; он реализуется за конечное число действий, но не вычисляет сам собственные значения, а только дает более простую матрицу для применения итераций.

3. *Число обусловленности матрицы.* В § 7.2 делается попытка оценить чувствительность, или «уязвимость» решения: если изменить немного  $A$  и  $b$ , то как сильно изменится решение  $x = A^{-1}b$ ? Прежде чем заняться этим вопросом, мы хотим отметить одно препятствие (которое легко обойти). Должен существовать способ измерения возмущений  $\delta A$  матрицы  $A$  и оценки размера матрицы  $A$ . Длина вектора уже определена, и теперь нам необходима *норма матрицы*. Тогда число обусловленности, а с ним и чувствительность матрицы, будут выражены через нормы матриц  $A$  и  $A^{-1}$ .

---

<sup>1)</sup> J. M. Wilkinson, C. Reinsch, Handbook for Automatic Computation, Linear Algebra, Springer, New York and Berlin, 1971.

## § 7.2. НОРМА И ЧИСЛО ОБУСЛОВЛЕННОСТИ МАТРИЦЫ

Ошибки бывают разные, и следует различать небольшие ошибки (погрешности), возможно, неизбежные даже для идеального математика и идеальной вычислительной машины, и грубые ошибки, которые опаснее и по величине больше по крайней мере на порядок. Когда вычислительная машина округляет число до девятой значащей цифры, то это небольшая ошибка. Если же задача настолько чувствительна, что ошибка округления полностью изменяет решение, то почти наверняка где-то допущена грубая ошибка. Наша цель в этом разделе — проанализировать влияние ошибок так, чтобы грубых ошибок можно было избежать.

На самом деле мы продолжаем обсуждение, начатое в гл. 1, матриц

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix} \quad \text{и} \quad A' = \begin{bmatrix} 0.0001 & 1 \\ 1 & 1 \end{bmatrix}.$$

Мы утверждаем, что  $A'$  хорошо обусловлена и не особенно чувствительна к ошибкам округления — за исключением того случая, когда исключение Гаусса проводится неразумным образом и матрица становится крайне «уязвимой». Грубой ошибкой будет принять 0.0001 за первый ведущий элемент, поэтому мы должны добиться большего и более «надежного» ведущего элемента посредством перестановки строк матрицы  $A'$ . Когда «частичный выбор ведущего элемента» введен в алгоритм исключения, так что компьютер автоматически ищет наибольший ведущий элемент, то естественной сопротивляемости ошибкам округления уже ничто не угрожает.

Как мы можем измерить эту естественную сопротивляемость и определить, является матрица хорошо обусловленной или плохо обусловленной? Если имеется малое возмущение в  $b$  или в  $A$ , как велики изменения, возникающие в решении  $x$ ?

Мы начнем с возмущения в правой части, от  $b$  до  $b + \delta b$ . Эта ошибка может возникнуть из экспериментальных данных и из-за округления; мы можем предположить, что  $\delta b$  мало, но его направления мы не знаем. Соответствующее решение изменится с  $x$  на  $x + \delta x$ , т. е., вычитая  $Ax = b$  из  $A(x + \delta x) = b + \delta b$ , получаем  $A(\delta x) = \delta b$ . Это особенно простой случай; мы рассмотрим все возмущения  $\delta b$  и оцениваем получаемое возмущение  $\delta x = A^{-1}\delta b$ . Большие изменения в решении будут тогда, когда  $A^{-1}$  велика ( $A$  почти вырожденная), и особенно велики, когда вектор  $\delta b$  имеет то направление, которое больше всего увеличивается матрицей  $A^{-1}$ .

Начнем с предположения, что матрица  $A$  симметрическая и что ее собственные значения положительны:  $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ .

Любой вектор  $\delta b$  является комбинацией соответствующих единичных собственных векторов  $x_1, \dots, x_n$ , и худшим возмущением будет ошибка в направлении первого собственного вектора  $x_1$ :

$$\text{если } \delta b = \varepsilon x_1, \text{ то } \delta x = \frac{\delta b}{\lambda_1}. \quad (1)$$

Погрешность правой части  $\|\delta b\|$  увеличивается на множитель  $1/\lambda_1$ , который является наибольшим собственным значением матрицы  $A^{-1}$ . Увеличение наибольшее, если  $\lambda_1$  близко к нулю, так что матрицы, близкие к вырожденным, наиболее чувствительны.

У этой меры чувствительности есть лишь один изъян, но очень серьезный. Предположим, что мы умножим все элементы матрицы  $A$  на 1000; тогда  $\lambda_1$  умножится на 1000 и кажется, что матрица стала менее вырожденной. Это оскорбляет наше чувство справедливости — такое простое масштабирование не может сделать плохо обусловленную матрицу хорошей.

Верно, что  $\delta x$  станет меньше в 1000 раз, но то же самое произойдет с решением  $x = A^{-1}b$  и относительная погрешность  $\|\delta x\|/\|x\|$  останется той же. Множитель  $\|x\|$  в знаменателе нормализует задачу по отношению к тривиальному изменению масштаба. В то же время имеется соответствующая нормализация для  $\delta b$ , и наша задача — сравнить относительное изменение  $\|\delta b\|/\|b\|$  с относительной погрешностью  $\|\delta x\|/\|x\|$ .

Наихудший случай — это когда числитель  $\|\delta x\|$  велик, т. е. возмущение ориентировано в направлении собственного вектора  $x_1$ , а знаменатель  $\|x\|$  мал (невозмущенное решение  $x$  оказывается очень мало в сравнении с невозмущенным  $b$ ). Это означает, что решение исходной задачи  $Ax = b$  лежит в другом экстремальном направлении — в направлении последнего собственного вектора  $x_n$ :

$$\text{если } b = x_n, \text{ то } x = A^{-1}b = \frac{b}{\lambda_n}. \quad (2)$$

Приведенная комбинация ( $b = x_n$  и  $\delta b = \varepsilon x_1$ ) делает относительную погрешность наибольшей. Экстремальные случаи описываются следующими неравенствами:

**7А.** Для положительно определенной матрицы решение  $x = A^{-1}b$  и погрешность  $\delta x = A^{-1}\delta b$  всегда удовлетворяют неравенствам

$$\|x\| \geq \frac{\|b\|}{\lambda_n}, \quad \|\delta x\| \leq \frac{\|\delta b\|}{\lambda_1}. \quad (3)$$

Поэтому для относительной погрешности справедлива оценка

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\lambda_n}{\lambda_1} \frac{\|\delta b\|}{\|b\|}. \quad (4)$$

Величина  $c = \frac{\lambda_n}{\lambda_1} = \frac{\lambda_{\max}}{\lambda_{\min}}$  называется **числом обусловленности матрицы  $A$** .

**Пример 1.** Собственные значения матрицы

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix}$$

приблизительно равны  $\lambda_1 = 10^{-4}/2$  и  $\lambda_2 = 2$ . Поэтому число обусловленности матрицы  $A$  примерно равно  $c = 4 \cdot 10^4$ , и мы должны ожидать сильного изменения решения при некоторых очень обычных изменениях начальных данных. На стр. 48 мы сравнили уравнения  $Ax = b$  и  $Ax' = b'$ :

$$\begin{array}{l} u + v = 2, \\ u + 1.0001v = 2.0001 \end{array} \quad \text{и} \quad \begin{array}{l} u' + v' = 2, \\ u' + 1.0001v' = 2.0002. \end{array}$$

Решение изменяется с  $x = (1, 1)$  до  $x' = (0, 2)$ :

$$\frac{\|x' - x\|}{\|x\|} = \frac{\|(-1, 1)\|}{\|(1, 1)\|} = 1, \quad \frac{\|b' - b\|}{\|b\|} = \frac{\|(0, 0.0001)\|}{\|(2, 2.0001)\|} \sim \frac{10^{-4}}{2\sqrt{2}}.$$

Относительное увеличение равно  $\|\delta x\|/\|x\| \sim 2\sqrt{2} \cdot 10^4 \|\delta b\|/\|b\|$ . Мы здесь не выбрали возмущение специальным образом (наше  $\delta b$  составляет угол  $45^\circ$  с собственным вектором  $x_1$ , что и объясняет разницу в  $\sqrt{2}$  между нашим множителем и наилучшим возможным увеличением  $c = 4 \cdot 10^4$ ) и тем не менее обнаруживаем громадные изменения в решении.

Отметим, что размер матрицы на число обусловленности прямого влияния не оказывает; если  $A = I$  или даже  $A = I/10$ , то  $c = \lambda_{\max}/\lambda_{\min} = 1$ . Для сравнения, *определитель — очень плохой показатель плохой обусловленности*, потому что зависит не только от масштабирования, но и от порядка  $n$ ; если  $A = I/10$ , то ее определитель равен  $10^{-n}$ . На самом же деле эта «почти вырожденная матрица» так хорошо обусловлена, как это только возможно.

**Пример 2.** Рассмотрим конечно-разностную матрицу

$$A = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & & \\ & & & \ddots & \\ & & & & -1 & 2 \end{bmatrix}$$

порядка  $n$ .

Приблизительно ее наибольшее собственное число  $\lambda_n = 4$ , а наименьшее  $\lambda_1 = \pi^2/n^2$ . Поэтому число обусловленности приблизительно равно  $c = n^2/2$ , в данном случае имеется фактическая зависимость от  $n$ . Мы лучше аппроксимируем задачу —  $u'' = f$ , увеличивая число неизвестных, но труднее становится вычислить

приближенное решение. Это делается не только дольше, но и наблюдается большее воздействие ошибок округления. С некоторого момента увеличение  $n$  фактически приводит к ухудшению ответа.

К счастью для инженеров, этот момент наступает тогда, когда точность уже достаточно хороша. Работая с обыкновенной точностью, обычная вычислительная машина может делать ошибки округления порядка  $10^{-9}$ . Если приближенное решение ищется для значения  $n=100$ , т. е.  $c=5000$ , то такая погрешность возрастает самое большее до порядка  $10^{-5}$ , т. е. точность получается выше, чем при любом обычном измерении. Но при 10 000 неизвестных или при рассмотрении конечно-разностной аппроксимации уравнений более высокого порядка, таких, как  $d^4u/dx^4 = f(x)$ , для которых число обусловленности растет как  $n^4$ , есть причины для беспокойства<sup>1)</sup>.

Наш анализ до сих пор применялся исключительно к симметрическим матрицам с положительными собственными значениями. Мы легко могли бы отбросить предположение о положительности и использовать абсолютные значения  $\lambda_i$ , числом обусловленности тогда стала бы величина  $c = \max |\lambda_i| / \min |\lambda_i|$ . Но для того чтобы обойтись без условия симметричности, что нам бы хотелось сделать, нужны более существенные изменения. Легко видеть, что для матриц

$$A = \begin{bmatrix} 1 & 100 \\ 0 & 1 \end{bmatrix} \quad \text{и} \quad A^{-1} = \begin{bmatrix} 1 & -100 \\ 0 & 1 \end{bmatrix} \quad (5)$$

все собственные значения равны единице, но совершенно неправильно, что относительное изменение  $x$  ограничено относительным изменением  $b$  — число обусловленности не равно  $\lambda_{\max}/\lambda_{\min} = 1$ . Сравним решения

$$x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \text{когда} \quad b = \begin{bmatrix} 100 \\ 1 \end{bmatrix}; \quad x' = \begin{bmatrix} 100 \\ 0 \end{bmatrix}, \quad \text{когда} \quad b' = \begin{bmatrix} 100 \\ 0 \end{bmatrix}.$$

Изменение на 1% в  $b$  вызывает изменения порядка сотен в  $x$ ; множитель увеличения равен  $100^2$ . Так как  $c$  представляет собой верхнюю границу этого увеличения, эта константа должна быть не менее 10 000. Плохим свойством таких матриц является то, что большие внедиагональные элементы в  $A$  означают существование таких же больших элементов в  $A^{-1}$  — это противоречит нашим интуитивным ощущениям, что  $A^{-1}$  должна была бы становиться меньше, если  $A$  становится больше.

<sup>1)</sup> Эмпирический закон, экспериментально проверенный, состоит в том, что при исключении Гаусса при наличии ошибок округления в решении может потеряться  $\log c$  десятичных знаков.

Чтобы дать подходящее определение числа обусловленности, мы снова обратимся к уравнению (3). Из него следует, что  $\lambda_n$  — это наибольшее из возможных значений отношения  $\|b\|$  к  $\|x\|$ ; мы пытались сделать  $x$  маленьким, а  $b = Ax$  большим. (Экстремальный случай имеет место на собственном векторе  $x_n$ , когда отношение  $\|Ax\|$  к  $\|x\|$  точно равно  $\lambda_n$ ). Единственное отличие в случае, когда матрица  $A$  уже не симметрическая, состоит в том, что максимум  $\|Ax\|/\|x\|$  может находиться на векторе  $x$ , который уже не является ни одним из собственных векторов. Этот максимум все еще является прекрасной мерой «величины» матрицы  $A$ . Он называется **нормой** матрицы  $A$  и обозначается через  $\|A\|$ .

**7B.** *Нормой матрицы  $A$  называется число, определяемое соотношением*

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}. \quad (6)$$

*Другими словами,  $\|A\|$  ограничивает «степень возрастания» при умножении на матрицу  $A$ :*

$$\|Ax\| \leq \|A\| \|x\| \quad \text{для всех векторов } x, \quad (7)$$

*причем равенство достигается по крайней мере на одном ненулевом векторе  $x$ .*

Матрицы  $A$  и  $A^{-1}$  из (5) имеют нормы, лежащие между 100 и 101. Через некоторое время мы сможем посчитать их точно, но сначала мы хотим установить связь между нормами и числами обусловленности. Так как  $b = Ax$  и  $\delta x = A^{-1}\delta b$ , непосредственно из определения (7) мы получаем, что

$$\|b\| \leq \|A\| \|x\| \quad \text{и} \quad \|\delta x\| \leq \|A^{-1}\| \|\delta b\|. \quad (8)$$

Это очевидная замена (3), когда  $A$  — несимметрическая матрица. В симметрическом случае  $\|A\|$  есть не что иное, как  $\lambda_n$ , а  $\|A^{-1}\|$  — не что иное, как  $1/\lambda_1$ . Поэтому очевидной заменой для  $\lambda_n/\lambda_1$  является произведение  $\|A\| \|A^{-1}\|$ .

**7C.** *Числом обусловленности матрицы  $A$  называется величина  $c = \|A\| \|A^{-1}\|$ , и относительная погрешность удовлетворяет неравенству*

$$\frac{\|\delta x\|}{\|x\|} \leq c \frac{\|\delta b\|}{\|b\|}. \quad (9)$$

*Если мы возмутим матрицу  $A$ , а не правую часть  $b$ , то*

$$\frac{\|\delta x\|}{\|x + \delta x\|} \leq c \frac{\|\delta A\|}{\|A\|}. \quad (10)$$

Неравенство (9) верно для любых  $b$  и  $\delta b$ , и оно является результатом двух неравенств (8). Замечательно то, что то же самое

число обусловленности появилось в (10), когда матрица сама возмущена. Если  $Ax = b$  и  $(A + \delta A)(x + \delta x) = b$ , то, вычитая одно из другого, получим

$$A\delta x + \delta A(x + \delta x) = 0, \quad \text{или} \quad \delta x = -A^{-1}(\delta A)(x + \delta x).$$

Умножение на  $\delta A$  увеличивает норму вектора не больше чем в  $\|\delta A\|$  раз, но тогда умножение на  $A^{-1}$  увеличивает ее не более чем в  $\|A^{-1}\|$  раз. Поэтому

$$\|\delta x\| \leq \|A^{-1}\| \|\delta A\| \|x + \delta x\|,$$

или

$$\frac{\|\delta x\|}{\|x + \delta x\|} \leq \|A^{-1}\| \|\delta A\| = c \frac{\|\delta A\|}{\|A\|}.$$

Эти неравенства означают, что ошибка округления возникает из двух источников: первая — естественная чувствительность задачи, которая измеряется числом обусловленности  $c$ ; и вторая — фактические погрешности  $\delta b$  или  $\delta A$ , которые вносят вклад в ее решение. Приведенные рассуждения послужили Уилкинсону базисом для анализа погрешностей. Так как алгоритм исключения фактически дает приближенные множители  $L'$  и  $U'$ , то он вместо задачи с исходной матрицей  $A = LU$  решает уравнение с ошибочной матрицей  $A + \delta A = L'U'$ . Уилкинсон доказал, что частичный выбор ведущего элемента обеспечивает контроль над  $\delta A$  (см. его книгу «Rounding Errors in Algebraic Processes»). *Итак, полная ошибка из-за ошибок округления определяется числом обусловленности  $c$ .*

**Упражнение 7.2.1.** Пусть  $A$  является ортогональной матрицей  $Q$ . Показать, что  $\|Q\| = 1$  и  $c(Q) = 1$ . Ортогональные матрицы (и матрицы вида  $\alpha Q$ ) прекрасно обусловлены.

**Упражнение 7.2.2.** Из какого «знаменитого» неравенства вытекает неравенство  $\|(A+B)x\| \leq \|Ax\| + \|Bx\|$  и почему из (6) следует, что  $\|A+B\| \leq \|A\| + \|B\|$ ?

**Упражнение 7.2.3.** Объяснить, почему  $\|ABx\| \leq \|A\| \|B\| \|x\|$ , и вывести из (6), что  $\|AB\| \leq \|A\| \|B\|$ . Показать, что отсюда также следует неравенство  $c(AB) \leq c(A) \cdot c(B)$ .

**Упражнение 7.2.4.** Для положительно определенной матрицы  $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$  вычислить  $\|A^{-1}\| = 1/\lambda_1$ ,  $\|A\| = \lambda_2$  и  $c(A) = \lambda_2/\lambda_1$ . Подобрать правую часть  $b$  и возмущения  $\delta b$ , такие, чтобы погрешность решения была наилучшей, т. е.  $\|\delta x\| / \|x\| = c \|\delta b\| / \|b\|$ .

**Упражнение 7.2.5.** Показать, что если  $\lambda$  — произвольное собственное значение матрицы  $A$ ,  $Ax = \lambda x$ , то  $|\lambda| \leq \|A\|$ .

## ФОРМУЛА ДЛЯ НОРМЫ

Норма матрицы  $A$  измеряет величину наибольшего изменения векторов (собственных или нет) при умножении их на эту матрицу:  $\|A\| = \max(\|Ax\|/\|x\|)$ . Норма единичной матрицы равна 1. Для того чтобы вычислить этот «увеличивающий множитель» в общем случае, возведем в квадрат обе части определяющего норму соотношения:

$$\|A\|^2 = \max \frac{\|Ax\|^2}{\|x\|^2} = \max \frac{x^T A^T A x}{x^T x}. \quad (11)$$

Полученные соотношения возвращают нас к симметрической матрице  $A^T A$  и ее «отношению Релея».

**7D.** *Норма матрицы  $A$  равна квадратному корню из наибольшего собственного значения матрицы  $A^T A$ :  $\|A\|^2 = \lambda_{\max}(A^T A)$ . В случае когда  $A$  симметрическая,  $A^T A = A^2$  и, следовательно, норма равна наибольшему по модулю собственному значению матрицы  $A$ :  $\|A\| = \max |\lambda_i|$ . В любом случае вектор, который увеличивается максимальным образом, является соответствующим собственным вектором матрицы  $A^T A$ :*

$$\frac{x^T A^T A x}{x^T x} = \frac{x^T (\lambda_{\max} x)}{x^T x} = \lambda_{\max} = \|A\|^2.$$

*Замечание 1.* Норма и число обусловленности матрицы  $A$  в практических задачах фактически не вычисляются, а только оцениваются. У нас нет возможности решать задачу на собственные значения для вычисления  $\lambda_{\max}(A^T A)$ .

*Замечание 2.* Число обусловленности объясняет, почему нормальные уравнения  $A^T A x = A^T b$  так трудно решать методом наименьших квадратов: число обусловленности  $c(A^T A)$  является квадратом  $c(A)$ . Формируя  $A^T A$ , можно превратить хорошую задачу в плохую, и гораздо лучше (исключая очень маленькие задачи) использовать сразу либо процесс Грама—Шмидта, либо сингулярное разложение  $A = Q_1 \Sigma Q_2^T$ .

*Замечание 3.* Элементы  $\sigma_i$  диагональной матрицы  $\Sigma$  называются *сингулярными числами* матрицы  $A$ , и по построению (см. стр. 171) их квадраты являются собственными значениями матрицы  $A^T A$ . Поэтому другой формулой для нормы будет  $\|A\| = \sigma_{\max}$ . В соотношении  $\|Ax\| = \|Q_1 \Sigma Q_2^T x\|$  ортогональные матрицы  $Q_1$  и  $Q_2$  не изменяют длины векторов и, следовательно, наибольший увеличивающий множитель равен наибольшему  $\sigma$ .

*Замечание 4.* Погрешности округления возникают не только при решении задачи  $Ax = b$ , но и задачи  $Ax = \lambda x$ . Поэтому возни-

кает новый вопрос: что такое «число обусловленности для задач на собственные значения»? Очевидный ответ неверен — это не число обусловленности самой матрицы  $A$ . Чувствительность собственных значений измеряется числом обусловленности диагонализующей матрицы  $S$ . Если  $\mu$  — собственное значение матрицы  $A + E$ , то расстояние от него до какого-нибудь собственного значения матрицы  $A$  равно

$$|\mu - \lambda| \leq \|S\| \|S^{-1}\| \|E\| = c(S) \|E\|. \quad (12)$$

В случае когда  $S$  — ортогональная матрица  $Q$ , задача на собственные значения прекрасно обусловлена ( $c(Q) = 1$ ) и изменение  $\mu - \lambda$  собственных значений не больше, чем возмущение  $E$  матрицы  $A$ . Это происходит тогда, когда собственные векторы  $A$  образуют ортонормированную систему — они являются столбцами матрицы  $S$ . Поэтому наилучшим является случай, когда  $A$  симметрическая или, более общо, когда  $A \cdot A^T = A^T \cdot A$ , т. е. когда  $A$  — нормальная матрица и диагонализующая ее матрица  $S$  является ортогональной матрицей  $Q$  (см. разд. 5.6). Собственные числа нормальной матрицы прекрасно обусловлены. Нетрудно выяснить, как матрица  $S$  присутствует в формуле для возмущений для отдельного собственного значения: если  $x_k$  является  $k$ -м столбцом  $S$ , а  $y_k$  является  $k$ -й строкой  $S^{-1}$ , то

$$\mu_k - \lambda_k = y_k E x_k + \text{члены порядка } \|E\|^2. \quad (13)$$

В практике  $y_k E x_k$  достаточно правильно оценивает изменение собственных значений, и в каждом хорошем алгоритме стремятся оставлять матрицу погрешностей  $E$  столь малой, насколько это возможно, обычно требуя (как в  $QR$ -алгоритме, который будет описан в следующем разделе) ортогональности матрицы на каждом шаге итераций.

**Упражнение 7.2.6.** Найти точные нормы матриц (5).

**Упражнение 7.2.7.** Для матрицы  $A$  из упр. 5.6.3 показать, сравнивая собственные значения матриц  $A^T A$  и  $A A^T$ , что  $\|A\| = \|A^T\|$ .

**Упражнение 7.2.8.** Для положительно определенной матрицы  $A$  разложение Холецкого имеет вид  $A = LDL^T = W^T W$ , где  $W = \sqrt{D} L^T$ . Показать, непосредственно используя 7D, что число обусловленности матрицы  $W$  равно корню квадратному из числа обусловленности матрицы  $A$ . Отсюда следует, что алгоритм Гаусса для положительно определенных матриц не требует перестановки строк. Обусловленность не ухудшается, так как  $c(A) = c(W^T) c(W)$ .

**Упражнение 7.2.9.** Построив с помощью матриц порядка 2 контрпримеры к неравенствам  $\lambda_{\max}(A + B) \leq \lambda_{\max}(A) + \lambda_{\max}(B)$ ,  $\lambda_{\max}(AB) \leq \lambda_{\max}(A) \lambda_{\max}(B)$ , показать, что наибольшее собственное значение, вообще говоря, не годится в качестве нормы.

**Упражнение 7.2.10.** Предположим, что под обозначением  $\|x\|$  понимается не евклидова норма  $\sqrt{x_1^2 + \dots + x_n^2}$ , а «максимальная норма», или « $L_\infty$ -норма»,  $\|x\|_\infty = \max |x_i|$  (пример:  $\|(1, -2, 1)\| = 2$ ). Вычислить соответствующую матричную норму

$$\|A\|_\infty = \max_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty} = \max_{x_i = \pm 1} \|Ax\|_\infty, \quad \text{если} \quad A = \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}.$$

### § 7.3. ВЫЧИСЛЕНИЕ СОБСТВЕННЫХ ЗНАЧЕНИЙ

Невозможно указать наилучший способ нахождения собственных значений матрицы. Однако существует несколько ужасных методов, которые никогда не следует применять, и несколько идей, которые следует использовать постоянно. Мы начнем обсуждение с одного очень грубого, но простого подхода, называемого *степенным методом*, свойства сходимости которого очень легко выявить. Затем мы изучим более сложный алгоритм, который начинается с приведения симметрической матрицы к трехдиагональной и заканчивается практически приведением к диагональному виду. Этот последний шаг реализуется при помощи процесса Грама—Шмидта, а весь метод известен как *QR-алгоритм*.

Обычный степенной метод реализуется точно по принципу разностных уравнений. Выбирается начальный вектор  $u_0$ , а затем последовательно строятся векторы  $u_1 = Au_0$ ,  $u_2 = Au_1$  и в общем случае  $u_{k+1} = Au_k$ . Каждый шаг представляет собой умножение вектора на матрицу, в результате чего через  $k$  шагов получаем  $u_k = A^k u_0$ , хотя матрица  $A^k$  нигде не используется. В самом деле, важно, чтобы умножение на  $A$  осуществлялось очень просто (если матрица большого порядка, ее лучше разредить), потому что сходимость к собственному вектору часто очень медленная. Если предположить, что  $A$  имеет полный набор собственных векторов  $x_1, \dots, x_n$ , то вектор  $u_k$  будет иметь обычное для разностных уравнений разложение

$$u_k = c_1 \lambda_1^k x_1 + \dots + c_n \lambda_n^k x_n. \quad (14)$$

Предположим, что собственные значения пронумерованы в возрастающем порядке и что наибольшее собственное значение единственно, т. е. нет другого собственного значения с таким же модулем и кратность  $\lambda_n$  равна 1, т. е.  $|\lambda_1| \leq \dots \leq |\lambda_{n-1}| < |\lambda_n|$ . Тогда если начальное приближение  $u_0$  содержит в качестве компоненты собственный вектор  $x_n$  с коэффициентом  $c_n \neq 0$ , то эта компонента постепенно становится доминирующей:

$$\frac{u_k}{\lambda_n^k} = c_1 \left(\frac{\lambda_1}{\lambda_n}\right)^k x_1 + \dots + c_{n-1} \left(\frac{\lambda_{n-1}}{\lambda_n}\right)^k x_{n-1} + c_n x_n. \quad (15)$$

Вектор  $u_k$  все более и более приближается к направлению  $x_n$ , и множителем, характеризующим сходимость, будет отношение  $r = |\lambda_{n-1}| / |\lambda_n|$ . Это напоминает сходимость к стационарному состоянию, которую мы изучали для марковских процессов, с тем отличием, что теперь наибольшее собственное значение  $\lambda_n$  может не равняться единице. В самом деле, мы не знаем масштабирующего множителя  $\lambda_n^k$  в (15), но и некоторый масштабирующий множитель необходимо ввести, так как в противном случае  $u_k$  может стать очень большим или очень маленьким, в зависимости от того, будет ли  $|\lambda_n| > 1$  или  $|\lambda_n| < 1$ . Обычно мы можем просто делить каждый вектор  $u_k$  на его первую компоненту  $\alpha_k$  перед тем, как сделать следующий шаг. С этим простым масштабированием степенной метод выглядит следующим образом:  $u_{k+1} = Au_k / \alpha_k$  и он сходится к кратному вектору  $x_n$ <sup>1</sup>).

**Пример** о Калифорнии (см. выше) с собственным значением единица и собственным вектором  $\begin{bmatrix} 0.667 \\ 0.333 \end{bmatrix}$ :

$A = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix}$  была матрицей перемещения населения;

$$u_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad u_1 = \begin{bmatrix} 0.9 \\ 0.1 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0.83 \\ 0.17 \end{bmatrix}, \quad u_3 = \begin{bmatrix} 0.781 \\ 0.219 \end{bmatrix}, \quad u_4 = \begin{bmatrix} 0.747 \\ 0.253 \end{bmatrix}.$$

Основные недостатки метода ясно видны из этого примера. Если  $r$  близко к 1, то сходимость очень медленная. Во многих приложениях  $r > 0.9$ , и это означает, что необходимо более 20 итераций, чтобы изменить  $(\lambda_2/\lambda_1)^k$  в 10 раз. (В нашем примере  $r = 0.7$ , но сходимость тем не менее медленная). Конечно, если  $r = 1$ , что означает  $|\lambda_{n-1}| = |\lambda_n|$ , то сходимости может вообще не быть. Есть много путей для того, чтобы обойти это препятствие, и мы опишем три из них:

(1) *Блочный степенной метод* работает с несколькими векторами одновременно вместо одного вектора  $u_k$ . Если мы начнем с  $p$  ортонормированных векторов, умножим их все на  $A$  и затем применим к ним ортогонализацию Грама—Шмидта (это единственный шаг метода), то в результате отношением, характеризующим сходимость процесса, станет величина  $r' = |\lambda_{n-p}| / |\lambda_n|$ . Более того, мы одновременно получим приближения к  $p$  различным собственным значениям и соответствующим им собственным векторам.

(2) *Обратный степенной метод* работает с  $A^{-1}$  вместо  $A$ . Единичный шаг процесса имеет вид  $v_{k+1} = A^{-1}v_k$ , и это означает, что

<sup>1</sup> Масштабирующие множители  $\alpha_k$  будут также сходиться, приближаясь к величине  $\lambda_n$ .

мы решаем линейную систему  $Au_{k+1} = v_k$  (и сохраняем множители  $L$  и  $U$ ). В данном случае теория гарантирует сходимость к *наименьшему собственному значению*, если множитель  $r'' = |\lambda_1|/|\lambda_2|$ , характеризующий сходимость, будет меньше 1. В приложениях часто требуется найти наименьшее собственное значение, и тогда автоматически выбирается обратный степенной метод.

(3) *Обратный степенной метод со сдвигом* является наилучшим из всех. Предположим, что  $A$  заменена на  $A - \alpha I$ . Тогда все собственные значения  $\lambda_i$  сдвигаются на эту же величину  $\alpha$  и множитель, характеризующий сходимость для обратного метода, становится равным  $r''' = |\lambda_1 - \alpha|/|\lambda_2 - \alpha|$ . Поэтому, если  $\alpha$  выбрано так, что оно является хорошей аппроксимацией для  $\lambda_1$ , то  $r'''$  будет очень маленьким и сходимость неизмеримо ускорится. Каждый шаг метода состоит в решении системы  $(A - \alpha I)w_{k+1} = w_k$ , и этому разностному уравнению удовлетворяет

$$w_k = \frac{c_1 x_1}{(\lambda_1 - \alpha)^k} + \frac{c_2 x_2}{(\lambda_2 - \alpha)^k} + \dots + \frac{c_n x_n}{(\lambda_n - \alpha)^k}.$$

Если  $\alpha$  близко к  $\lambda_1$ , то знаменатель первого члена этого представления будет близок к нулю и, следовательно, потребуется только один или два шага, чтобы сделать первый член полностью доминирующим. В частности, если  $\lambda_1$  уже вычислено с помощью другого алгоритма (например,  $QR$ ), то в качестве  $\alpha$  можно взять это вычисленное значение. Стандартная процедура состоит в разложении  $A - \alpha I$  в произведение  $LU^1$ ) и решении системы  $Ux_1 = (1, 1, \dots, 1)^T$  путем обратной подстановки.

Если для  $\lambda_1$  еще не получено приближение при помощи независимого алгоритма, то степенной метод со сдвигом должен сам производить выбор  $\alpha$  или (так как мы можем, если хотим, менять сдвиг на каждом шаге) он должен выбирать  $\alpha_k$ , входящее в соотношении  $(A - \alpha_k I)w_{k+1} = w_k$ . Простейшая возможность состоит в том, чтобы просто работать с масштабирующим множителем, который приводит  $w_k$  к приемлемым размерам. Существуют, однако, и другие, лучшие приемы. В симметрическом случае  $A = A^T$  наилучшим выбором, по-видимому, является отношение Релея

$$\alpha_k = R(u_k) = \frac{u_k^T A u_k}{u_k^T u_k}.$$

<sup>1)</sup> Эта процедура может показаться исключительно плохо обусловленной, так как матрица  $A - \alpha I$  близка к вырожденной настолько, насколько это можно было сделать вычитанием  $\alpha I$  из  $A$ . К счастью, погрешность главным образом сосредоточена в направлении собственного вектора матрицы  $A$ . Так как любой собственный вектор, умноженный на число, тоже является собственным вектором, то это единственное направление, которое мы хотим вычислить.

Мы уже знаем, что это отношение достигает минимума на собственном векторе: производные  $R$  равны нулю, его график представляет собой дно параболы. Поэтому погрешность  $\lambda - \alpha_k$  собственного значения приблизительно равна квадрату погрешности собственного вектора. Множитель сходимости  $r''' = |\lambda_1 - \alpha_k| / |\lambda_2 - \alpha_k|$  изменяется на каждом шаге и в действительности стремится к нулю. Окончательным результатом при выборе в качестве сдвига отношения Релея является *кубическая сходимость*<sup>1)</sup>  $\alpha_k$  к  $\lambda_1$ .

**Упражнение 7.3.1.** Для матрицы  $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$  с собственными значениями  $\lambda_1 = 1$  и  $\lambda_2 = 3$  проделать три итерации степенного метода  $u_{k+1} = Au_k$  с начальным приближением  $u_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ . Каким будет предельный вектор  $u_\infty$ ?

**Упражнение 7.3.2.** Для этой же матрицы  $A$  и начального приближения  $u_0 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$  сравнить результаты:  
(i) трех шагов обратного степенного метода

$$u_{k+i} = A^{-1}u_k = \frac{1}{3} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} u_k;$$

(ii) одного шага  $u_1 = (A - \alpha I)^{-1} u_0$  обратного степенного метода со сдвигом, где  $\alpha = u_0^T A u_0 / u_0^T u_0$ .  
Теперь предельный вектор  $u_\infty$  будет кратным другому собственному вектору  $(1, 1)$  матрицы  $A$ .

### ТРЕХДИАГОНАЛЬНАЯ И ХЕССЕНБЕРГОВА ФОРМЫ МАТРИЦЫ

Степенной метод рекомендуется только для сильно разреженных матриц большого размера. Когда большинство элементов матрицы ненулевые, то применять степенной метод невыгодно. Поэтому возникает вопрос, существует ли простой прием получения нулей в матрице? Ответ на него будет основной целью этого раздела.

Заметим сначала, что после нахождения подобной матрицы  $U^{-1}AU$  с большим количеством нулей, чем в  $A$ , мы не предполагаем возвращаться к степенному методу. Существуют более сложные разновидности и лучшей из них, по-видимому, будет  $QR$ -алгоритм. (Обратный степенной метод со сдвигом, однако,

<sup>1)</sup> Линейная сходимость означает, что на каждом шаге погрешность умножается на фиксированный множитель  $r < 1$ . Квадратичная сходимость означает, что погрешность возводится в квадрат на каждом шаге, как в методе Ньютона  $x_{k+1} - x_k = -f(x_k)/f'(x_k)$  для решения уравнения  $f(x) = 0$ . Кубическая сходимость означает, что погрешность возводится в куб на каждом шаге, убывая, например, за два шага сначала от  $10^{-1}$  до  $10^{-3}$ , а затем до  $10^{-9}$ .

применяется в самом конце для нахождения собственных векторов.) В любом случае первый шаг состоит в получении у матрицы как можно большего количества нулевых элементов и осуществления этого по возможности наиболее быстрым способом. Наше ограничение на скорость состоит только в том, что мы будем использовать лишь унитарные (или ортогональные) преобразования, которые сохраняют симметричность и длину. Иначе говоря, если  $A$  — симметрическая матрица, то симметрической будет и матрица  $U^{-1}AU$  и ни один ее элемент не становится опасо большим.

Есть по крайней мере две возможности преобразовать  $A$  в  $U^{-1}AU$ : либо мы получаем один нуль на каждом шаге (как в исключении Гаусса), либо мы можем работать с целыми столбцами. Для получения одного нуля достаточно использовать вращение плоскости, как это показано ниже в формуле (21). Матрица вращения имеет блок порядка 2, в котором стоят  $\cos \theta$  и  $\sin \theta$ . Аналогично мы поступаем со всеми элементами, стоящими ниже диагонали, выбирая на каждом шаге угол вращения  $\theta$  так, чтобы получить нуль. Это основной принцип *метода Якоби*.

К сожалению, мы не в состоянии диагонализировать  $A$  за конечное число вращений, так как нули, полученные на более ранних шагах, исчезнут на последующих.

Для того чтобы сохранить эти полученные нули, мы должны ограничиться приведением матрицы к почти треугольной форме, сохраняя *одну ненулевую диагональ ниже главной диагонали*. Это так называемая форма Хессенберга. Если матрица симметрическая, то верхний треугольник является зеркальным отражением нижнего треугольника этой матрицы и, следовательно, матрица будет трехдиагональной.

Обе эти формы получаются с помощью серии вращений в соответствующих плоскостях, но Хаусхолдер нашел новый способ решения той же задачи. Его идея дает «подготовительный шаг» для  $QR$ -алгоритма<sup>1)</sup>.

*Преобразование Хаусхолдера*, или элементарное отражение, осуществляет матрица вида

$$H = I - 2 \frac{vv^T}{\|v\|^2}.$$

Часто  $v$  нормируют и используют вектор единичной длины  $u = v/\|v\|$ . Тогда  $H$  записывается в виде  $I - 2uu^T$ . В обоих случаях  $H$  — симметрическая и ортогональная матрица:

$$H^T H = (I - 2uu^T)(I - 2uu^T) = I - 4uu^T + 4uu^T uu^T = I,$$

<sup>1)</sup> Можете, если хотите, пропустить этот подготовительный этап и прямо перейти к  $QR$ -алгоритму, изложенному на стр. 340. Только при вычислениях будет необходимо предварительное получение нулей.

и, следовательно,  $H = H^T = H^{-1}$ . В комплексном случае соответствующая матрица  $I - 2uu^H$  эрмитова и унитарна. Метод Хаусхолдера позволяет получать нули с помощью этих матриц и существенно опирается на следующее тождество.

|| **7E.** *Предположим, что  $z$  — это вектор-столбец  $(1, 0, \dots, 0)^T$ ,  $\sigma = \|x\|$  и  $v = x + \sigma z$ . Тогда  $Hx = -\sigma z = (-\sigma, 0, \dots, 0)^T$ .*

**Доказательство.**

$$Hx = x - \frac{2vv^T x}{\|v\|^2} = x - (x + \sigma z) \frac{2(x + \sigma z)^T x}{(x + \sigma z)^T (x + \sigma z)} = \\ = x - (x + \sigma z) \quad (\text{так как } x^T x = \sigma^2) = -\sigma z. \quad (16)$$

Это тождество может быть сразу же использовано. Мы начнем с первого столбца  $A$  и будем помнить, что в результате должны получить трехдиагональную форму или форму Хессенберга  $U^{-1}AU$  матрицы  $A$ . Поэтому сейчас мы рассмотрим только  $n-1$  элементов ниже диагонали:

$$x = \begin{bmatrix} a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{bmatrix}, \quad z = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad Hx = \begin{bmatrix} -\sigma \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (17)$$

В данном случае матрица Хаусхолдера  $H$  порядка только  $n-1$ , и поэтому она вкладывается в нижний правый угол матрицы  $U_1$  полного порядка. Итак,

$$U_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & & & & \\ 0 & H & & & \\ 0 & & & & \\ 0 & & & & \end{bmatrix} = U_1^{-1}, \quad \text{и} \quad U_1^{-1}AU_1 = \begin{bmatrix} a_{11} & * & * & * & * \\ -\sigma & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \end{bmatrix}.$$

Так как в левом верхнем углу стоит 1, матрица  $U_1$  оставляет элемент  $a_{11}$  неизменным и, что более важно, не трогает нулей, которые появляются в (17). Поэтому первый этап полностью завершен, т. е. матрица  $U_1^{-1}AU_1$  имеет требуемый первый столбец.

Второй этап аналогичен первому:  $x$  состоит из последних  $n-2$  элементов второго столбца,  $z$  — единичный вектор соответствующей размерности и  $H_2$  — матрица порядка  $n-2$ . Если мы ее включим в  $U_2$ , то в результате получим

$$U_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & & & \\ 0 & 0 & H_2 & & \\ 0 & 0 & & & \end{bmatrix} = U_2^{-1}, \quad U_2^{-1}(U_1^{-1}AU_1)U_2 = \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \end{bmatrix}.$$

Наконец,  $U_3$  преобразовывает третий столбец, и мы получаем форму Хессенберга для матрицы порядка 5. В общем случае  $U$  является произведением всех матриц  $U_1 U_2 \dots U_{n-2}$  и количество операций, требующихся для ее вычисления, является величиной порядка  $n^3$ .

**Пример.**

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \quad x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad H = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}.$$

Вкладывая  $H$  в  $U$ , получаем трехдиагональную матрицу

$$U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix}, \quad U^{-1}AU = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Матрица  $U^{-1}AU$  имеет удобный вид для нахождения собственных значений; здесь можно применить  $QR$ -алгоритм, но мы позволим себе небольшое отступление, чтобы обратить внимание на другие применения преобразования Хаусхолдера.

1. *Разложение  $A = QR$ .* Мы дадим краткое описание процесса Грама—Шмидта из гл. 3, излагая его более просто и более четко. Напомним, что через  $R$  обозначается верхняя треугольная матрица (мы уже не должны допускать существование ненулевой диагонали ниже главной, так как у нас теперь не будет множителей  $U$  или  $H$  справа (как в преобразовании  $U^{-1}AU$ ), «портящих» уже полученные нули). Итак, на первом шаге построения  $Q$  мы имеем дело со всем первым столбцом матрицы  $A$ :

$$x = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix}, \quad z = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad v = x + \|x\|z, \quad H_1 = I - 2 \frac{vv^T}{\|v\|^2}.$$

При этом первый столбец матрицы  $H_1 A$  будет как раз таким, как нам надо: он равен  $-\|x\|z$ , элементы, расположенные под главной диагональю, равны нулю и это первый столбец  $R$ . Второй шаг осуществляет преобразование второго столбца матрицы  $H_1 A$  от диагонального элемента и ниже. В результате получаем матрицу  $H_2 H_1 A$ , у которой все элементы второго столбца ниже диагонального равны нулю. (Весь алгоритм очень похож на метод исключения Гаусса, и на самом деле это несколько более медленный его вариант.) Результатом  $n-1$  шага снова будет верхняя треугольная матрица  $R$ , но матрица, осуществляющая шаги, не будет нижней треугольной  $L$ . Напротив, эта матрица ортого-

нальна и имеет вид  $Q = H_1 H_2 \dots H_{n-1}$ . Ее можно хранить как последовательность сомножителей и не вычислять в явном виде. Этим заканчивается описание процесса Грама—Шмидта.

II. *Сингулярное разложение*  $Q_1^T A Q_2 = \Sigma$ . На стр. 170 это разложение немедленно давало для любой задачи оптимальное решение  $\bar{x}$  по методу наименьших квадратов. Напомним, что  $\Sigma$ —диагональная матрица того же порядка, что и  $A$ , и что ее элементы (сингулярные числа  $\mu$ ) являются собственными значениями матрицы  $A^T A$ . Поскольку преобразования Хаусхолдера позволяют только *подготовить* матрицу для решения задачи на собственные значения, но не решают эту задачу, мы не можем надеяться, что их использование дает нам матрицу  $\Sigma$ . Вместо этого они позволяют построить *двухдиагональную матрицу* с нулями везде, кроме главной диагонали и одной диагонали выше нее. Конечно, этот процесс численно устойчив, так как матрицы  $H$  ортогональны.

Первый шаг процесса такой же, как и в описанном выше  $QR$ -разложении: в качестве  $x$  выбирается первый столбец  $A$ , и тогда все элементы первого столбца  $H_1 A$ , расположенные ниже диагонального элемента, равны нулю. Следующий шаг состоит в умножении  $H_1 A$  справа на матрицу  $H^{(1)}$ , которая зануляет элементы первой строки:

$$A \rightarrow H_1 A = \begin{bmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{bmatrix} \rightarrow H_1 A H^{(1)} = \begin{bmatrix} * & * & 0 & 0 \\ 0 & * & * & * \\ 0 & * & * & * \end{bmatrix}. \quad (18)$$

Второе преобразование Хаусхолдера дает

$$H_2 H_1 A H^{(1)} = \begin{bmatrix} * & * & 0 & 0 \\ 0 & * & * & * \\ 0 & 0 & * & * \end{bmatrix} \quad \text{и} \quad H_2 H_1 A H^{(1)} H^{(2)} = \begin{bmatrix} * & * & 0 & 0 \\ 0 & * & * & 0 \\ 0 & 0 & * & * \end{bmatrix}.$$

В результате получаем ту двухдиагональную форму, о которой говорили. Итак, мы проиллюстрировали снова, как можно быстро получить нули на нужных местах при помощи преобразований Хаусхолдера.

**Упражнение 7.3.3.** Показать, что для любых двух различных векторов  $x, y$  одинаковой длины  $\|x\| = \|y\|$  выбор  $v = x - y$  приводит к преобразованию Хаусхолдера, такому, что  $Hx = y$  и  $Hv = x$ .

**Упражнение 7.3.4.** Для

$$x = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \quad \text{и} \quad z = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

вычислить  $\sigma = \|x\|$ ,  $v = x + \sigma z$  и соответствующую матрицу Хаусхолдера  $H$ . Проверить, что  $Hx = -\sigma z$ .

Упражнение 7.3.5. Используя 7.3.4, найти трехдиагональную матрицу  $U^{-1}AU$  для матрицы

$$A = \begin{bmatrix} 1 & 3 & 4 \\ 3 & 1 & 0 \\ 4 & 0 & 0 \end{bmatrix}.$$

### QR-АЛГОРИТМ

Этот алгоритм почти мистически прост. Он начинается с матрицы  $A_0$ , которая разлагается по Граму—Шмидту в произведение  $Q_0R_0$ , а затем меняются местами сомножители:  $A_1 = R_0Q_0$ . Эта матрица подобна начальной,  $Q_0^{-1}A_0Q_0 = Q_0^{-1}(Q_0R_0)Q_0 = A_1$ . Этот процесс продолжается, причем собственные значения не изменяются:

$$A_k = Q_kR_k \quad \text{и} \quad A_{k+1} = R_kQ_k. \quad (19)$$

Эта формула описывает *QR-алгоритм без сдвигов*, и при довольно общих предположениях он сходится:  $A_k$  сходятся к треугольной матрице, и, следовательно, их диагональные элементы стремятся к ее собственным значениям, которые в свою очередь являются собственными значениями исходной матрицы  $A_0$ <sup>1)</sup>.

Установлено, что рассмотренный алгоритм хорош, но не очень. Для того чтобы сделать его лучше, надо ввести два усовершенствования: а) мы должны использовать сдвиги и (б) должны быть уверены, что *QR*-разложение на каждом шаге производится очень быстро.

(а) *Алгоритм со сдвигом*. Если число  $\alpha_k$  близко к собственному значению, то шаг (19) немедленно заменяется на следующий:

$$A_k - \alpha_k I = Q_kR_k \quad \text{и} \quad \text{затем} \quad A_{k+1} = R_kQ_k + \alpha_k I. \quad (20)$$

Такой переход не меняет собственных значений, так как  $A_{k+1}$  подобна  $A_k$ :

$$Q_k^{-1}A_kQ_k = Q_k^{-1}(Q_kR_k + \alpha_k I)Q_k = A_{k+1}.$$

На практике оказывается, что элемент в позиции  $(n, n)$  матрицы  $A_k$  (в нижнем правом углу) является первым приближением для собственного значения. Поэтому этот элемент используется в качестве простейшего и наиболее популярного выбора величины сдвига  $\alpha_k$ . Обычно такой подход дает квадратичную сходимость, а в симметрическом случае — кубическую сходимость к наименьшему собственному значению. После, вероятно, трех-четырёх

<sup>1)</sup>  $A_0$  является матрицей, с которой начинается *QR*-алгоритм. Если она уже сама получена при помощи преобразований Хаусхолдера, которые привели ее к трехдиагональному виду, то  $A_0$  связана с начальной матрицей  $A$  равенством  $U^{-1}AU = A_0$ .

шагов алгоритма со сдвигом матрица  $A_k$  выглядит примерно так:

$$A_k = \left[ \begin{array}{ccc|c} * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \\ \hline 0 & 0 & \varepsilon & \lambda_1' \end{array} \right] \quad \text{с} \quad \varepsilon \ll 1.$$

Мы предполагаем, что вычисленное  $\lambda_1'$ — очень близкое приближение к искомому числу  $\lambda_1$ . Для того чтобы найти следующее собственное значение,  $QR$ -алгоритм продолжают с меньшей матрицей (порядка три на рисунке) в верхнем левом углу. Ее поддиагональные элементы уже несколько изменены на первых шагах  $QR$ -алгоритма, и следующих двух шагов достаточно для нахождения  $\lambda_2$ . Это дает систематическую процедуру для нахождения всех собственных значений. **Теперь  $QR$ -алгоритм полностью описан** и осталось определить только собственные векторы, что делается за один шаг обратного степенного метода, и пользоваться обращением в нуль соответствующих элементов матрицы в процедуре Хаусхолдера.

(b) Цель предварительных преобразований Хаусхолдера, которые придают  $A_0$  трехдиагональный вид или форму Хессенберга— сделать каждый шаг  $QR$ -алгоритма очень быстрым. Обычно процесс Грама—Шмидта (т. е.  $QR$ -разложение) требует  $O(n^3)$  действий, но для хессенберговой матрицы требуется  $O(n^2)$ , а для трехдиагональной  $O(n)$  действий. Без этого улучшения алгоритм будет невозможно медленным, и, если не обеспечивается, чтобы каждая новая матрица  $A_k$  снова имела трехдиагональный вид или форму Хессенберга, указанное улучшение можно применить только на первом шаге и далее оно бесполезно.

К счастью, этого не случается. Чтобы показать, что  $A_1$  имеет ту же форму, что и  $A_0$ , рассмотрим матрицу

$$Q_0 = A_0 R_0^{-1} = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \end{bmatrix} \begin{bmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{bmatrix}.$$

Вы можете легко проверить, что это умножение сохраняет  $Q_0$  те три нуля, которые имеет  $A_0$ , т. е.  $Q_0$  сама имеет форму Хессенберга. Отсюда следует, что матрица  $A_1$ , которая формируется посредством перестановки сомножителей, имеет вид

$$A_1 = R_0 Q_0 = \begin{bmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{bmatrix} \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \end{bmatrix},$$

т. е. эти три нуля снова появляются в произведении, и, следовательно,  $A_1$  имеет форму Хессенберга всякий раз, когда  $A_0$  имеет эту форму. Симметрический случай даже лучше, так как  $A_1 = Q_0^{-1}A_0Q_0$  остается симметрической:

$$A_1^T = Q_0^T A_0^T (Q_0^{-1})^T = Q_0^{-1} A_0 Q_0 = A_1.$$

По только что изложенной причине построенная матрица  $A_1$  имеет форму Хессенберга, и поскольку она одновременно симметрическая, то она является *трехдиагональной* матрицей. То же самое относится к каждой из последующих матриц  $A_2, A_3, \dots$ . Итак, каждый шаг  $QR$ -алгоритма имеет дело с трехдиагональной матрицей.

В заключение поясним процесс построения матриц  $Q_0$  и  $R_0$  для начальной матрицы  $A_0$  (и  $Q_k$ , и  $R_k$  для каждой  $A_k$  или на самом деле для  $A_k - \alpha_k I$ ). Здесь снова можно использовать преобразования Хаусхолдера, но проще уничтожать каждый поддиагональный элемент простым «вращением плоскости». Первое из них имеет вид

$$P_{21}A_0 = \begin{bmatrix} \cos \theta & -\sin \theta & & \\ \sin \theta & \cos \theta & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \begin{bmatrix} a_{11} & * & * & * \\ a_{21} & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \end{bmatrix}. \quad (21)$$

Элемент, стоящий в позиции (2, 1) этого произведения, равен  $a_{11} \sin \theta + a_{21} \cos \theta$ , и мы выбираем угол  $\theta$  так, чтобы он сделал эту комбинацию равной нулю. Затем  $P_{32}$  выбираем таким же способом, чтобы элемент матрицы  $P_{32}P_{21}A_0$  в позиции (3,2) был равен нулю. После  $n-1$  элементарных вращений в качестве конечного результата получаем верхнюю треугольную матрицу  $R_0$ :

$$R_0 = P_{nn-1} \dots P_{32}P_{21}A_0. \quad (22)$$

Это все, что мы позволили себе сказать (более подробно этот метод изложен в книге Уилкинсона и Стюарта) об одном из наиболее замечательных алгоритмов вычислительной математики.

**Упражнение 7.3.6.** Показать, что для начальной матрицы  $A_0 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$  алгоритм без сдвига дает лишь более чем скромное улучшение. Здесь  $A_1 = \frac{1}{5} \begin{bmatrix} 14 & -3 \\ -3 & 6 \end{bmatrix}$ .

**Упражнение 7.3.7.** Сделать простой шаг  $QR$ -алгоритма для матрицы

$$A = \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & 0 \end{bmatrix}$$

со сдвигом  $\alpha = a_{22}$ , который в данном случае совпадает с  $QR$ -алгоритмом без сдвигов, поскольку  $a_{22} = 0$ . Показать, что внедиагональные элементы изменяются от  $\sin \theta$  до  $-\sin^3 \theta$  (пример кубической сходимости).

**Упражнение 7.3.8.** Показать, что трехдиагональная матрица  $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$  остается неизменной на каждом шаге  $QR$ -алгоритма и потому является (редким) контрпримером к сходимости. Это явление обходится путем использования произвольного сдвига.

**Упражнение 7.3.9.** Показать по индукции, что для  $QR$ -алгоритма без сдвигов произведение  $(Q_0 Q_1 \dots Q_k)(R_k \dots R_1 R_0)$  является в точности  $QR$ -разложением матрицы  $A^{k+1}$ . Это совпадение связывает  $QR$ -алгоритм со степенным методом и объясняет его сходимость: если  $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$ , то эти собственные значения постепенно появляются в убывающем порядке на главной диагонали матрицы  $A_k$ .

## § 7.4. ИТЕРАЦИОННЫЕ МЕТОДЫ РЕШЕНИЯ СИСТЕМЫ $Ax=b$

В отличие от задач на собственные значения, где у нас нет выбора, для решения системы  $Ax=b$  нам не столь уж необходим итерационный метод. Исключение Гаусса дает решение  $x$  за конечное число шагов, и, пока это число приемлемо, мы можем применять этот метод. С другой стороны, когда  $n^3/3$  велико, мы можем получить приближение для  $x$  более быстрым способом. И нет смысла начинать процедуру исключения, а затем бросать. Наша цель — описать методы, которые начинаются с некоторого начального приближения  $x_0$ , а затем позволяют строить улучшенные приближения  $x_{k+1}$ , используя предыдущее приближение  $x_k$ , и остановить этот процесс когда угодно.

Такие методы легко построить путем простого *расщепления матрицы*  $A$ . Если  $A = S - T$ , то система  $Ax=b$  эквивалентна системе  $Sx = Tx + b$  и, следовательно, мы можем попытаться итерировать, используя формулу

$$Sx_{k+1} = Tx_k + b. \quad (23)$$

Конечно, нет гарантии, что любой такой метод хорош, и, для того чтобы он был успешным, расщепление должно удовлетворять двум различным требованиям:

(i) Новый вектор  $x_{k+1}$  должен легко вычисляться. Поэтому  $S$  должна быть простой (и обратимой!) матрицей. Она может быть диагональной или треугольной матрицей.

(ii) Последовательность  $x_k$  должна сходиться к точному решению  $x$ . Если мы вычтем (23) из исходного уравнения  $Sx = Tx + b$ ,

то в результате получим формулу, включающую в себя только ошибки  $e_k = x - x_k$ :

$$Se_{k+1} = Te_k. \quad (24)$$

Это совсем другое разностное уравнение. Для него итерационный процесс начинается с начальной ошибки  $e_0$ , и после  $k$  шагов мы получаем новую ошибку  $e_k = (S^{-1}T)^k e_0$ . Вопрос сходимости здесь эквивалентен вопросу устойчивости:  $x_k \rightarrow x$  только тогда, когда  $e_k \rightarrow 0$ .

**7F.** Итерационный метод (23) сходится тогда и только тогда, когда каждое собственное значение  $\lambda$  матрицы  $S^{-1}T$  удовлетворяет неравенству  $|\lambda| < 1$ . Его скорость сходимости зависит от максимального значения  $|\lambda|$ , которое известно как **спектральный радиус** матрицы  $S^{-1}T$ :

$$\rho(S^{-1}T) = \max_i |\lambda_i|. \quad (25)$$

Напомним, что обычно решение уравнения  $e_{k+1} = S^{-1}Te_k$  представимо в виде

$$e_k = c_1 \lambda_1^k x_1 + \dots + c_n \lambda_n^k x_n.$$

Отсюда следует, что наибольшее из  $|\lambda_i|$  в конце концов станет доминирующим и будет определять скорость, с которой  $e_k$  сходится к нулю.

Два требования, предъявляемые к итерационному процессу, до некоторой степени противоречивы. С одной стороны, мы могли бы достичь немедленной сходимости, выбирая следующее расщепление:  $S = A$  и  $T = 0$ . При этом первый (и последний) шаг итерации был бы  $Ax_1 = b$ . В этом случае матрица перехода  $S^{-1}T$  равна нулю, ее собственные значения и спектральный радиус равны нулю и, следовательно, скорость сходимости (обычно определяемая как  $-\log \rho$ ) бесконечна. Но, разумеется, матрицу  $S = A$  не всегда легко обратить, что было основным побуждением для расщепления. С другой стороны, мы можем взять в качестве  $S$  диагональную часть матрицы  $A$  и итерации становятся хорошо известным *методом Якоби*:

$$\begin{aligned} a_{11}(x_1)_{k+1} &= (-a_{12}x_2 - a_{13}x_3 - \dots - a_{1n}x_n)_k + b_1, \\ &\vdots \\ a_{nn}(x_n)_{k+1} &= (-a_{n1}x_1 - a_{n2}x_2 - \dots - a_{nn-1}x_{n-1})_k + b_n. \end{aligned} \quad (26)$$

Если диагональные элементы  $a_{ii}$  все ненулевые и если  $A$  — разреженная матрица, так что большинство членов в правой части отсутствует, то переход от  $x_k$  к  $x_{k+1}$  прост. Важным является вопрос о том, сходится ли итерационный метод, и если сходится, то как быстро.

## Пример 1.

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad S = \begin{bmatrix} 2 & \\ & 2 \end{bmatrix}, \quad T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad S^{-1}T = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix}.$$

Если компоненты вектора  $x$  обозначить через  $v$  и  $w$ , то метод Якоби  $Sx_{k+1} = Tx_k + b$  имеет вид

$$\begin{aligned} 2v_{k+1} &= w_k + b_1, \\ 2w_{k+1} &= v_k + b_2, \end{aligned} \quad \text{или} \quad \begin{bmatrix} v \\ w \end{bmatrix}_{k+1} = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix}_k + \begin{bmatrix} b_1/2 \\ b_2/2 \end{bmatrix}.$$

Собственными значениями матрицы  $S^{-1}T$ , имеющей решающее значение, служат  $\pm 1/2$ . Последнее означает, что ошибка убывает в два раза (приближение становится точнее на один двоичный знак) на каждом шаге. В этом примере, который слишком прост, чтобы быть типичным, сходимость быстрая. Если мы попытаемся взять матрицу  $A$  большего порядка, то немедленно возникают практические трудности при реализации метода Якоби (26). *Требуется хранить в памяти все компоненты  $x_k$ , пока не вычислено полностью приближение  $x_{k+1}$ .* Более естественная идея, которая требует только половину памяти по сравнению с методом Якоби, — это начинать использовать каждую компоненту нового вектора  $x_{k+1}$ , как только она вычислена; вектор  $x_{k+1}$  постепенно, по одной компоненте, становится на место  $x_k$  и поэтому  $x_k$  исчезает, как только завершается построение вектора  $x_{k+1}$ . Сказанное означает, что первое уравнение остается таким же, как и раньше:

$$a_{11}(x_1)_{k+1} = (-a_{12}x_2 - a_{13}x_3 - \dots - a_{1n}x_n)_k + b_1.$$

Следующее уравнение немедленно оперирует с этим новым значением  $x_1$ :

$$a_{22}(x_2)_{k+1} = -a_{21}(x_1)_{k+1} + (-a_{23}x_3 - \dots - a_{2n}x_n)_k + b_2,$$

а последнее уравнение будет использовать исключительно новые значения:

$$a_{nn}(x_n)_{k+1} = (-a_{n1}x_1 - a_{n2}x_2 - \dots - a_{n,n-1}x_{n-1})_{k+1} + b_n.$$

Этот метод носит название *метода Гаусса—Зейделя*, даже не смотря на то, что он, по-видимому, не был известен Гауссу и не предлагался Зейделем. Это забавно, поскольку сам метод является неплохим. Заметим, что если все компоненты  $x_{k+1}$  перенести в левую часть, то очевидно, что *матрица  $S$  станет нижней треугольной частью  $A$* , а в правой части остается матрица  $T$  из представления  $A = S - T$ , которая является верхней строго треугольной.

## Пример 2.

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad S = \begin{bmatrix} 2 & 0 \\ -1 & 2 \end{bmatrix}, \quad T = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad S^{-1}T = \begin{bmatrix} 0 & \frac{1}{2} \\ 0 & \frac{1}{4} \end{bmatrix}.$$

Один шаг метода Гаусса—Зейделя преобразует компоненты  $v_k$  и  $w_k$  по формулам

$$\begin{aligned} 2v_{k+1} &= w_k + b_1, \\ 2w_{k+1} &= v_{k+1} + b_2, \end{aligned} \quad \text{или} \quad \begin{bmatrix} 2 & 0 \\ -1 & 2 \end{bmatrix} x_{k+1} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x_k + b.$$

Собственные значения матрицы  $S^{-1}T$  снова имеют решающее значение, и их, конечно, легко найти. Они равны 0 и  $1/4$ . Ошибка делится на 4 каждый раз, так что *один шаг метода Гаусса—Зейделя эквивалентен двум шагам метода Якоби*<sup>1)</sup>. Так как оба метода требуют одного и того же количества операций (мы будем использовать новые значения вместо старых и практически экономим память), то метод Гаусса—Зейделя лучше.

Существует способ сделать метод Гаусса—Зейделя еще лучше. Еще во времена ручных вычислений было замечено (возможно, случайно), что сходимость становится быстрее, если способ вычисления поправки  $x_{k+1} - x_k$  несколько отличается от используемого в методе Гаусса—Зейделя. Точнее говоря, обычный метод сходится монотонно, т. е. приближения остаются с одной и той же стороны от решения  $x$ . Поэтому естественно попытаться ввести *релаксационный множитель*  $\omega$ , чтобы подойти к решению ближе. При  $\omega = 1$  такой способ совпадает с методом Гаусса—Зейделя, а при  $\omega > 1$  известен как метод *последовательной верхней релаксации* (SOR). Оптимальный выбор  $\omega$  зависит от задачи, но никогда не превосходит 2 и часто приблизительно равен 1.9.

Чтобы описать этот метод более подробно, обозначим через  $D$ ,  $L$  и  $U$  диагональную, нижнюю строго треугольную и верхнюю строго треугольную части  $A$ . (Это расщепление не имеет ничего общего с разложением Гаусса  $A = LDU$  и на самом деле имеет вид  $A = L + D + U$ .) В методе Якоби  $S = D$  в левой части (23) и  $T = -L - U$  в правой части, тогда как в методе Гаусса—Зейделя выбирается расщепление  $S = D + L$  и  $T = -U$ . Сейчас, чтобы ускорить сходимость, мы рассмотрим итерационный метод

$$(D + \omega L) x_{k+1} = [(1 - \omega)D - \omega U] x_k + \omega b. \quad (27)$$

<sup>1)</sup> Это правило верно для очень большого класса матриц, даже несмотря на то, что можно сконструировать и другие примеры, в которых метод Якоби сходится, а метод Гаусса—Зейделя нет (или наоборот). Симметрический случай наиболее ясный: если  $a_{ii} > 0$  при всех  $i$ , то метод Гаусса—Зейделя сходится тогда и только тогда, когда  $A$  положительно определена.

Заметим, что при  $\omega=1$  мы не получаем ускорения, а возвращаемся точно к методу Гаусса—Зейделя. Но независимо от  $\omega$  матрица в левой части нижняя треугольная, а в правой части верхняя треугольная. Поэтому мы все еще можем покомпонентно заменять  $x_k$  на  $x_{k+1}$ , по мере того как эти компоненты вычисляются. Один шаг метода Гаусса—Зейделя имеет вид

$$a_{ii}(x_i)_{k+1} = a_{ii}(x_i)_k + \omega [(-a_{i1}x_1 - \dots - a_{i,i-1}x_{i-1})_{k+1} + (-a_{i,i}x_i - \dots - a_{in}x_n)_k + b_i].$$

Если старое приближение  $x_k$  совпадает с истинным решением  $x$ , то новое приближение  $x_{k+1}$  также совпадет с  $x$  и, следовательно, выражение в скобках будет равно нулю.

**Пример 3.** Для той же матрицы  $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$  каждый шаг метода последовательной верхней релаксации имеет вид

$$\begin{bmatrix} 2 & 0 \\ -\omega & 2 \end{bmatrix} x_{k+1} = \begin{bmatrix} 2(1-\omega) & \omega \\ 0 & 2(1-\omega) \end{bmatrix} x_k + \omega b.$$

Если мы разделим на  $\omega$  обе части уравнения, то получим две матрицы  $S$  и  $T$  расщепления  $A = S - T$  и итерации снова запишутся как  $Sx_{k+1} = Tx_k + b$ . Таким образом, матрицей  $S^{-1}T$ , собственные значения которой определяют скорость сходимости метода, будет матрица

$$L_\omega = \begin{bmatrix} 2 & 0 \\ -\omega & 2 \end{bmatrix}^{-1} \begin{bmatrix} 2(1-\omega) & \omega \\ 0 & 2(1-\omega) \end{bmatrix} = \begin{bmatrix} 1-\omega & \frac{1}{2}\omega \\ \frac{1}{2}\omega(1-\omega) & 1-\omega + \frac{1}{4}\omega^2 \end{bmatrix}.$$

Оптимальным значением параметра  $\omega$  является такое значение, при котором наибольшее собственное значение матрицы  $L_\omega$  (другими словами, спектральный радиус) будет наименьшим из возможных. Основной проблемой метода последовательной верхней релаксации является выбор оптимального параметра  $\omega$ . Начнем с замечания, что произведение собственных значений должно быть равно определителю. Рассмотрим два треугольных сомножителя, входящих в произведение, задающее  $L_\omega$ ; определитель первого из них равен  $1/4$  (после обращения  $S$ ), а второго  $4(1-\omega)^2$ . Поэтому

$$\lambda_1 \lambda_2 = \det L_\omega = (1-\omega)^2.$$

Это общее свойство: вклад первой матрицы  $(D + \omega L)^{-1}$  в определитель равен  $\det D^{-1}$ , так как все ненулевые элементы  $L$  лежат ниже главной диагонали, а вклад второй матрицы равен  $\det(1-\omega)D$ , поскольку ненулевые элементы  $U$  лежат только выше главной диагонали. Произведение этих определителей в случае

матриц порядка  $n$  равно  $\det L_\omega = (1 - \omega)^n$ . (Это объясняет, почему нельзя доходить до значения  $\omega = 2$ : произведение собственных значений было бы настолько большим, что неравенство  $|\lambda_i| < 1$  не могло бы выполняться для всех собственных значений и метод не сходился бы.) Проведенные выкладки дают ключ к пониманию поведения собственных значений матрицы  $L_\omega$ . При  $\omega = 1$  это собственные значения метода Гаусса—Зейделя, равные 0 и 1/4, а когда  $\omega$  возрастает, то собственные значения приближаются друг к другу. При оптимальном  $\omega$  оба собственных значения равны по модулю. Следовательно, в этот момент они оба должны быть равны по модулю  $\omega - 1$ , чтобы их произведение было равно определителю матрицы  $L_\omega$ <sup>1)</sup>.

Оптимальное значение  $\omega$  вычислить легко, потому что сумма собственных значений всегда равна сумме диагональных элементов матрицы (следу  $L_\omega$ ). Поэтому наилучший параметр  $\omega_{\text{опт}}$  определяется из уравнения

$$\lambda_1 + \lambda_2 = (\omega_{\text{опт}} - 1) + (\omega_{\text{опт}} - 1) = 2 - 2\omega_{\text{опт}} + \frac{1}{4} \omega_{\text{опт}}^2. \quad (28)$$

Квадратное уравнение дает  $\omega_{\text{опт}} = 4(2 - \sqrt{3}) \approx 1.07$ . Поэтому модули собственных значений приблизительно равны  $\omega - 1 = 0.07$ , что является значительным изменением по сравнению с методом Гаусса—Зейделя с его собственным значением  $\lambda = 1/4$  при  $\omega = 1$ . В этом примере правильный выбор параметра снова удвоил скорость сходимости, потому что  $(1/4)^2 \approx 0.07$ .

Обнаружение того, что это усовершенствование так легко сделать (почти как по волшебству), положило начало двадцатилетнему периоду громадной активности в численном анализе. Первой задачей было развить и расширить теорию верхней релаксации. Диссертация Янга в 1950 году содержала ее решение—простую формулу для оптимального  $\omega$ . Ключевым моментом его диссертации было установление связи между собственными значениями  $\lambda$  матрицы  $L_\omega$  и собственными значениями  $\mu$  исходной матрицы Якоби  $D^{-1}(-L-U)$ . Эта связь выражается формулой

$$(\lambda + \omega - 1)^2 = \lambda \omega^2 \mu^2. \quad (29)$$

Она верна для широкого класса конечно-разностных матриц, и если мы положим  $\omega = 1$  (метод Гаусса—Зейделя), то это дает  $\lambda^2 = \lambda \mu^2$ , откуда  $\lambda = 0$  и  $\lambda = \mu^2$ . Это было подтверждено примерами 1 и 2, в которых  $\mu = \pm 1/2$ , а  $\lambda = 0$  и  $\lambda = 1/4$ . В самом деле, ситуация здесь полностью отражает соотношение между методами Якоби и Гаусса—Зейделя: собственные значения  $\mu$  всех матриц класса Янга разбиты на «плюс-минус-пары», и тогда (29) пока-

<sup>1)</sup> Если  $\omega$  растет дальше, то собственные значения становятся комплексно-сопряженной парой: для обоих справедливо  $|\lambda| = \omega - 1$ , т. е. их произведение по-прежнему равно  $(\omega - 1)^2$  и их модули растут при возрастании  $\omega$ .

зывает, что соответствующие  $\lambda$  равны либо нулю, либо  $\mu^2$ . Скорость сходимости метода удваивается.

Важная задача—еще больше улучшить метод. Мы хотим выбрать  $\omega$  так, чтобы минимизировался модуль наибольшего собственного значения  $\lambda$ . К счастью, эта проблема уже решена. Уравнение Янга (29) есть не что иное, как характеристическое уравнение для примера матрицы второго порядка  $L_\omega$ , и лучшим было то  $\omega$ , которое делает оба корня  $\lambda$  равными по модулю  $\omega-1$ . Точно так же, как в (28), где  $\mu^2 = 1/4$ , это приводит к уравнению

$$(\omega-1) + (\omega-1) = 2 - 2\omega + \mu^2\omega^2, \quad \text{или} \quad \omega = \frac{2(1 - \sqrt{1 - \mu^2})}{\mu^2}.$$

Единственная разница состоит в том, что для больших матриц это будет повторяться для различных пар  $\pm \mu_i$ , и нам остается только сделать правильный выбор  $\omega$ . Наибольшая из этих пар собственных значений  $\mu$  дает наибольшее собственное значение в методе Якоби  $\mu_{\max}$ , и оно также дает наибольшее значение  $\omega$  и  $\lambda = \omega - 1$ . Поэтому, так как наша цель—сделать  $\lambda_{\max}$  как можно меньше, эта экстремальная пара определяет наилучший выбор  $\omega_{\text{опт}}$ :

$$\omega_{\text{опт}} = \frac{2(1 - \sqrt{1 - \mu_{\max}^2})}{\mu_{\max}^2} \quad \text{и} \quad \lambda_{\max} = \omega_{\text{опт}} - 1. \quad (30)$$

Эта формула Янга оптимального множителя для метода последовательной верхней релаксации.

Для нашей конечно-разностной матрицы с элементами  $-1, 2, -1$  вдоль трех основных диагоналей можно вычислить улучшение, которое вносит  $\omega_{\text{опт}}$ . В примерах это были матрицы порядка два. Теперь предположим, что порядок равен  $n$ , т. е. соответствует шагу сетки  $h = 1/(n+1)$ . Наибольшее собственное значение матрицы метода Якоби в соответствии с упр. 7.4.3 равно  $\mu_{\max} = \cos \pi h$ . Поэтому наибольшее собственное значение матрицы метода Гаусса—Зейделя равно  $\mu_{\max}^2 = \cos^2 \pi h$ , а максимальное собственное значение для метода последовательной верхней релаксации находится подстановкой  $\mu_{\max}$  в (30):

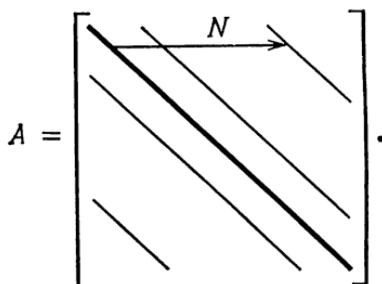
$$\lambda_{\max} = \frac{2(1 - \sin \pi h)}{\cos^2 \pi h} - 1 = \frac{(1 - \sin \pi h)^2}{\cos^2 \pi h} = \frac{1 - \sin \pi h}{1 + \sin \pi h}.$$

Оценить эти величины можно только для конкретного примера. Предположим, что  $A$  имеет порядок 21, что совсем не много. Тогда  $h = 1/22$ ,  $\cos \pi h = 0.99$  и метод Якоби сходится медленно;  $\cos^2 \pi h = 0.98$  означает, что даже метод Гаусса—Зейделя требует проведения большого числа итераций. Но так как  $\sin \pi h = \sqrt{0.02} = 0.14$ , то множитель, определяющий сходимость оптимального варианта метода верхней релаксации, равен

$$\lambda_{\max} = \frac{0.86}{1.14} = 0.75 \quad \text{при} \quad \omega_{\text{опт}} = 1 + \lambda_{\max} = 1.75.$$

На каждом шаге ошибка уменьшается на 25%, и **один шаг метода верхней релаксации эквивалентен 30 шагам по методу Якоби**:  $(0.99)^{30} = 0.75$ .

Это очень удивительный результат для такой простой идеи. В действительности этот метод применяется не к одномерным задачам (обыкновенные дифференциальные уравнения), так как система  $Ax = b$  с трехдиагональной матрицей легко решается. Верхняя релаксация наиболее важна для случая больших размерностей, что соответствует уравнениям с частными производными. Если мы заменим единичный интервал  $0 \leq x \leq 1$  на единичный квадрат  $0 \leq x, y \leq 1$ , а уравнение  $-u_{xx} = f$  на  $-u_{xx} - u_{yy} = f$ , то естественным конечно-разностным аналогом для этого уравнения будет «пятиточечная схема». Коэффициенты  $-1, 2, -1$  по направлению  $x$  комбинируются с коэффициентами  $-1, 2, -1$  по направлению  $y$ . В результате имеем главную диагональ с элементами  $+4$  и четыре других диагонали с элементами  $-1$ . Но **ширина ленты матрицы  $A$  не равна пяти**, и не существует способа нумерации  $N^2$  узловых точек в квадрате так, чтобы каждая точка располагалась близко ко всем четырем своим соседним. Если упорядочение проводится строка за строкой, то каждая точка на протяжении целой строки должна ожидать свою верхнюю соседнюю точку. В результате «пятиточечная матрица» имеет ширину ленты  $N$ :



Эта матрица привлекает большее внимание и изучается более детально, чем матрицы любых других линейных уравнений  $Ax = b$ . Я думаю, что сейчас возникла тенденция возвращения к прямым методам, базирующимся на идеях Голуба и Хоккнея, хотя ряд специальных матриц они не охватывают. (Это можно сравнить с так называемым «быстрым преобразованием Фурье».)

Ранее появились итерационные методы *переменных направлений*, в которых расщепление осуществлялось путем разделения, на трехдиагональную матрицу в направлении  $x$  и такую же в направлении  $y$ . Еще раньше появился метод верхней релаксации, поскольку собственное значение матрицы метода Якоби  $\mu_{\max} = \cos \pi h$  такое же, как и в одномерном случае, и таким же ока-

зывается оптимальный параметр  $\omega_{\text{опт}}$ . Во всех случаях трудности возникают при переходе от модельных задач к реальным, и каждый метод имеет свои особенности при решении уравнений более общих, чем  $-u_{xx}-u_{yy}=f$ , и для областей более общей геометрии, чем квадрат.

Мы не можем закончить изложение без упоминания метода сопряженных градиентов, который, казалось, умер, но внезапно вновь ожил; скорее его можно отнести к прямым, чем к итерационным, но в отличие от исключения он может быть остановлен на одном из шагов. И нет нужды говорить, что могут появиться и победить другие, совершенно новые идеи. Но тем не менее нужно сказать, что именно изменение с 0.99 до 0.75 явилось революцией в решении системы  $Ax=b$ .

**Упражнение 7.4.1.** Для матрицы

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

с собственными значениями  $2-\sqrt{2}$ ,  $2$ ,  $2+\sqrt{2}$  найти матрицу Якоби  $D^{-1}(-L-U)$  и ее собственные значения, матрицу Гаусса—Зейделя  $(D+L)^{-1}(-U)$  и ее собственные значения и числа  $\omega_{\text{опт}}$  и  $\lambda_{\text{max}}$  для метода последовательной верхней релаксации. Вычислять матрицу  $L_{\omega}$  нет необходимости.

**Упражнение 7.4.2.** Для матрицы

$$A = \begin{bmatrix} 2 & -1 & & & \\ -1 & \cdot & \cdot & & \\ & \cdot & \cdot & -1 & \\ & & & -1 & 2 \end{bmatrix}$$

порядка  $n$  описать матрицу Якоби  $J=D^{-1}(-L-U)$ . Показать, что вектор  $x_1=(\sin \pi h, \sin 2\pi h, \dots, \sin n\pi h)$ —ее собственный вектор, соответствующий ее собственному значению  $\lambda_1=\cos \pi h=\cos \pi/(n+1)$ .

**Упражнение 7.4.3.** Для той же матрицы  $A$  показать, что другие собственные векторы имеют вид  $x_k=(\sin k\pi h, \sin 2k\pi h, \dots, \sin nk\pi h)$  и соответствуют собственным значениям  $\lambda_k=\cos k\pi h$ , которые, как видно, составляют плюс-минус-пары ( $\lambda_n=-\lambda_1$ ,  $\lambda_{n-1}=-\lambda_2$ , ...), причем  $\lambda_{\text{max}}=\cos \pi h$ .

Следующие упражнения используют теорему о кругах Гершгорина: *каждое собственное значение матрицы  $A$  лежит в одном из кругов  $C_1, \dots, C_n$ , где круг  $C_i$  имеет центром диагональный элемент  $a_{ii}$ , а его радиус  $r_i=\sum_{j \neq i} |a_{ij}|$  равен сумме абсолютных значений оставшихся элементов строки.*

**Доказательство.** Уравнение  $Ax = \lambda x$  в покомпонентной записи имеет вид  $(\lambda - a_{ii})x_i = \sum_{j \neq i} a_{ij}x_j$ , откуда  $|\lambda - a_{ii}| \leq \leq \sum_{j \neq i} |a_{ij}| \frac{|x_j|}{|x_i|}$ . Если наибольшей по модулю компонентой  $x$  является компонента  $x_i$ , то  $|x_j|/|x_i| \leq 1$  для всех  $j \neq i$  и, следовательно,  $|\lambda - a_{ii}| \leq r_i$ , т. е.  $\lambda$  лежит в  $i$ -м круге.

**Упражнение 7.4.4.** Матрица

$$A = \begin{bmatrix} 3 & 1 & 1 \\ 0 & 4 & 1 \\ 2 & 2 & 5 \end{bmatrix}$$

называется матрицей с *диагональным преобладанием*, так как для каждого  $i$   $|a_{ii}| > r_i$ . Показать, что нуль не лежит ни в одном из кругов Гершгорина и, следовательно, матрица  $A$  невырожденная.

**Упражнение 7.4.5.** Выписать матрицу Якоби  $J$  для положительно определенной матрицы  $A$  из упр. 7.4.4 и найти для  $J$  три круга Гершгорина. Показать, что все радиусы удовлетворяют неравенству  $r_i < 1$ , и объяснить, почему метод Якоби будет сходиться.

**Упражнение 7.4.6.** Точное решение системы  $Ax = b$  отличается от решения системы  $LUx_0 = b$  так как матрица  $A - LU$  ненулевая из-за ошибок округления. Одна из возможностей повышения точности приближенного решения состоит в использовании везде двойной точности вычислений, но лучшим и более быстрым способом является *итеративное уточнение*. В этом способе с двойной точностью вычисляется только вектор  $r = b - Ax_0$ . Затем решается система  $LUy = r$  и вычисленное решение добавляется как коррекция к предыдущему приближенному решению. Задача заключается в следующем: умножить вектор  $x_1 = x_0 + y$  на матрицу  $LU$ , выписать результат в виде итерационной процедуры  $Sx_1 = Tx_0 + b$ , указав матрицы  $S$  и  $T$  соответствующего расщепления, и объяснить, почему  $T$  оказывается очень малой. Эта простая операция коррекции приводит нас к почти точному решению.

ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ  
И ТЕОРИЯ ИГР

## § 8.1. ЛИНЕЙНЫЕ НЕРАВЕНСТВА

Линейная алгебра, в отличие от анализа, связана с решением только уравнений и не имеет дела с неравенствами. Это всегда казалось очевидным, но в конце концов я понял, что линейное программирование представляет собой контрпример: оно связано с неравенствами, но безусловно является частью линейной алгебры. То же самое справедливо и в отношении теории игр, и есть три подхода к этим предметам: либо интуитивный — через геометрию, либо вычислительный — через симплекс-метод, либо алгебраический — через теорию двойственности. Для линейного программирования эти подходы развиваются в данном параграфе и в § 8.2 и 8.3. Затем в § 8.4 рассматриваются задачи, имеющие целочисленное решение (например, задача о бракосочетании). В параграфе 8.5 обсуждаются покер и другие матричные игры, а в конце объясняется теорема о минимаксе и ее связь с теоремой двойственности из линейного программирования.

Ключом к этой главе служит понимание геометрического смысла линейных неравенств. Неравенство делит  $n$ -мерное пространство на два **полупространства**, в одном из которых неравенство выполняется, а в другом не выполняется. Типичным примером является неравенство  $x + 2y \geq 4$ . Границей, разделяющей эти два полупространства, является прямая  $x + 2y = 4$ , на которой неравенство обращается в равенство. Над этой прямой расположено полупространство, где  $x + 2y \geq 4$  (рис. 8.1); под ней находится противоположное полупространство, в котором неравенство неверно. В трехмерном случае рисунок будет почти таким же; граница превращается в плоскость, например  $x + 2y + z = 4$ , а над ней расположено полупространство  $x + 2y + z \geq 4$ . В  $n$ -мерном случае мы также будем называть  $(n - 1)$ -мерную границу «плоскостью».

В дополнение к неравенствам этого типа в линейном программировании существует еще одно основное ограничение: величины  $x$

и  $y$  должны быть неотрицательными. Разумеется, это ограничение само по себе дает пару линейных неравенств:  $x \geq 0$  и  $y \geq 0$ . Поэтому у нас есть еще два полупространства с границами, ле-

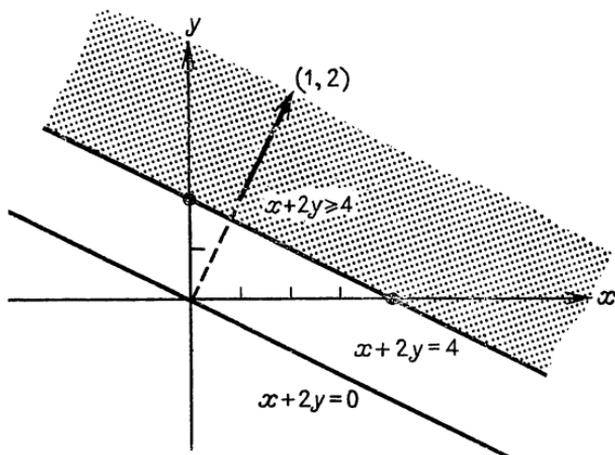


Рис. 8.1. Уравнения и неравенства.

жащими прямо на осях координат: неравенство  $x \geq 0$  допускает все точки, лежащие по правую сторону от прямой  $x=0$ , а  $y \geq 0$  дает полупространство над прямой  $y=0$ .

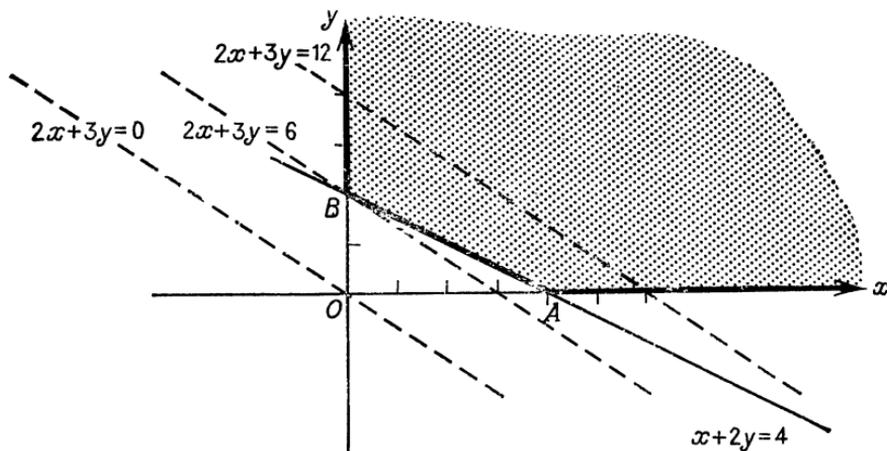


Рис. 8.2. Допустимое множество и функция стоимости  $2x+3y$

Важный шаг состоит в совместном рассмотрении всех трех неравенств  $x+2y \geq 4$ ,  $x \geq 0$ ,  $y \geq 0$ . Геометрически их сочетание дает выделенную область на рис. 8.2. Эту область можно рассматривать как *пересечение* трех полупространств. Она уже не является полупространством, а представляет собой типичный пример того, что в линейном программировании называется *допу-*

**стимым множеством.** Другими словами, допустимое множество образуется из решений набора линейных неравенств.

Нетрудно заметить, что система вида  $Ax = b$ , состоящая из  $m$  уравнений с  $n$  неизвестными, на самом деле описывает пересечение  $m$  различных плоскостей — по одной на каждое из уравнений. (Когда  $m = n$  и плоскости являются независимыми, они пересекаются в одной точке, являющейся решением  $x = A^{-1}b$ .) Аналогичным образом система  $m$  неравенств  $Ax \geq b$  описывает пересечение  $m$  полупространств. Если, кроме того, потребовать, чтобы каждая компонента вектора  $x$  была неотрицательна (что записывается в виде векторного неравенства  $x \geq 0$ ), то в результате возникает еще  $n$  полупространств. Чем больше ограничений мы накладываем, тем меньше будет допустимое множество.

Может случиться, что допустимое множество окажется ограниченным или даже пустым. Если мы перейдем к примеру с полупространством  $x + 2y \leq 4$ , сохранив неравенства  $x \geq 0$ ,  $y \geq 0$ , то в результате получим треугольник  $OAB$ . При соединении вместе неравенств  $x + 2y \geq 4$  и  $x + 2y \leq 4$  допустимое множество вырождается в прямую: два противоположных ограничения дают равенство  $x + 2y = 4$ . Если же мы добавим противоречащее ограничение, например  $x + 2y \leq -2$ , то допустимое множество будет пустым<sup>1)</sup>.

Алгебра линейных неравенств (или допустимых множеств) является составной частью нашего предмета. В линейном программировании, однако, имеется другая существенная составляющая: нас интересует не множество всех допустимых точек, а определенная точка, максимизирующая или минимизирующая некоторую «функцию стоимости». К примеру  $x + 2y \geq 4$ ,  $x \geq 0$ ,  $y \geq 0$  мы добавляем функцию стоимости (или целевую функцию)  $2x + 3y$ . Тогда настоящая задача линейного программирования состоит в отыскании точки  $x$ ,  $y$ , лежащей в допустимом множестве и минимизирующей стоимость.

Геометрическая интерпретация задачи приведена на рис. 8.2. Семейство стоимостей  $2x + 3y$  дает семейство параллельных прямых, и мы должны найти минимальную стоимость, т. е. первую прямую, которая пересекает допустимое множество. Очевидно, что это пересечение происходит в точке  $B$ , где  $x^* = 0$  и  $y^* = 2$ ; минимальная стоимость есть  $2x^* + 3y^* = 6$ . Вектор  $(0, 2)$  называется *допустимым*, поскольку он лежит в допустимом множестве, и он *оптимален*, поскольку он минимизирует функцию стоимости, и минимальная стоимость 6 называется *значением программы*. Мы будем помечать оптимальные векторы звездочкой.

<sup>1)</sup> Полупространство  $x + 2y \leq -2$  не пересекается с первым квадрантом  $x \geq 0$ ,  $y \geq 0$ .

Нетрудно видеть, что оптимальный вектор находится в углу допустимого множества. Это гарантируется геометрией, поскольку прямые, задающие функцию стоимости (или плоскости, когда мы перейдем к большему числу неизвестных), непрерывно движутся

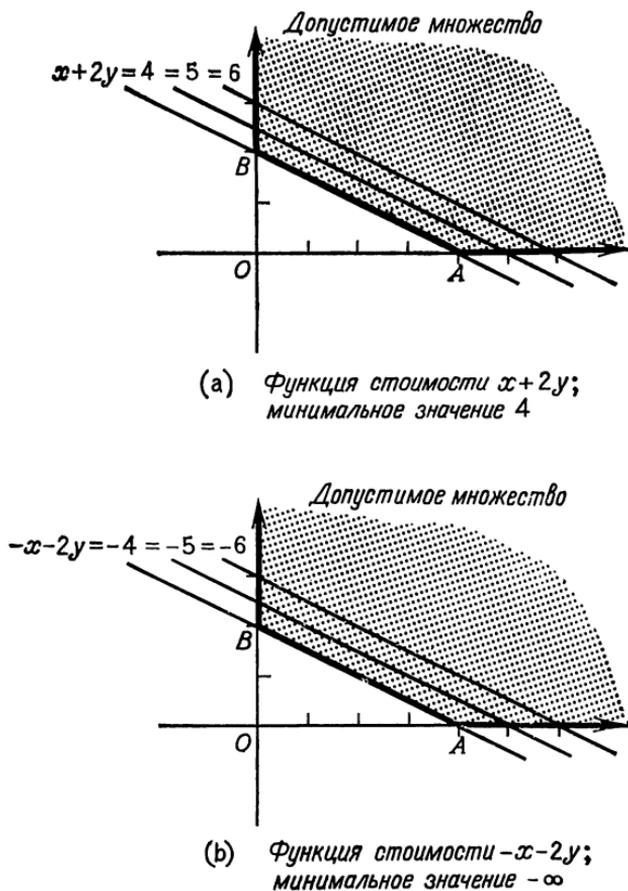


Рис. 8.3. Частные случаи: бесконечное число оптимальных векторов и ни одного оптимального вектора.

вверх, пока не пересекут допустимое множество. Первый контакт должен произойти на его границе. Однако при другой функции стоимости пересечение может и не быть одной точкой: если стоимость будет выражаться функцией  $x+2y$ , то все ребро между  $B$  и  $A$  попадет в пересечение одновременно, и мы получим бесконечное множество оптимальных векторов на этом отрезке (рис. 8.3а). Значение по-прежнему остается единственным ( $x^*+2y^*$  равняется 4 для всех оптимальных векторов), и потому задача на минимум по-прежнему имеет определенное решение. С другой

стороны, задача на максимум вообще не будет иметь решения! При нашем допустимом множестве стоимость может подниматься сколь угодно высоко и максимальная стоимость бесконечна. Для рассмотрения этой возможности в рамках задач на минимум следует изменить стоимость на  $-x-2y$ . Тогда минимум равняется  $-\infty$ , как на рис. 8.3b, и вновь решения нет. Каждая задача линейного программирования относится к одной из трех возможных категорий: либо ее допустимое множество является пустым, либо ее функция стоимости на допустимом множестве неограниченна, либо имеется единственное значение для линейной программы (с одним оптимальным вектором или с бесконечным множеством оптимальных векторов). Первые две ситуации должны быть весьма нетипичными для реальной экономической или инженерной задачи.

Существует простой способ превращения ограничений-неравенств в равенства, заключающийся во введении *переменной невязки*  $w = x + 2y - 4$ . Это и есть наше уравнение! Исходное ограничение  $x + 2y \geq 4$  превращается в  $w \geq 0$ , что прекрасно согласуется с остальными ограничениями  $x \geq 0$ ,  $y \geq 0$ . Именно с этого шага начинается симплекс-метод: введение переменных невязки для каждого из неравенств дает в результате только уравнения и простые ограничения на неотрицательность.

#### ИНТЕРПРЕТАЦИЯ ИСХОДНОЙ ЗАДАЧИ И ДВОЙСТВЕННОЙ К НЕЙ

Пожалуй, хватит о необычных случаях. Мы хотим вернуться к нашему исходному примеру со стоимостью  $2x + 3y$  и описать его словесно. Он является иллюстрацией «задачи о диете» в линейном программировании с двумя источниками протеина — например, бифштексом и арахисовым маслом. Каждый фунт арахисового масла дает единицу протеина, каждый фунт бифштекса дает две единицы, и диета должна содержать по меньшей мере четыре единицы. Поэтому диета, содержащая  $x$  фунтов арахисового масла и  $y$  фунтов бифштекса, ограничивается неравенствами  $x + 2y \geq 4$ ,  $x \geq 0$ ,  $y \geq 0$ . (Мы не можем иметь отрицательного количества бифштекса или арахисового масла). Это и есть допустимое множество, и задача состоит в минимизации стоимости. Если фунт арахисового масла стоит 2 доллара, а фунт бифштекса 3 доллара, то стоимость всей диеты есть  $2x + 3y$ . К счастью, оптимальная диета состоит из бифштекса:  $x^* = 0$  и  $y^* = 2$ .

Каждая линейная программа, в том числе и эта, имеет *двойственную*. Если исходная задача состоит в минимизации, то двойственная к ней — в максимизации и решение одной прямо приводит к решению другой; фактически минимум в данной «основной задаче» должен быть равен максимуму в двойственной к ней. На

самом деле это центральный результат в теории линейного программирования, и он будет разъяснен в § 8.3. Здесь же мы будем придерживаться задачи о диете и попытаемся дать интерпретацию двойственной к ней.

Для покупателя, выбирающего между бифштексом и арахисовым маслом из условия минимальной стоимости получаемого протеина, двойственной задачей является та, которая стоит перед продавцом, продающим *синтетический протеин*. Он намерен конкурировать с бифштексом и арахисовым маслом, и мы немедленно встречаемся с двумя ингредиентами типичной задачи линейного программирования: он хочет максимизировать цену  $p$ , но эта цена подчиняется линейным ограничениям. Прежде всего синтетический протеин не должен стоить больше, чем протеин в арахисовом масле (2 доллара за единицу) или протеин в бифштексе (3 доллара за две единицы). В то же время цена должна быть неотрицательной, иначе продавец не будет продавать. Поскольку диетой требовались четыре единицы протеина, доход продавца будет равен  $4p$  и двойственная задача в точности такова: *Максимизировать  $4p$ , если  $p \leq 2$ ,  $2p \leq 3$  и  $p \geq 0$* . Это пример, в котором двойственную задачу решить легче, чем исходную; она имеет лишь одну неизвестную. Ясно, что фактически действует лишь одно ограничение  $2p \leq 3$  и максимальная цена синтетического протеина равняется 1.50 доллара. Поэтому максимальный доход равен  $4p = 6$  долларов. Это была минимальная стоимость в исходной задаче, и покупатель платит одно и то же как за натуральный, так и за синтетический протеин. В этом состоит смысл теоремы двойственности.

**Упражнение 8.1.1.** Изобразить допустимое множество с ограничениями  $x + 2y \geq 6$ ,  $2x + y \geq 6$ ,  $x \geq 0$ ,  $y \geq 0$ . Какие точки находятся в трех «углах» этого множества?

**Упражнение 8.1.2** (рекомендуемое). Чему равняется минимальное значение функции стоимости  $x + y$  для предыдущего допустимого множества? Нарисовать прямую  $x + y = \text{const}$ , которая первой касается допустимого множества. Как насчет функций стоимости  $3x + y$  и  $x - y$ ?

**Упражнение 8.1.3.** Показать, что допустимое множество, ограниченное неравенствами  $x \geq 0$ ,  $y \geq 0$ ,  $2x + 5y \leq 3$ ,  $-3x + 8y \leq -5$ , является пустым.

**Упражнение 8.1.4.** Показать, что следующая задача допустима, но не ограничена и не имеет поэтому оптимального решения: максимизировать  $x + y$  при  $x \geq 0$ ,  $y \geq 0$ ,  $-3x + 2y \leq -1$ ,  $x - y \leq 2$ .

**Упражнение 8.1.5.** Добавить одно ограничительное неравенство к  $x \geq 0$ ,  $y \geq 0$  так, чтобы допустимое множество содержало лишь одну точку.

**Упражнение 8.1.6.** Какую форму имеет допустимое множество  $x \geq 0$ ,  $y \geq 0$ ,  $z \geq 0$ ,  $x + y + z = 1$  и чему равен максимум выражения  $x + 2y + 3z$ ?

## ТИПИЧНЫЕ ПРИЛОЖЕНИЯ

В следующем параграфе мы сконцентрируем внимание на решении (и опустим формулировки) задач линейного программирования. Поэтому пора описать некоторые практические ситуации, в которых может возникнуть следующая математическая задача: *минимизировать или максимизировать линейную функцию стоимости, подчиняющуюся линейным ограничениям.*

1. *Выбор портфеля ценных бумаг.* Предположим, что имеются три типа ценных бумаг: федеральные стоимостью  $A$ , дающие 5% годовых, муниципальные стоимостью  $B$ , дающие 6% годовых, и акции урановой компании стоимостью  $C$  и дающие 9% годовых. Мы можем купить этих бумаг на суммы  $x$ ,  $y$  и  $z$  соответственно с тем, чтобы общая сумма не превосходила 100 тысяч долларов. Задача состоит в максимизации процентного дохода при условиях, что

- i) в уран можно вложить не более 20 тысяч долларов;
- ii) средняя стоимость выбранных бумаг должна быть не меньше  $B$ .

Задача: *максимизировать величину  $5x + 6y + 9z$  при условиях  $x + y + z \leq 100\,000$ ,  $z \leq 20\,000$ ,  $z \leq x$ ,  $x, y, z \geq 0$ .*

2. *Планирование производства.* Пусть корпорация «Дженерал Моторс» получает прибыль в размере 100 долларов на каждый «шевроле», 200 долларов на каждый «бьюик» и 400 долларов на каждый «кадиллак». Они проходят соответственно по 20, 17 и 14 миль на одном галлоне бензина, а Конгресс постановил, что в среднем производимый автомобиль должен проходить 18 миль на галлоне бензина. Завод может собрать «шевроле» за 1 минуту, «бьюик» — за 2 минуты, а «кадиллак» — за 3 минуты. Чему равняется максимальная прибыль за восьмичасовой рабочий день?

Задача: *максимизировать величину  $100x + 200y + 400z$  при условиях*

$$20x + 17y + 14z \geq 18(x + y + z), \quad x + 2y + 3z \leq 480, \quad x, y, z \geq 0.$$

**Упражнение 8.1.7.** Решить задачу о портфеле ценных бумаг.

**Упражнение 8.1.8.** В допустимом множестве для второго примера условие неотрицательности  $x, y, z \geq 0$  выделяет одну восьмую трехмерного пространства (положительный октант). Как он разрезается двумя плоскостями, соответствующими ограничениям, и какую форму имеет допустимое множество? Каким образом его углы показывают, что при этих двух ограничениях в оптимальном решении будет только две марки автомашин?

**Упражнение 8.1.9.** *Задача о перевозках.* Пусть Техас, Калифорния и Аляска производят по миллиону баррелей нефти. Город Чикаго потребляет 800 тысяч баррелей и расположен на расстоянии 1000, 2000 и 3000 миль от соответствующих месторождений. Оставшиеся 2200 тысяч баррелей транспор-

тируются в Новую Англию на расстояние 1500, 3000 и 3700 миль соответственно. Если стоимость перевозки барреля составляет одну единицу на милю, то какую линейную программу с пятью ограничениями-равенствами нужно решить, чтобы оптимизировать величины  $x$ ,  $y$ ,  $z$  для Чикого и  $u$ ,  $v$ ,  $w$  для Новой Англии?

## § 8.2. СИМПЛЕКС-МЕТОД

Этот параграф посвящен линейному программированию с  $n$  неизвестными и  $m$  ограничениями. В предыдущем параграфе мы имели  $n=2$  и  $m=1$ ; там были две неотрицательные переменные и одно ограничение  $x+2y \geq 4$ . Более общий случай нетрудно объяснить, но нелегко решить.

Наилучшим способом для описания задачи является ее матричная запись.  $n$  неизвестных образуют неотрицательный вектор-столбец  $x = [x_1 \ x_2 \ \dots \ x_n]^T$ . Чтобы быть «допустимым», этот вектор должен удовлетворять  $m$  ограничениям помимо  $x \geq 0$ ; мы записываем эти ограничения в виде  $Ax \geq b$ . Матрица  $A$  имеет размер  $m \times n$ , и каждая из ее строк дает одно неравенство (точно так же, как в остальной части книги каждая строка системы  $Ax = b$  давала одно уравнение). Матрица  $A$  и вектор  $b$  заданы — для ограничения  $x+2y \geq 4$  мы имели  $A = [1 \ 2]$  и  $b = [4]$ . Функция стоимости тоже задана, и она также является линейной; она равняется  $c_1x_1 + c_2x_2 + \dots + c_nx_n$ , или  $sx^1$ ). Задача состоит в отыскании допустимого вектора, минимизирующего стоимость; этот вектор и есть оптимальный.

**Задача на минимум:** минимизировать  $sx$  при условиях  $x \geq 0$ ,  $Ax \geq b$ .

Геометрическая интерпретация очевидна. Первое условие  $x \geq 0$  ограничивает вектор «положительным квадрантом» в  $n$ -мерном пространстве — это есть область, общая для всех полупространств  $x_j \geq 0$ . В двумерном случае это четверть плоскости, в трехмерном — восьмая часть пространства; в общем случае имеется один шанс из  $2^n$  для того, чтобы вектор был неотрицательным. Другие  $m$  ограничений дают  $m$  дополнительных полупространств, и допустимыми векторами являются те, которые удовлетворяют всем  $m+n$  условиям; они лежат в квадранте  $x \geq 0$  и одновременно удовлетворяют  $Ax \geq b$ . Другими словами, допустимое множество есть пересечение  $m+n$  полупространств; оно может быть ограниченным (с плоскими сторонами!), неограниченным или пустым.

Функция стоимости  $sx$  вводит в задачу полное семейство параллельных плоскостей. Один из элементов этого семейства, а именно содержащий начало координат, является плоскостью  $sx = 0$ ; если вектор  $x$  удовлетворяет этому уравнению, то он имеет

<sup>1</sup>)  $s$  — вектор-строка.

«нулевую стоимость». Остальные плоскости  $sx = \text{const}$  дают все возможные стоимости. При изменении стоимости эти плоскости перемещаются в  $n$ -мерном пространстве и оптимальным вектором  $x^*$  будет тот, при котором плоскость впервые коснется допустимого множества. Этот вектор  $x^*$  является допустимым, и соответствующая стоимость  $sx^*$  является минимальной внутри допустимого множества; таким образом,  $x^*$  решает стандартную задачу на минимум в линейном программировании.

Наша цель в этом параграфе заключается в вычислении  $x^*$ . В принципе мы можем сделать это путем отыскания всех углов допустимого множества и вычисления стоимостей в них; тогда угол с минимальной стоимостью будет оптимальным. Однако практическая реализация данного подхода невозможна, поскольку таких углов насчитываются миллионы и мы не можем производить вычисления во всех них. Вместо этого мы обратимся к *симплекс-методу*, который без сомнения можно считать наиболее плодотворной идеей современной вычислительной математики. Он был разработан Данцигом как систематический способ для решения задач линейного программирования и либо по счастливой случайности, либо благодаря выдающимся достоинствам имел потрясающий успех. Этапы этого метода приведены на стр. 372, но сначала мы попытаемся разъяснить их насколько возможно.

Мне кажется, что лучше всего сделать это геометрически. На первом шаге мы просто находим некоторый угол допустимого множества, что составляет первую фазу. Сердцем метода является его вторая фаза, заключающаяся в *движении от угла к углу вдоль ребер допустимого множества*. В типичном случае нам придется выбирать для движения одно из  $n$  ребер; некоторые из них будут вводить от оптимального (но неизвестного)  $x^*$ , другие будут постепенно приближать нас к этому вектору. Данциг предложил двигаться вдоль того ребра, которое гарантирует убывание стоимости. Это ребро приводит к новому углу с меньшей стоимостью, и уже нет возможности вернуться к чему-либо более дорогому. Постепенно мы достигнем особого угла, для которого уже все направления станут недопустимыми и стоимость минимизирована. Этот угол дает оптимальный вектор  $x^*$ , и процесс останавливается.

Реальная трудность состоит в переходе от этой идеи к ее алгебраической реализации. Сначала нужно интерпретировать слова *угол* и *ребро* в  $n$ -мерном пространстве. Угол является точкой пересечения  $n$  различных плоскостей, каждая из которых задается одним уравнением — в точности, как три плоскости (или передняя стена, боковая стена и пол) образуют трехмерный угол. Напомним, что допустимое множество в линейном программировании определяется  $m$  неравенствами  $Ax \geq b$  и  $n$  неравенствами  $x \geq 0$ . Каждый из углов множества образуется путем замены  $n$  из  $n + m$  неравенств на равенства и отыскания точки пересечения этих

плоскостей. В частности, одна возможность состоит в выборе  $n$  уравнений  $x_1=0, \dots, x_n=0$ , что даст начало координат. Как и для всех остальных случаев, эта точка пересечения будет истинным углом только в том случае, когда она удовлетворяет также и  $m$  остальным ограничениям. В противном случае она даже не лежит внутри допустимого множества и является ложной. Пример в упр. 8.1.1 содержал  $n=2$  переменных и  $m=2$  ограничений; существует шесть возможных пересечений, только три из которых на самом деле являются углами допустимого множества (см. рис. 8.4). Эти углы обозначены через  $P, Q, R$ ; они

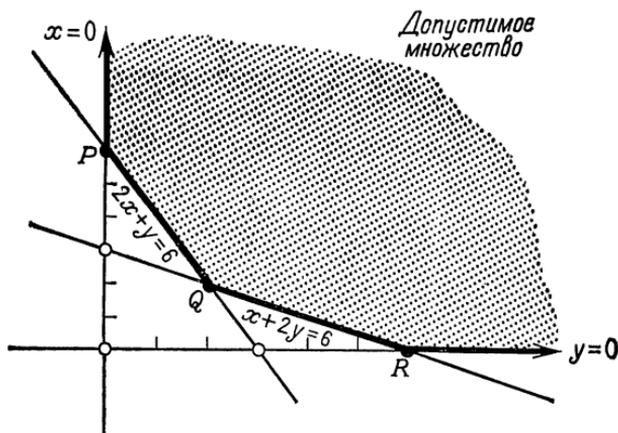


Рис. 8.4. Углы и ребра допустимого множества.

дают векторы  $(0, 6)$ ,  $(2, 2)$ ,  $(6, 0)$ , один из которых должен быть оптимальным (кроме случая, когда минимальная стоимость равняется  $-\infty$ ). Остальные три угла, включая начало координат, являются ложными.

В общем случае имеется  $(n+m)!/n!m!$  возможных пересечений, что равняется числу способов, которыми можно выбрать  $n$  из  $n+m$  плоскостей<sup>1)</sup>. Если допустимое множество является пустым, то ни одно из этих пересечений не будет настоящим углом допустимого множества. Отыскание настоящего угла или установление пустоты допустимого множества является задачей первой фазы; в дальнейшем предполагается, что некоторый угол уже найден.

Уберем одну из  $n$  пересекающихся плоскостей; в результате останется  $n-1$  уравнений и, следовательно, одна степень свободы. Точки, удовлетворяющие этим  $n-1$  уравнениям, образуют ребро, выходящее из угла. Геометрически это ребро является

<sup>1)</sup> Величина этого числа и объясняет, почему вычисление стоимости для всех углов является совершенно нереальной задачей при больших  $n$ .

пересечением  $n - 1$  плоскостей. Поскольку мы должны оставаться в допустимом множестве, то мы не можем выбирать направление движения вдоль ребра: только одно направление остается допустимым. Но мы можем выбирать любое из  $n$  различных ребер, и этот выбор осуществляется фазой 2.

Для описания этой фазы мы должны переписать систему  $Ax \geq b$  в форме, совершенной аналогичной  $n$  простым ограничениям  $x_1 \geq 0, \dots, x_n \geq 0$ . Для этого служат переменные невязки  $w = Ax - b$ . С их помощью  $m$  ограничений  $Ax \geq b$  переходят в  $w_1 \geq 0, \dots, w_m \geq 0$  с одной переменной невязки для каждой строки матрицы  $A$ . Тогда уравнение  $w = Ax - b$ , или  $Ax - w = b$ , в матричной форме записывается как

$$[A \quad -I] \begin{bmatrix} x \\ w \end{bmatrix} = b.$$

Допустимое множество определяется этими  $m$  уравнениями и  $n + m$  простыми неравенствами  $x \geq 0, w \geq 0$ . Другими словами, обычная задача привелась к каноническому виду с ограничениями в виде равенств.

Для завершения этого перехода остается устранить различие между исходным вектором  $x$  и новым вектором  $w$ , поскольку они не различаются в симплекс-методе и сохранять обозначения

$$[A \quad -I] \quad \text{и} \quad \begin{bmatrix} x \\ w \end{bmatrix}$$

не имеет смысла. Поэтому мы изменим обозначения и будем обозначать через  $A$  расширенную матрицу, а через  $x$  расширенный вектор. Тогда ограничения-равенства запишутся в виде  $Ax = b$  и  $n + m$  простых неравенств перейдут в  $x \geq 0$ <sup>1)</sup>. Начальный вектор стоимости  $c$  должен быть расширен путем добавления  $m$  нулевых компонент; тогда стоимость  $cx$  при новой интерпретации обозначений  $x$  и  $c$  совпадает с исходной. Единственное отличие, вызванное введением переменных невязки, состоит в том, что матрица  $A$  будет иметь размер  $m \times (n + m)$ , а вектор  $x$  будет иметь  $n + m$  компонент. Мы сохраняем старые обозначения  $m$  и  $n$  в качестве напоминания о произведенной замене.

**Пример.** В задаче, проиллюстрированной на рис. 8.4, с ограничениями  $x + 2y \geq 6, 2x + y \geq 6$  и функцией стоимости  $x + y$  новая система имеет вид

$$A = \begin{bmatrix} 1 & 2 & -1 & 0 \\ 2 & 1 & 0 & -1 \end{bmatrix}, \quad b = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \quad \text{и} \quad c = [1 \quad 1 \quad 0 \quad 0].$$

<sup>1)</sup> Разумеется, в экономических или физических задачах ограничения могут иметь вид равенств с самого начала. В этом случае переменные невязки не нужны и симплекс-метод будет сразу работать с системой  $Ax = b$ .

После этого перехода к ограничениям-равенствам можно применять симплекс-метод. Вспомним, что фазой 1 уже найден угол допустимого множества, т. е. точка пересечения  $n$  плоскостей:  $n$  исходных неравенств  $x \geq 0$  и  $Ax \geq b$  (или  $w \geq 0$ ) преобразованы в уравнения. Другими словами, *угол есть точка, в которой  $n$  компонент нового вектора  $x$  (или исходного вектора  $\begin{bmatrix} x \\ w \end{bmatrix}$ ) равны нулю.* В системе  $Ax = b$  мы будем считать эти  $n$  компонент вектора  $x$  свободными переменными, а остальные  $m$  компонент — базисными переменными. Тогда, полагая  $n$  свободных переменных равными нулю, из  $m$  уравнений системы  $Ax = b$  находим  $m$  базисных переменных. Это решение  $x$  называется *базисным* в соответствии с тем, что оно полностью зависит от базисных переменных. Если, кроме того, ненулевые компоненты вектора  $x$  неотрицательны, то соответствующий угол будет истинным, т. е. принадлежащим допустимому множеству.

**8А.** Углы допустимого множества являются **базисными допустимыми решениями** системы  $Ax = b$ : решение  $x$  является базисным, когда  $n$  из  $m+n$  его компонент равны нулю, и оно допустимое, если  $x \geq 0$ . Первая фаза симплекс-метода дает одно базисное допустимое решение, а вторая фаза служит для последовательного приближения к оптимальному решению.

**Пример.** Угловая точка  $P$  на рис. 8.4 дается пересечением прямых  $x = 0$  и  $2x + y - 6 = 0$ , так что две компоненты вектора  $x$  равны нулю, а две другие положительны. Следовательно, он является базисным и допустимым:

$$Ax = \begin{bmatrix} 1 & 2 & -1 & 0 \\ 2 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 6 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} = b.$$

Остается принять главное решение: в какой угол следует теперь переходить? Мы нашли один угол, в котором  $m$  компонент вектора  $x$  являются положительными, а остальные — нулями, и хотим двигаться вдоль ребра допустимого множества в соседний угол. Тот факт, что эти два угла являются соседними, означает, что из  $m$  базисных переменных  $m-1$  останутся базисными и лишь одна станет свободной (нулем). Одновременно одна из свободных переменных станет базисной; ее значение из нуля станет положительным, а остальные  $m-1$  основных переменных изменятся, оставаясь положительными. Задача состоит в том, чтобы решить, какую переменную следует исключить из базисных и

какую добавить. Если мы знаем, какие  $m$  переменных являются базисными в новом углу, их значения вычисляются путем решения системы  $m$  уравнений  $Ax = b$ , где оставшиеся компоненты вектора  $x$  полагаются равными нулю.

Поясним сказанное на примере:

*Минимизировать  $7x_3 - x_4 - 3x_5$  при условиях*

$$\begin{aligned} x_1 + x_3 + 6x_4 + 2x_5 &= 8, \\ x_2 + x_3 + 3x_5 &= 9. \end{aligned}$$

Начинаем с угла, где  $x_1 = 8$  и  $x_2 = 9$  (они являются базисными переменными) и  $x_3 = x_4 = x_5 = 0$ . Ограничения удовлетворяются, но стоимость может не быть минимальной. Фактически именно функция стоимости определяет, какую свободную переменную нужно ввести в базис. Глупо делать положительной переменную  $x_3$ , поскольку коэффициент при ней равен  $+7$ , а мы пытаемся уменьшить стоимость. Целесообразно выбирать переменную  $x_4$  или  $x_5$ , и мы выберем  $x_5$ , поскольку коэффициент  $-3$  при ней является более отрицательным. Если бы все коэффициенты были положительны, то исходное решение уже являлось бы оптимальным.

Если переменная  $x_5$  включается в базис, то либо  $x_1$ , либо  $x_2$  следует исключить из него. Если  $x_5$  увеличивается в первом из уравнений, а  $x_1$  уменьшается, с тем чтобы сохранить равенство  $x_1 + 2x_5 = 8$ , то  $x_1$  окажется равной нулю при  $x_5 = 4$ . К этому моменту, однако, величина  $x_2$  во втором уравнении уже станет отрицательной:  $x_5 = 4$  влечет за собой равенство  $x_2 = -3$ . Правило состоит в том, чтобы увеличивать  $x_5$  до тех пор, пока одна из базисных переменных не станет равной нулю, после чего она становится свободной. В нашем случае  $x_2$  равняется нулю при  $x_5 = 3$  и из первого уравнения получаем, что  $x_1 = 2$ . Стоимость при этом падает до  $-9$ . В общем случае мы вычисляем отношения правых частей к коэффициентам при включаемой переменной,  $8/2$  и  $9/3$ , и наименьшее отношение показывает, какая из переменных станет свободной.

Это правило учитывает лишь положительные отношения, поскольку если коэффициент при  $x_5$  будет равен  $-3$  вместо  $+3$ , то увеличение  $x_5$  будет одновременно *увеличивать*  $x_2$  (при  $x_5 = 10$  мы получим  $x_2 = +39$  во втором уравнении) и мы никогда не получим нулевого  $x_2$ . Если все коэффициенты при  $x_5$  являются отрицательными, то мы имеем *неограниченный* случай:  $x_5$  можно неограниченно увеличивать и при этом стоимость будет убывать до  $-\infty$ , причем мы не покинем допустимое множество. Ребро является бесконечно длинным.

В нашем случае шаг заканчивается в углу  $x_1 = 2$ ,  $x_2 = x_3 = x_4 = 0$ ,  $x_5 = 3$ . Однако следующий шаг будет простым только

в том случае, когда уравнения имеют подходящий вид — когда базисные переменные выделены, как это было в исходной ситуации с  $x_1$  и  $x_2$ . Для этого мы осуществляем «выделение ведущего элемента», решая второе уравнение относительно  $x_5$  и подставляя результат в функцию стоимости и в первое уравнение:  $3x_5 = 9 - x_2 - x_3$ , так что новая задача имеет вид:

минимизировать  $7x_3 - x_4 - (9 - x_2 - x_3) = x_2 + 8x_3 - x_4 - 9$  при условиях

$$\begin{aligned} x_1 - \frac{2}{3}x_2 + \frac{1}{3}x_3 + 6x_4 &= 2, \\ \frac{1}{3}x_2 + \frac{1}{3}x_3 &+ x_5 = 3. \end{aligned}$$

Следующий шаг осуществляется просто. Единственный отрицательный коэффициент в функции стоимости означает, что должна включаться переменная  $x_4$ . Отношения  $2/6$  и  $3/0$  означают, что исключаться должна переменная  $x_1$ . Новый угол задается соотношениями  $x_1 = x_2 = x_3 = 0$ ,  $x_4 = 1/3$ ,  $x_5 = 3$ , и новая стоимость равна  $-9 \frac{1}{3}$ . Она и является оптимальной.

В задачах большей размерности исключаемая переменная впоследствии может вновь попасть в базис. Однако поскольку стоимость продолжает снижаться (исключая вырожденный случай), ситуация, когда все  $m$  базисных переменных вернуться, невозможна. Ни один угол не посещается дважды, и процесс должен остановиться в оптимальном углу (или на  $-\infty$ , если стоимость является неограниченной). Что самое замечательное, так это быстрота, с которой обычно находится оптимальный угол.

**Замечание о вырожденности.** Угол называется *вырожденным*, если в нем равны нулю не  $n$  компонент вектора  $x$ , а больше, т. е. через этот угол проходит больше чем  $n$  плоскостей и некоторые из базисных переменных обращаются в нуль. В симплекс-методе признаком этого служит появление нуля в правой части уравнений. В результате этого отношения, определяющие исключаемую переменную, будут содержать нули и базис может изменяться без выхода из данного угла; на самом деле мы могли бы до бесконечности кружиться на месте, выбирая базис.

Такую ситуацию можно устранить либо путем соответствующей перестройки симплекс-метода, либо предполагая, что заикливания не происходит. К счастью, последнее согласуется с вычислительной практикой.

**Упражнение 8.2.1.** Минимизировать  $x_1 + x_2 - x_3$  при условиях

$$\begin{aligned} 2x_1 - 4x_2 + x_3 + x_4 &= 4, \\ 3x_1 + 5x_2 + x_3 &+ x_5 = 2. \end{aligned}$$

Какая из переменных  $x_1, x_2, x_3$  должна быть включена в базис и какая из  $x_4, x_5$  должна его покинуть? Вычислить новую пару базисных переменных и найти стоимость в новом углу.

**Упражнение 8.2.2.** Сделав первый шаг, подготовить и осуществить следующий.

### ТАБЛИЦА

Каждый шаг симплекс-метода состоит из принятия решений и операций над строками матрицы: нужно выбрать входящую и исключаемую переменные и осуществить это включение и исключение. Один из способов организации этого состоит во включении данных в большую матрицу, называемую *таблицей*. Чтобы объяснить все операции, мы сначала осуществим их в терминах матриц, присутствующих в формуле для стоимости (и для критерия остановки процесса, которому удовлетворяет лишь оптимальный угол). Нам задана матрица  $A$ , правая часть  $b$  и вектор стоимости  $c$ , так что исходная таблица имеет вид

$$T = \left[ \begin{array}{c|c} A & b \\ \hline c & 0 \end{array} \right].$$

Эта матрица имеет размер  $(m+1) \times (m+n+1)$ . В таком ее виде базисные переменные легко спутать с другими и поэтому первый шаг заключается в том, чтобы оставить по одной базисной переменной в каждой строке, после чего станет легко принимать решения. Предположим, что  $x_1, \dots, x_m$  являются базисными переменными (при необходимости можно произвести перенумерацию), а остальные  $n$  переменных являются свободными. Тогда первые  $m$  столбцов матрицы  $A$  образуют квадратную матрицу  $B$  (базисную матрицу для этого угла) и оставшиеся  $n$  столбцов дают матрицу  $F$  размера  $m \times n$ . Аналогичным образом вектор стоимости  $c$  разбивается на  $[c_B \ c_F]$ , а вектор  $x$  — на  $[x_B \ x_F]^T$ . Основным моментом состоит в том, что в этом углу  $x_F = 0$  и уравнение  $Ax = b$  превращается в  $Bx_B = b$ , откуда находятся базисные переменные  $x_B$ . Таким образом, стоимость равняется  $cx = c_B x_B$ . Для работы с таблицей мы мысленно разбиваем ее на клетки:

$$T = \left[ \begin{array}{c|c|c} B & F & b \\ \hline c_B & c_F & 0 \end{array} \right].$$

Базисные переменные будут разделены, если мы сможем заменить матрицу  $B$  единичной матрицей. Это достигается при помощи преобразований строк, и в результате получаем

$$T' = \left[ \begin{array}{c|c|c} I & B^{-1}F & B^{-1}b \\ \hline c_B & c_F & 0 \end{array} \right].$$

Естественно, матрица, преобразующая  $B$  в  $I$ , есть  $B^{-1}$ , хотя она нигде не присутствует явно. Поэтому все вычитания одной строки из другой в совокупности эквивалентны умножению на  $B^{-1}$ . (Фактически мы производим шаг метода Гаусса — Жордана, аналогичный описанному на стр. 45; ведущие элементы устраняются путем деления на них, так что на диагонали остаются единицы, а над ними и под ними получаются нули.) Для завершения мы используем эти единичные ведущие элементы, чтобы получить нули также и в нижней строке, где теперь расположен вектор  $c_B$ . Для этого первая строка умножается на  $c_1$  и вычитается из последней, затем вторая строка умножается на  $c_2$  и т. д., т. е.  $c_B$  умножается на верхнюю часть матрицы и вычитается из нижней:

$$T'' = \left[ \begin{array}{c|c|c} I & B^{-1}F & B^{-1}b \\ \hline 0 & c_F - c_B B^{-1}F & -c_B B^{-1}b \end{array} \right].$$

На этом построение таблицы заканчивается и остается лишь правильно ее интерпретировать. Наша задача заключалась в минимизации  $cx$  при  $x \geq 0$ ,  $Ax = b$ . Уравнение  $Ax = b$  умножалось на  $B^{-1}$ , давая в результате

$$x_B + B^{-1}F x_F = B^{-1}b, \quad (1)$$

и стоимость  $cx = c_B x_B + c_F x_F$  превращалась в

$$cx = (c_F - c_B B^{-1}F) x_F + c_B B^{-1}b. \quad (2)$$

Важно то, что таблица содержит все нужные величины. Справа расположены базисные переменные  $x_B = B^{-1}b$  (свободные переменные суть  $x_F = 0$ ). Текущая стоимость равняется  $cx = c_B B^{-1}b$  и находится внизу справа со знаком минус. И, что самое важное, мы можем определить, является ли данный угол оптимальным, взглянув на функцию стоимости, расположенную в середине нижней строки. Если какой-либо из ее элементов является отрицательным, то стоимость может быть уменьшена. Если все коэффициенты положительны или равны нулю, то любое изменение может только увеличить стоимость и наилучший угол найден. Это и есть **критерий останова**, или **условие оптимальности**:

**8В.** Если вектор  $r = c_F - c_B B^{-1}F$  является неотрицательным, то уменьшить стоимость невозможно: угол уже является оптимальным и минимальная стоимость равняется  $c_B B^{-1}b$ . Следовательно, условие  $r \geq 0$  является критерием останова симплекс-метода. Если это условие не выполняется, то каждая отрицательная компонента вектора  $r$  соответствует ребру, вдоль которого стоимость может быть уменьшена и обычной стратегией является выбор наибольшей по модулю отрицательной компоненты вектора  $r$ .

Если по этому правилу была выбрана  $i$ -я компонента вектора  $r$ , то исходный угол не является оптимальным и  $i$ -я компонента (скажем,  $\alpha$ ) вектора  $x_F$  станет положительной. Остается вопрос, насколько большой становится эта компонента? Мы решили, какая компонента из свободной станет базисной, а теперь мы должны решить, какая из базисных компонент должна стать свободной. При увеличении  $\alpha$  некоторые компоненты вектора  $x_B$  могут начать уменьшаться, и мы попадаем в следующий угол, когда какая-либо компонента вектора  $x_B$  окажется равной нулю. Именно эта компонента (скажем,  $k$ -я) становится свободной переменной. В этой точке мы получим новый вектор  $x$ , который одновременно является допустимым и базисным: он является допустимым, так как по-прежнему  $x \geq 0$ , и он базисный, поскольку мы вновь имеем  $n$  нулевых компонент. При этом  $i$ -я компонента вектора  $x_F$ , которая изменилась от нуля до  $\alpha$ , заменяет  $k$ -ю компоненту вектора  $x_B$ , которая обратилась в нуль. Остальные компоненты вектора  $x_B$  также изменятся, но останутся положительными.

8С. Пусть  $u$  есть  $i$ -й столбец матрицы  $F$ , который был выбран по 8В, чтобы стать из свободного базисным. Тогда соответствующая компонента в новом углу  $x$  равняется

$$\alpha = \min \frac{(B^{-1}b)_j}{(B^{-1}u)_j} = \frac{(B^{-1}b)_k}{(B^{-1}u)_k}. \quad (3)$$

Минимум берется только по положительным компонентам вектора  $B^{-1}u$ . Если положительные компоненты отсутствуют, то следующий угол расположен бесконечно далеко и стоимость можно уменьшать безгранично; минимальная стоимость равняется  $-\infty$ . В противном случае старый  $k$ -й столбец матрицы  $B$  заменяется новым столбцом и симплекс-метод начинает работать в новом углу.

Столбцы  $B^{-1}b$  и  $B^{-1}u$  присутствуют в окончательной таблице ( $B^{-1}u$  является столбцом матрицы  $B^{-1}F$  и расположен над самым отрицательным элементом в нижней строке  $r$ ), так что отношения сразу вычисляются.

**Пример.** При функции стоимости вида  $x + y$  и ограничениях  $x + 2y - u = 6$ ,  $2x + y - v = 6$ , которые мы рассматривали выше, первая таблица имеет вид

$$\begin{bmatrix} 1 & 2 & -1 & 0 & 6 \\ 2 & 1 & 0 & -1 & 6 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Если мы начинаем с того же угла  $P$  на рис. 8.4, в котором прямая  $x=0$  пересекается с  $2x+y=6$  (что эквивалентно равенству  $v=0$ ), то базисными будут оставшиеся две переменные  $y$  и  $u$ . Для единообразия переставим местами первый и третий столбцы, после чего вектор  $u$  будет расположен до  $x$ :

$$T = \left[ \begin{array}{cc|cc|c} -1 & 2 & 1 & 0 & 6 \\ 0 & 1 & 2 & -1 & 6 \\ \hline 0 & 1 & 1 & 0 & 0 \end{array} \right].$$

Для процесса исключения умножим первую строку на  $-1$ , что даст единичный ведущий элемент, и после этого при помощи второй строки получим нули во втором столбце:

$$T'' = \left[ \begin{array}{cc|cc|c} 1 & 0 & 3 & -2 & 6 \\ 0 & 1 & 2 & -1 & 6 \\ \hline 0 & 0 & -1 & 1 & -6 \end{array} \right].$$

Первым делом посмотрим на  $r = [-1 \ 1]$  в нижней строке. Он имеет отрицательную компоненту, так что текущий угол  $u = y = 6$  и текущая стоимость  $+6$  не являются оптимальными. Отрицательная компонента находится в третьем столбце, так что третья переменная  $x$  будет включаться в базис. Столбец над этой отрицательной компонентой есть  $B^{-1}u = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ , и его отношения с последним столбцом равны  $6/3$  и  $6/2$ . Поскольку первое отношение меньше, первая неизвестная  $u$  (и первый столбец таблицы) будет исключаться из базиса.

В новой таблице мы меняем местами первый и третий столбец, и выделение ведущих элементов по методу исключения в этой таблице дает

$$\left[ \begin{array}{cc|cc|c} 3 & 0 & 1 & -2 & 6 \\ 2 & 1 & 0 & -1 & 6 \\ \hline -1 & 0 & 0 & 1 & -6 \end{array} \right] \longrightarrow \left[ \begin{array}{cc|cc|c} 1 & 0 & \frac{1}{3} & -\frac{2}{3} & 2 \\ 0 & 1 & -\frac{2}{3} & \frac{1}{3} & 2 \\ \hline 0 & 0 & \frac{1}{3} & \frac{1}{3} & -4 \end{array} \right].$$

Теперь  $r = [1/3 \ 1/3]$  является положительным и критерий остановки проходит. Угол  $x = y = 2$  и соответствующая стоимость  $+4$  будут оптимальными.

**Упражнение 8.2.3.** Пусть исходная функция стоимости имеет вид  $3x + y$ , так что после перестановки  $c = (0, 1, 3, 0)$  в углу  $P$ . Показать, что  $r \geq 0$  и, следовательно, что этот угол является оптимальным.

**Упражнение 8.2.4.** Пусть исходная функция стоимости имеет вид  $x-y$ , так что после перестановки  $c=(0, -1, 1, 0)$  в углу  $P$ . Вычислить  $r$  и определить, какой столбец  $u$  должен быть включен в базис. После этого вычислить  $B^{-1}u$  и показать (используя его знак), что другой угол никогда не будет достигнут. Мы поднимаемся вдоль оси  $y$  (см. рис. 8.4) и величина  $x-y$  убывает к  $-\infty$ .

**Упражнение 8.2.5.** В том же самом примере изменить стоимость на  $x+3y$ . Проверить, что симплекс-метод будет переводить нас от  $P$  к  $Q$  и к  $R$  и что угол  $R$  является оптимальным.

**Упражнение 8.2.6.** Первая фаза состоит в отыскании базисного допустимого решения системы  $Ax=b$ . После изменения знаков, чтобы сделать  $b \geq 0$ , рассмотреть дополнительную задачу минимизации  $\omega_1+\omega_2+\dots+\omega_m$  при условиях  $x \geq 0, \omega \geq 0, Ax+\omega=b$ . Если система  $Ax=b$  имеет неотрицательное решение, то минимальная стоимость в этой задаче будет равняться нулю —  $\omega^*=0$ .

а) Показать, что для новой задачи угол  $x=0, \omega=b$  является одновременно базисным и допустимым. Поэтому для нее первая фаза уже выполнена и симплекс-метод может начинать работу по отысканию оптимальной пары  $x^*, \omega^*$ . Если  $\omega^*=0$ , то  $x^*$  есть требуемый угол исходной задачи.

б) для  $A=[1 \ -1]$  и  $b=[3]$  выписать дополнительную задачу, вектор ее первой фазы из предыдущего пункта (а) и оптимальный вектор. Найти угол допустимого множества  $x_1-x_2=3, x_1 \geq 0, x_2 \geq 0$  и нарисовать это множество.

**Упражнение 8.2.7.** Если мы хотим максимизировать, а не минимизировать стоимость (по-прежнему для  $Ax=b$  и  $x \geq 0$ ), то каким будет критерий останова для  $r$ ? По каким правилам должны выбираться столбцы матрицы  $F$ , которые следует сделать базисными, и столбцы матрицы  $B$ , которые должны становиться свободными?

**Упражнение 8.2.8.** Минимизировать  $2x_1+x_2$  при условиях  $x_1+x_2 \geq 4, x_1+3x_2 \geq 12, x_1-x_2 \geq 0, x_1 \geq 0, x_2 \geq 0$ .

### ОРГАНИЗАЦИЯ ШАГА СИМПЛЕКС-МЕТОДА

Мы осуществили переход от геометрии симплекс-метода к его алгебре, т. е. от языка «углов» к языку «базисных допустимых решений». Теперь мы знаем, что основную роль играют вектор  $r$  и отношение  $\alpha$ , и хотим вновь вернуться к их вычислению. Это является основой симплекс-метода и может быть организовано тремя способами:

- (1) составлением таблицы;
- (2) вычислением  $B^{-1}$  всякий раз, когда  $k$ -й столбец матрицы  $B$  заменяется  $i$ -м столбцом  $u$  из  $F$ ;
- (3) вычислением  $B=LU$  и пересчетом этих множителей вместо  $B^{-1}$ .

Этот список представляет собой краткую историю симплекс-метода. В некотором смысле наиболее удивительным его этапом был первый — *таблица*, — который был доминирующим в течение многих лет. Для большинства из нас он привносил в линейное

программирование ощущение тайны главным образом потому, что он ухитрялся почти полностью обходиться без матричных обозначений (поскольку все матрицы выписывались целиком!). Однако с вычислительной точки зрения пора таблиц кончилась (кроме простых задач в учебниках).

Дело в том, что, после того как вектор  $r$  вычислен и его максимальная по модулю отрицательная компонента показала, какой столбец будет включен в базис, ни один из оставшихся столбцов, расположенных над  $r$ , не будет использоваться. Их вычисление является пустой тратой времени. В большой задаче сотни столбцов будут вычисляться снова и снова в ожидании своей очереди для включения в базис. Столь тщательное проведение всех исключений дает теоретическую строгость, но является неоправданным для практики.

Пора заканчивать описание табличного метода. Значительно быстрее и значительно проще оказывается рассмотрение симплекс-метода и выяснение того, какие вычисления действительно являются необходимыми. На каждом шаге мы должны переставлять столбец матрицы  $F$  со столбцом матрицы  $B$  и решать (в соответствии с  $r$  и  $\alpha$ ), какие именно столбцы должны меняться местами. Этот шаг состоит в выполнении следующего набора операций, начинающегося с данной базисной матрицы  $B$  и решения  $x_B = B^{-1}b$ :

### Таблица 8.1. Шаг симплекс-метода

- 1) Вычисляем вектор-строку  $\lambda = c_B B^{-1}$  и вектор  $r = c_F - \lambda F$ .
- 2) Если  $r \geq 0$ , то процесс останавливается; текущее решение является оптимальным. В противном случае если  $r_i$  есть наибольшая по модулю отрицательная компонента, то  $i$ -й столбец матрицы  $F$  будет включаться в базис. Обозначим его через  $u$ .
- 3) Вычисляем  $v = B^{-1}u$ .
- 4) Вычисляем отношения  $B^{-1}b$  к  $B^{-1}u$  для положительных компонент  $B^{-1}u$ . Если положительные компоненты отсутствуют, то минимальная стоимость равна  $-\infty$ ; если же наименьшим является отношение  $k$ -х компонент, то будет заменяться  $k$ -й столбец матрицы  $B$ .
- 5) Переписываем  $B$  (или  $B^{-1}$ ) и решение  $x_B = B^{-1}b$  и переходим к первому пункту таблицы.

Эта методика иногда называется *исправленным* симплекс-методом, чтобы отличать ее от действий с таблицей. На самом же деле она и представляет собой (в сжатом виде) симплекс-метод.

Для завершения этого обсуждения остается решить, каким образом должны осуществляться первый, третий и пятый шаги,



различных углов, т. е. требует примерно  $m^2n$  действий — величина, сравнимая с объемом вычислений при использовании метода исключения для системы  $Ax=b$ , — чем и объясняется успех симплекс-метода. Но математические расчеты показывают, что длина пути не всегда может быть ограничена числом, кратным  $m$ , или фиксированной степенью числа  $m$ . Сейчас известно, что наилучшие допустимые множества могут дать путь, состоящий более чем из  $ct^{n/2}$  ребер, где число  $c$  не зависит от  $m$ , но по какой-то причине, лежащей в геометрических свойствах многомерных многогранников, симплекс-метод оказывается «везучим».

**Упражнение 8.2.9.** Проверить, что

$$E^{-1} = \begin{bmatrix} 1 & -v_1/v_k & & & \\ & \cdot & \cdot & & \\ & \cdot & \cdot & & \\ & & 1/v_k & & \\ & & \cdot & \cdot & \\ & & \cdot & \cdot & \\ & & & \cdot & \cdot \\ & & & & -v_n/v_k & 1 \end{bmatrix}, \quad \text{если} \quad E = \begin{bmatrix} 1 & v_1 & & & \\ & \cdot & \cdot & & \\ & \cdot & \cdot & & \\ & & \cdot & \cdot & \\ & & & v_k & \\ & & & \cdot & \cdot \\ & & & \cdot & \cdot \\ & & & & v_n & 1 \end{bmatrix}.$$

**Упражнение 8.2.10.** Показать, что матрица  $BE$  совпадает с  $B$ , кроме  $k$ -го столбца, который заменяется на  $Bv$  (что равняется искомому вектору  $u$ ). Поэтому  $BE$  действительно является базисной матрицей  $B$  для следующего шага, обратная к ней есть  $E^{-1}B^{-1}$  и сомножитель  $E^{-1}$  правильно корректирует обратную матрицу.

**Упражнение 8.2.11.** Можно ли найти в табл. 8.1 место, где требуется  $mn$  операций?

### § 8.3. ТЕОРИЯ ДВОЙСТВЕННОСТИ

Вторая глава начиналась с утверждения о том, что, хотя метод исключения и дает способ для решения системы  $Ax=b$ , возможен другой и более глубокий подход. То же самое справедливо и для линейного программирования. Механика симплекс-метода позволяет решить линейную программу, но центром теории линейного программирования является идея двойственности. Это красивая идея, и в то же время она очень важна для приложений. Объясним, как мы ее понимаем.

Начнем со стандартной задачи:

Исходная задача:

Минимизировать  $cx$  при условиях  $x \geq 0$  и  $Ax \geq b$ .

Но теперь вместо введения эквивалентной формулировки, содержащей уравнения вместо неравенств, двойственность ставит совершенно иную задачу. В двойственной задаче мы начинаем с тех же самых  $A$ ,  $b$  и  $c$ , но затем все делается наоборот.

В исходной задаче  $c$  было функцией стоимости и  $b$  был ограничением; в двойственной задаче эти векторы меняются местами. Более того, знак неравенства изменяется на противоположный и новые неизвестные составляют вектор-строку  $y$ , так что допустимое множество будет удовлетворять неравенству  $yA \leq c$  вместо  $Ax \geq b$ . Наконец, вместо минимизации мы будем максимизировать. Лишь требование неотрицательности остается неизменным; неизвестный вектор  $y$ , содержащий  $m$  компонент, должен удовлетворять условию  $y \geq 0$ . Коротче говоря, двойственной к задаче на минимум является следующая задача на максимум:

Двойственная задача.

Максимизировать  $yb$  при условиях  $y \geq 0$  и  $yA \leq c$ .

Двойственная к этой задаче совпадает с исходной задачей на минимум<sup>1)</sup>.

Естественно, я должен как-то проинтерпретировать все эти обращения. Они заключают в себе нечто вроде конкуренции между «минимизатором» и «максимизатором». Легче всего будет объяснить это, вернувшись к задаче о диете; я надеюсь, что вы проследите за ходом рассуждений еще раз. В задаче на минимум  $n$  неизвестных, соответствующих  $n$  различным продуктам питания, которые потребляются в (неотрицательных) количествах  $x_1, \dots, x_n$ . При этом  $m$  ограничений соответствуют  $m$  требуемым витаминам вместо одного ограничения на количество протеина, которое делалось в исходной задаче о диете. Элемент  $a_{ij}$  представляет собой количество  $i$ -го витамина в  $j$ -м продукте, и  $i$ -я строка системы  $Ax \geq b$  означает, что диета должна содержать этот витамин в количестве, не меньшем чем  $b_i$ . Наконец, если  $c_j$  есть стоимость  $j$ -го продукта, то  $c_1x_1 + \dots + c_nx_n = cx$  является стоимостью всей диеты. Минимизация этой стоимости и составляет цель исходной задачи.

В двойственной задаче аптекарь реализует не еду, а витамины в таблетках. Цены на витамины не отрицательны, и их можно варьировать. Основное ограничение здесь состоит в том, что он не может запрашивать за витамины больше, чем стоит настоящая еда. Поскольку  $j$ -й продукт питания содержит витамины в количествах  $a_{ij}$ , цена, назначаемая аптекарем за эквивалент в витаминах (и равная  $a_{1j}y_1 + \dots + a_{mj}y_m$ ), не может превосходить цену  $c_j$  настоящего продукта. Это и дает ограничение  $yA \leq c$ . С учетом этого ограничения аптекарь может продавать каждый витамин в количестве  $b_j$ , получая доход в размере  $y_1b_1 + \dots + y_mb_m = yb$ , который он стремится максимизировать.

<sup>1)</sup> Эти две задачи полностью симметричны. Мы начали с минимизации, но симплекс-метод в равной мере может быть применен к задаче максимизации — и в любом случае решаются обе задачи.

Следует отметить, что допустимые множества в этих двух задачах совершенно различны. Первое является подмножеством пространства  $\mathbb{R}^n$ , характеризуемым матрицей  $A$  и вектором ограничений  $b$ ; второе является подмножеством пространства  $\mathbb{R}^m$ , определяемым транспонированной к  $A$  матрицей и другим вектором  $c$ . Тем не менее при включении функций стоимости обе задачи будут содержать одни и те же входные данные  $A$ ,  $b$  и  $c$ . Вся теория линейного программирования базируется на соотношениях между ними, и мы приходим к фундаментальному результату:

**8D. Теорема двойственности.** *Если исходная задача имеет решение, то двойственная к ней задача также разрешима, и наоборот. При этом значения решений совпадают: минимум величины  $sx$  равняется максимуму величины  $yb$ . В противном случае, если оптимальных векторов не существует, либо оба допустимых множества являются пустыми, либо одно множество является пустым и другая задача имеет неограниченное решение (максимум равен  $+\infty$  или минимум равен  $-\infty$ ).*

Если оба допустимых множества являются непустыми, то реализуется первый и наиболее важный случай:  $sx^* = y^*b$ .

Теоретически этот факт устраняет конкуренцию между аптекарем и бакалейщиком, поскольку *результат всегда «ничейный»*. Мы получим похожую «теорему о минимаксе» и аналогичное равновесие в теории игр. Эти теоремы не означают, что покупатель ничего не платит за оптимальную диету или что матричная игра дает равные шансы обоим игрокам. Они означают, что для покупателя нет смысла предпочитать витамины обычной еде, несмотря на то что аптекарь гарантирует наличие заменителя для любого продукта и даже продает по более низкой цене заменитель такого дорогого продукта, как арахисовое масло. Мы хотим показать, что дорогие продукты не включаются в оптимальную диету, так что результат все-таки будет одним и тем же.

Это положение может показаться безвыходным, но я надеюсь, что оно не смутит вас, ведь решающая информация содержится в оптимальных векторах. В исходной задаче  $x^*$  указывает стратегию для покупателя; в двойственной задаче вектор  $y^*$  фиксирует естественные цены (или «теневые цены»), которых должна придерживаться экономика. В той мере, в которой наша линейная модель отражает реальную экономику, эти векторы представляют те решения, которые действительно следует принимать; их надо найти при помощи симплекс-метода, и теорема двойственности всего лишь говорит нам об их наиболее важном свойстве.

Приступим к доказательству теоремы. В некотором смысле может показаться очевидным, что аптекарь в состоянии поднимать цены до цен бакалейщика, но это не так. Точнее, верна лишь первая часть: поскольку каждый отдельный продукт может быть заменен своим витаминным эквивалентом без увеличения стоимости, все адекватные продуктовые диеты должны стоить не меньше, чем любая витаминная диета у аптекаря. Это одностороннее неравенство—*цена у аптекаря*  $\leq$  *цена у бакалейщика*,— но оно является основополагающим. Оно называется *слабой двойственностью* и легко доказывается для любой линейной программы и двойственной к ней:

8E. Если  $x$  и  $y$  являются допустимыми векторами в задачах на минимум и на максимум соответственно, то  $yb \leq cx$ .

Доказательство. Поскольку указанные векторы являются допустимыми, они удовлетворяют неравенствам

$$Ax \geq b \quad \text{и} \quad yA \leq c.$$

Более того, поскольку допустимость подразумевает также и неравенства  $x \geq 0$  и  $y \geq 0$ , мы можем перейти к скалярным произведениям, не нарушая при этом знаков неравенств:

$$yAx \geq yb \quad \text{и} \quad yAx \leq cx. \quad (7)$$

Поскольку левые части неравенств совпадают, мы получаем слабую двойственность  $yb \leq cx$ .

Одностороннее неравенство легко можно использовать. Во-первых, оно запрещает возможность того, что решения обеих задач являются неограниченными: если  $yb$ —произвольно большое число, то не существует даже допустимого вектора  $x$ , поскольку это противоречило бы неравенству  $yb \leq cx$ . Аналогичным образом, если исходная задача была неограниченной, т. е. стоимость  $cx$  могла убывать до  $-\infty$ , то двойственная задача не имеет допустимых векторов  $y$ .

Второй и столь же важный момент заключается в том, что любые векторы, для которых достигается равенство  $yb = cx$ , должны быть оптимальными. В этом случае цена у бакалейщика совпадает с ценой у аптекаря, и мы узнаем цены оптимальной продуктовой и оптимальной витаминной диет по тому факту, что покупателю не из чего выбирать:

8F. Если векторы  $x$  и  $y$  являются допустимыми и если  $cx = yb$ , то эти векторы должны быть оптимальными.

Доказательство. Согласно 8E, не существует такого допустимого вектора  $y$ , что  $yb$  становится больше, чем  $cx$ . По-

скольку вектор  $y$  достигает этого значения, он является оптимальным. Аналогично никакой допустимый вектор  $x$  не может сделать величину  $sx$  меньше, чем  $yb$ , и любой  $x$ , для которого этот минимум достигается, должен быть оптимальным.

Сформулируем пример с двумя продуктами и двумя витаминами. Заметим, что в двойственной задаче появляется матрица  $A^T$ , поскольку неравенство  $yA \leq c$  для вектор-строк эквивалентно  $A^T y^T \leq c^T$  для столбцов.

Исходная задача:

Минимизировать  $x_1 + 4x_2$   
 при условиях  $x_1 \geq 0, x_2 \geq 0,$   
 $2x_1 + x_2 \geq 6,$   
 $5x_1 + 3x_2 \geq 7.$

Двойственная задача:

Максимизировать  $6y_1 + 7y_2$   
 при условиях  $y_1 \geq 0, y_2 \geq 0,$   
 $2y_1 + 5y_2 \leq 1,$   
 $y_1 + 3y_2 \leq 4.$

Выбор  $x_1 = 3, x_2 = 0$  является допустимым и дает стоимость  $x_1 + 4x_2 = 3$ . В двойственной задаче числа  $y_1 = 1/2$  и  $y_2 = 0$  дают то же самое значение  $6y_1 + 7y_2 = 3$ . Следовательно, эти векторы должны быть оптимальными.

Все это кажется исключительно простым. Тем не менее имеет смысл взглянуть на проблему более внимательно и выяснить, что происходит в тот момент, когда неравенство  $yb \leq sx$  переходит в равенство. Это похоже на ситуацию в дифференциальном исчислении, где каждому известно условие максимума или минимума функции, состоящее в равенстве нулю первых производных, и тем не менее часто забывают, что это условие совершенно меняется при наличии ограничений. Наилучшим примером является наклонная прямая; производная никогда не обращается в нуль, дифференциальное исчисление почти бесполезно и максимум определенно достигается на правом конце интервала. Это в точности соответствует ситуации, с которой мы сталкиваемся в линейном программировании. Хотя переменных здесь больше и обычный интервал заменяется допустимым множеством нескольких измерений, максимум всегда достигается в углу допустимого множества. Говоря на языке симплекс-метода, существует оптимальный вектор  $x$ , являющийся *базисным*, т. е. он имеет лишь  $m$  ненулевых компонент.

Реальная трудность линейного программирования состоит в том, чтобы определить, какой именно угол является оптимальным. Следует отметить, что в решении этой задачи дифференциальное исчисление может оказаться полезным и даже весьма полезным, поскольку метод «множителей Лагранжа» вновь возвращает нас к нулевым производным в точках максимума, а двойственные переменные  $y$  как раз и являются множителями Лагранжа в задаче минимизации  $sx$ . Этот метод играет важную роль и в нелинейном программировании, но мы не будем касаться этого вопроса,

поскольку он выходит за рамки обсуждаемых задач. В математической форме условия на максимум и минимум с ограничениями даются ниже в (8), а сейчас мы хотим сформулировать их в экономических терминах: *диета  $x$  и цены на витамины  $y$  являются оптимальными, если*

(i) бакалейщик не продает продукт, цена которого больше цены соответствующего витаминного эквивалента;

(ii) аптекарь не берет денег за витамины, которые оказываются в излишке в диете.

В нашем примере  $x_2 = 0$ , поскольку второй продукт слишком дорог; его цена превосходит цену у аптекаря, так как  $y_1 + 3y_2 \leq 4$  превращается в строгое неравенство  $1/2 + 0 < 4$ . Аналогичным образом  $y_i = 0$ , если  $i$ -й витамин находится в избытке; он нам не требуется. В примере нам было необходимо семь единиц второго витамина, но диета фактически давала  $5x_1 + 3x_2 = 15$ , откуда мы нашли, что  $y_2 = 0$ . Легко видеть, что двойственность стала полной; мы получаем оптимальную пару лишь тогда, когда оба условия удовлетворяются.

Эти условия легко понять, если мы обратимся к матричной записи. Мы сравниваем вектор  $Ax$  с вектором  $b$  (вспомним, что условие допустимости есть  $Ax \geq b$ ) и ищем те компоненты, для которых равенство нарушается. Это соответствует избытку данного витамина, так что его цена становится равной  $y_i = 0$ . Одновременно мы сравниваем  $yA$  с  $c$  и хотим, чтобы все строгие неравенства (соответствующие дорогим продуктам) давали  $x_j = 0$  (исключение из диеты). Эти условия называются «условиями совместности невязок» в линейном программировании и «условиями Куна — Такера» в нелинейном программировании.

**8G. Теорема о равновесии.** Пусть допустимые векторы  $x$  и  $y$  удовлетворяют следующим условиям совместности невязок: если  $(Ax)_i > b_i$ , то  $y_i = 0$ , и если  $(yA)_j < c_j$ , то  $x_j = 0$ . (8)

Тогда векторы  $x$  и  $y$  являются оптимальными. И наоборот, если векторы являются оптимальными, то они должны удовлетворять условиям (8).

Доказательство. Основными уравнениями являются

$$yb = y(Ax) = (yA)x = cx. \quad (9)$$

Обычно мы можем быть уверены лишь в выполнении среднего из равенств. Для первого не вызывают сомнений неравенства  $y \geq 0$  и  $Ax \geq b$ , откуда  $yb \leq y(Ax)$ . Более того, равенство может выполняться лишь в одном случае: как только возникает неравенство  $b_i < (Ax)_i$ , множитель  $y_i$  при этой компоненте должен

обращаться в нуль. Тогда это неравенство не вносит вклад в скалярные произведения и равенство сохраняется.

То же самое справедливо и для оставшегося равенства: условие допустимости дает  $x \geq 0$  и  $yA \leq c$ , откуда получаем неравенство  $yAx \leq cx$ . Равенство достигается лишь при выполнении второго условия невязки: если цена оказывается недопустимой,  $(yA)_j < c_j$ , то для исправления положения мы должны умножить это неравенство на  $x_j = 0$ . В результате получаем в (9) равенство  $yb = cx$ , что гарантирует оптимальность  $x$  и  $y$  (и гарантируется их оптимальностью).

Пожалуй, об одностороннем неравенстве  $yb \leq cx$  было сказано достаточно. Доказать его было легко, оно дало нам простой тест для проверки векторов на оптимальность (они превращают неравенство в равенство), и, наконец, мы получили из него набор необходимых и достаточных условий невязки. Единственное, что еще осталось сделать, — это показать, что равенство  $yb = cx$  действительно возможно. До тех пор пока оптимальные векторы не будут получены, а сделать это непросто, теорема двойственности не доказана.

Чтобы получить оптимальные векторы, вернемся к симплекс-методу, предполагая, что оптимальный вектор  $x$  уже вычислен с его помощью. Наша цель состоит в том, чтобы одновременно найти оптимальный вектор  $y$ , показав тем самым, что метод остановился на нужном месте и для двойственной задачи, хотя решалась исходная. Для этого вернемся к началу процесса. Вводя переменные невязки  $w = Ax - b$  и переписывая условия допустимости в виде

$$[A - I] \begin{bmatrix} x \\ w \end{bmatrix} = b, \quad \begin{bmatrix} x \\ w \end{bmatrix} \geq 0, \quad (10)$$

мы преобразовали  $m$  неравенств  $Ax \geq b$  в уравнения. Затем на каждом шаге метода мы выбирали  $m$  столбцов прямоугольной матрицы  $[A - I]$  в качестве базисных и сдвигали их (по крайней мере теоретически) влево. В результате этого получалась матрица  $[B \ F]$  и соответствующий сдвиг в расширенном векторе стоимости  $[c \ 0]$  превращал его в  $[c_B \ c_F]$ . Условие остановки симплекс-метода имело вид  $r = c_F - c_B B^{-1} F \geq 0$ .

Мы знаем, что это условие в конце концов выполняется, поскольку число углов является конечным. В этот момент стоимость равнялась

$$cx = [c_B \ c_F] \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix} = c_B B^{-1}b - \text{минимальная стоимость.} \quad (11)$$

Если мы можем выбрать  $y = c_B B^{-1}$  в двойственной задаче, то, разумеется, будет выполняться равенство  $yb = cx$ ; минимум и

максимум совпадают. Поэтому нужно показать, что  $y$  удовлетворяет двойственным ограничениям  $yA \leq c$  и  $y \geq 0$ , т. е.

$$y[A - I] \leq [c \ 0]. \quad (12)$$

Когда в симплекс-методе мы преобразуем матрицу  $[A - I]$ , чтобы сделать базисные переменные первыми, это перегруппировывает ограничения так, что они приобретают вид

$$y[B \ F] \leq [c_B \ c_F]. \quad (13)$$

При выборе  $y = c_B B^{-1}$  первая половина ограничений превращается в равенство, а вторая — в  $c_B B^{-1} F \leq c_F$ , что в точности совпадает с условием останова  $r \geq 0$ , которое, как мы знаем, удовлетворяется. Поэтому вектор  $y$  является допустимым и теорема двойственности доказана. Выделив соответствующую матрицу  $B$  порядка  $m$ , которая является невырожденной по предположению, симплекс-метод выработал одновременно векторы  $y^*$  и  $x^*$ .

**Упражнение 8.3.1.** Найти двойственную к следующей задаче: минимизировать  $x_1 + x_2$  при условиях  $x_1 \geq 0$ ,  $x_2 \geq 0$ ,  $2x_1 \geq 4$ ,  $x_1 + 3x_2 \geq 11$ . Найти решение этой задачи и двойственной к ней и вычислить их общее значение.

**Упражнение 8.3.2.** Что является двойственной к следующей задаче: максимизировать  $y_2$  при условиях  $y_1 \geq 0$ ,  $y_2 \geq 0$ ,  $y_1 + y_2 \leq 3$ ? Решить эту задачу и двойственную к ней.

**Упражнение 8.3.3.** Пусть задана единичная матрица  $A$  (так что  $m = n$ ) и неотрицательные векторы  $b$  и  $c$ . Объяснить, почему вектор  $x^* = b$  является оптимальным для задачи на минимум, найти оптимальный вектор  $y^*$  для задачи на максимум и проверить, что значения совпадают. Что произойдет с векторами  $x^*$  и  $y^*$ , если некоторые компоненты вектора  $b$  станут отрицательными?

**Упражнение 8.3.4.** Построить одномерную задачу, для которой условия  $Ax \geq b$ ,  $x \geq 0$  являются недопустимыми и двойственная задача оказывается неограниченной.

**Упражнение 8.3.5.** Для матрицы второго порядка  $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$  найти такие векторы  $b$  и  $c$ , чтобы оба допустимых множества  $Ax \geq b$ ,  $x \geq 0$  и  $yA \leq c$ ,  $y \geq 0$  были пустыми.

**Упражнение 8.3.6.** Показать, что если все элементы в  $A$ ,  $b$  и  $c$  являются положительными, то исходная и двойственная задачи автоматически являются допустимыми.

**Упражнение 8.3.7.** Показать, что векторы  $x = (1, 1, 1, 0)$  и  $y = (1, 1, 0, 1)$  являются допустимыми для обычных двойственных задач, если

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad c = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 3 \end{bmatrix}.$$

Вычислив  $sx$  и  $yb$ , объяснить, почему они являются оптимальными.

**Упражнение 8.3.8.** Проверить, что векторы из предыдущего упражнения удовлетворяют условиям совместности невязок (8), и найти одно неравенство для невязок в исходной и двойственной задачах.

**Упражнение 8.3.9.** Пусть  $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ,  $b = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$  и  $c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ . Найти оптимальные  $x$  и  $y$  и проверить выполнение условий совместности невязок (а также условия  $yb = cx$ )

**Упражнение 8.3.10.** Если в исходной задаче ограничения-неравенства заменить равенствами, т. е. *минимизировать*  $sx$  при условиях  $Ax = b$  и  $x \geq 0$ , то требование  $y \geq 0$  в двойственной задаче снимается: *максимизировать*  $yb$  при  $yA \leq c$ . Показать, что одностороннее неравенство  $yb \leq sx$  по-прежнему остается в силе. Почему условие  $y \geq 0$  является необходимым в (7), а здесь становится излишним?

### ТЕНЕВЫЕ ЦЕНЫ

Как будет изменяться минимальная стоимость при изменении правой части  $b$  или вектора стоимости  $c$ ? Этот вопрос относится к анализу устойчивости и позволяет нам выжать из симплекс-метода много дополнительной информации, содержащейся в нем. Эти вопросы, связанные с *предельной стоимостью*, являются наиболее важными для экономиста или руководителя компании.

Если допустить сильные изменения в векторах  $b$  и  $c$ , то оптимальное решение будет изменяться скачкообразно. При повышении цены на яйца они в определенный момент будут исключены из диеты; говоря языком линейного программирования, переменная  $x_{\text{egg}}$  скачкообразно перейдет из базисной в свободную. Чтобы проследить за этим изменением, необходимо ввести так называемое параметрическое программирование. Но если изменения являются небольшими, как обычно и бывает, то *оптимальный угол будет оставаться оптимальным*; базисные переменные останутся прежними. Другими словами, матрицы  $B$  и  $F$  не изменятся. Геометрически это означает, что мы слегка сдвигаем допустимое множество (изменяя вектор  $b$ ) и наклоняем семейство плоскостей, чтобы учесть этот сдвиг (изменяя вектор  $c$ ); если эти изменения невелики, то касание произойдет в том же самом углу (слегка сдвинутом).

В конце работы симплекс-метода, когда набор базисных переменных уже известен, соответствующие  $m$  столбцов матрицы  $A$  будут образовывать матрицу  $B$ . В этом углу

$$\text{минимальная стоимость} = c_B B^{-1} b = y^* b.$$

Таким образом, сдвиг на  $\Delta b$  изменяет минимальную стоимость на  $y^* \Delta b$ . Решение  $y^*$  двойственной задачи дает *скорость изменения минимальной стоимости* (т. е. производную) при изменении вектора  $b$ . Компоненты вектора  $y^*$  называются *теневыми ценами*, и они играют определенную роль. Если потребность в витамине  $B_1$  возрастает на  $\Delta$ , а его цена у аптекаря равняется

$y_1^*$  и он полностью конкурентоспособен с бакалейщиком, то стоимость диеты (аптекаря или бакалейщика) поднимается на  $y_1^* \Delta$ . Если  $y_1^*$  равняется нулю, то витамин  $B_1$  является «бесплатным» и небольшие изменения в его потреблении не окажут никакого эффекта; диета уже содержит его более чем достаточно.

Нас интересует другой вопрос. Пусть мы требуем, чтобы диета содержала по крайней мере небольшое количество яиц, т. е. условие неотрицательности  $x_{\text{egg}} \geq 0$  заменяется на  $x_{\text{egg}} \geq \delta$ . Как это скажется на стоимости?

Если яйца уже присутствовали в оптимальной диете, то никаких изменений не произойдет; новое требование уже удовлетворено и не вызывает дополнительных расходов. Но если они не содержались в диете, то включение количества  $\delta$  обойдется в некоторую сумму. Увеличение не будет равняться полной цене  $c_{\text{egg}} \delta$ , поскольку мы можем уменьшить количество остальных продуктов питания и сэкономить на этом. На самом деле увеличение стоимости зависит от *уменьшенной стоимости* яиц — их собственной цены без цены, которую мы платим за эквивалентное количество более дешевых продуктов. Для ее вычисления вернемся к уравнению (2): любая новая диета имеет

$$\text{стоимость} = (c_F - c_B B^{-1} F) x_F + c_B B^{-1} b = r x_F + c_B B^{-1} b.$$

Если количество яиц является первой свободной переменной, то увеличение первой компоненты вектора  $x_F$  на величину  $\delta$  увеличит стоимость на  $r_1 \delta$ . Поэтому уменьшенная стоимость равняется  $r_1$ . Аналогичным образом вектор  $r$  дает уменьшенные стоимости для всех остальных продуктов, не являющихся базисными, т. е. изменение общей стоимости при движении нижней границы для вектора  $x$  (ограничений на неотрицательность) вверх. Мы знаем, что  $r \geq 0$ , поскольку это условие было критерием останова. Из экономических соображений следует то же самое, т. е. что уменьшенная стоимость яиц не может быть отрицательной, поскольку в противном случае они уже были бы включены в оптимальную диету.

**Упражнение 8.3.11.** Если исходная задача имеет единственное оптимальное решение  $x^*$  и вектор  $c$  слегка изменяется, то почему при этом  $x^*$  по-прежнему будет оптимальным решением?

**Упражнение 8.3.12.** Если стоимость  $c_1$  бифштекса равна 3 долларам, а стоимость арахисового масла  $c_2$  равна 2 долларам и они дают соответственно две и одну единицы протеина (а требуется четыре единицы), то чему равняется теневая цена протеина и уменьшенная стоимость арахисового масла?

### ТЕОРИЯ НЕРАВЕНСТВ

Двойственность можно изучать различными способами. Метод, которым мы воспользовались — доказательство неравенства  $yb \leq cx$  и последующее использование симплекс-метода для получения

равенства,— был удобен тем, что для него все уже было готово. Но доказательство получилось слишком длинным. Правда, оно также было и *конструктивным*, поскольку мы фактически нашли оптимальные векторы  $x^*$  и  $y^*$ . Теперь мы кратко опишем иной подход, в котором используется не техника симплекс-метода, а геометрические соображения. Я думаю, что основные идеи не станут менее ясными (а, быть может, станут яснее), если мы опустим некоторые детали.

Наилучшей иллюстрацией этого подхода является основная теорема линейной алгебры. Вспомним, что задачей второй главы было решение системы  $Ax=b$ , или, другими словами, отыскание вектора  $b$  в пространстве столбцов матрицы  $A$ . После метода исключения и после рассмотрения четырех основных подпространств вопрос о разрешимости системы был решен совершенно иным способом в упр. 2.5.11:

|| **8N.** Либо уравнение  $Ax=b$  имеет решение, либо существует такой вектор  $y$ , что  $yA=0$ ,  $yb \neq 0$ .

Это — *альтернативная теорема*, поскольку одновременно найти  $x$  и  $y$  невозможно: если  $Ax=b$ , то  $yAx=yb \neq 0$ , что противоречит равенству  $yAx=0x=0$ . В терминах подпространств это означает, что либо  $b$  находится в пространстве столбцов матрицы  $A$ , либо он имеет ненулевую компоненту, лежащую в ортогональном подпространстве, которое является левым нуль-пространством матрицы  $A$ . Эта компонента и дает требуемый вектор  $y^1$ .

Мы хотим найти подобную теорему для неравенств. Для этого целесообразно начать с той же системы  $Ax=b$ , но при дополнительном ограничении  $x \geq 0$ , т. е. с ответа на вопрос, когда система  $Ax=b$  имеет не просто решение, а *неотрицательное решение*? Другими словами, нужно найти условия, при которых допустимое множество задачи с ограничениями-равенствами является непустым.

Для ответа на этот вопрос придется обратиться к понятию достижимых векторов, т. е. векторов  $b$ , для которых система  $Ax=b$  разрешима. В главе 2 было показано, что при произвольном векторе  $x$  вектор  $Ax$  может быть любой комбинацией столбцов матрицы  $A$ ; вектор  $b$  был достижимым, если он лежал в пространстве столбцов. Теперь мы допускаем лишь неотрицательные комбинации и достижимые векторы  $b$  уже не образуют подпространство. Вместо этого они образуют конусообразную область (рис. 8.5). Для матрицы размера  $m \times n$  мы будем иметь  $n$  столбцов в  $m$ -мерном пространстве, и конус превращается в некое

<sup>1)</sup> Легко видеть, что это доказательство не является конструктивным. Мы знаем только, что такая компонента должна существовать либо вектор  $b$  должен лежать в пространстве столбцов.

подобие пирамиды без основания. На рисунке изображены четыре столбца в двумерном пространстве, и матрица  $A$  будет иметь размер  $2 \times 4$ . Если вектор  $b$  лежит в этом конусе, то система  $Ax = b$  имеет неотрицательное решение; в противном случае такого решения не существует.

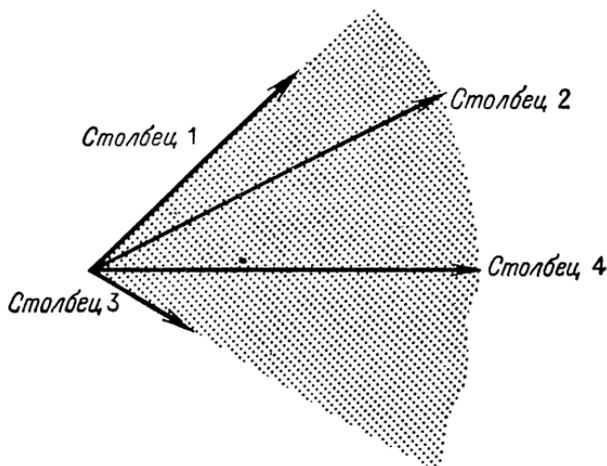


Рис. 8.5. Конус, образованный неотрицательными комбинациями столбцов.

Задача состоит в отыскании альтернативы: что произойдет, если вектор  $b$  будет лежать вне конуса? Эта ситуация изображена на рис. 8.6 и ее геометрия очевидна. Существует «разде-

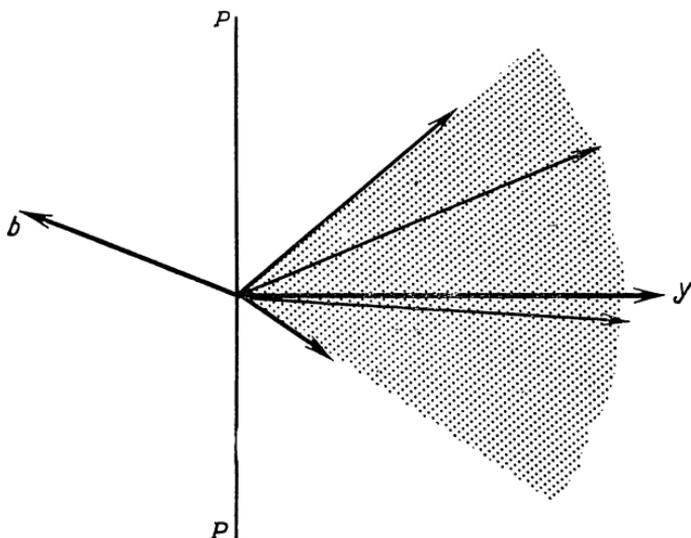


Рис. 8.6. Вектор  $b$  расположен вне конуса и по другую сторону от разделяющей гиперплоскости  $P$ .

ляющая гиперплоскость», проходящая через начало координат и такая, что вектор  $b$  лежит по одну сторону, а весь конус — по другую. (Приставка *гипер* означает, что число измерений может быть большим; плоскость же состоит, как обычно, из всех векторов, перпендикулярных фиксированному вектору  $y$ .) Скалярное произведение векторов  $y$  и  $b$  отрицательно, поскольку угол между ними больше  $90^\circ$ , а скалярное произведение вектора  $y$  на любой столбец матрицы  $A$  положительно. В терминах матриц это означает, что  $yb < 0$  и  $yA \geq 0$ , что и является искомым альтернативой:

|| 81. Либо система  $Ax=b$  имеет неотрицательное решение, либо существует вектор  $y$ , такой, что  $yA \geq 0$ ,  $yb < 0$ .

Это теорема о разделяющей гиперплоскости. Она является фундаментальной для математической экономики, и доказательство ее можно найти в прекрасной книге Гейла по теории линейных экономических моделей.

**Пример.** Если  $A=I$ , то соответствующим конусом является положительный квадрант. Каждый вектор  $b$  из этого квадранта является неотрицательной линейной комбинацией вектор-столбцов:

$$\text{если } b = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \text{ то } b = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Для произвольного вектора  $b$ , лежащего вне этого квадранта, должна выполняться вторая возможность:

$$\text{если } b = \begin{bmatrix} 2 \\ -3 \end{bmatrix}, \text{ то } y = [0 \ 1], \text{ так что } yA \geq 0, \text{ но } yb = -3.$$

Эта теорема непосредственно приводит к целой последовательности подобных вариантов (см. книгу Гейла). Мы почти готовы поверить, что как только мы встречаемся с двумя взаимоисключающими предположениями, то или другое из них должно выполняться. Например, подпространство  $S$  и его ортогональное дополнение  $S^\perp$  не могут одновременно содержать положительные векторы: их скалярное произведение будет положительным, в то время как ортогональные векторы дают нулевое скалярное произведение. С другой стороны, совсем не факт, что  $S$  либо  $S^\perp$  вообще должны содержать положительный вектор: если  $S$  есть ось  $x$ , а  $S^\perp$  — ось  $y$ , то они содержат лишь «полуположительные» векторы  $[1 \ 0]$  и  $[0 \ 1]$ . Интересно, что несколько более слабая альтернатива работает: либо  $S$  содержит положительный вектор  $x$ , либо  $S^\perp$  содержит полуположительный вектор  $y$ . Если  $S$  и  $S^\perp$  представляют собой перпендикулярные прямые на плоскости, то

легко видеть, что одна или другая должна проходить через первый квадрант, но мне трудно представить, что будет в случае большего числа измерений.

Для линейного программирования наиболее важной является альтернатива для ограничений в виде неравенств:

|| 8J. Либо система  $Ax \geq b$  имеет неотрицательное решение, либо существует такой вектор  $y$ , что  $yA \geq 0$ ,  $yb < 0$ ,  $y \leq 0$ .

Это следует из 8I, если мы введем переменные невязки  $w = Ax - b$ , чтобы преобразовать неравенство в равенство:

$$\begin{bmatrix} A & -I \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} = b.$$

Если не существует неотрицательного решения, то, согласно 8I, существует такой вектор  $y$ , что

$$y[A \ -I] \geq [0 \ 0], \quad yb < 0.$$

Это и дает вторую возможность в 8J, откуда получается «неконструктивное доказательство» теоремы двойственности. Но мы обещали придерживаться геометрических рассуждений, опуская алгебраические, и сдерживаем это обещание.

**Упражнение 8.3.13.** Для матрицы  $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$  описать конус, образованный ее столбцами. Для вектора  $b$ , лежащего внутри этого конуса, скажем  $b = (3, 2)^T$ , найти единственный допустимый вектор  $x$ . Если  $b$  лежит вне конуса, скажем  $b = (0, 1)^T$ , то какой вектор  $y$  будет удовлетворять второй возможности?

**Упражнение 8.3.14.** Если задано трехмерное пространство, то можно ли найти шесть векторов, неотрицательные комбинации которых образуют конус, совпадающий со всем пространством?

**Упражнение 8.3.15.** Используя 8H показать, что (в силу справедливости альтернативы) задача

$$\begin{bmatrix} 2 & 2 \\ 4 & 4 \end{bmatrix} x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

не имеет решения.

**Упражнение 8.3.16.** Используя 8I показать, что (в силу справедливости альтернативы) задача

$$\begin{bmatrix} 1 & 3 & -5 \\ 1 & -4 & -7 \end{bmatrix} x = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

не имеет неотрицательного решения.

**Упражнение 8.3.17.** Показать, что две возможности 8J не могут выполняться одновременно.

## § 8.4. СЕТЕВЫЕ МОДЕЛИ

Некоторые линейные программы имеют такую структуру, что их решение получается легко. Мы встречались уже с такой возможностью при решении линейных уравнений  $Ax = b$  в гл. 1, где для ленточных матриц, ненулевые элементы которых концентрируются вокруг главной диагонали, затраты на реализацию метода исключений уменьшались до порядка  $n$  вместо  $n^3$ . В линейном же программировании мы будем выделять специальный класс задач, известных под названием *сетевых*. Здесь также матрица  $A$  будет содержать большое число нулей, а ненулевые ее элементы обычно равняются  $+1$  или  $-1$ . (Но диагональ ничем не выделяется, поскольку матрицы являются прямоугольными.) Для этих задач шаги метода исключения требуют лишь простых операций сложения и умножения, и здесь могут решаться значительно большие задачи, чем обычно. Следует отметить, однако, что число шагов в методе по-прежнему может экспоненциально возрастать в наихудшем случае (см. статью в журнале *Scientific American* за январь 1978 г.<sup>1)</sup>), но, как и в общем симплекс-методе, это очень редко становится реальной опасностью.

Разумеется, высокая скорость решения интересует нас только тогда, когда сами задачи являются важными. Сетевые задачи имеют приложения ко всем отраслям индустриального и экономического планирования, например к распределению потоков товаров, людей, газа, нефти или воды. В качестве примера можно взять задачу о перегрузке товаров: готовая продукция должна распределяться по складам, а оттуда попадать к потребителям, и все это должно осуществляться с минимальными расходами. Источники, склады и потребители представляют собой *узлы* сети, а пути между ними называются *дугами*.

Одна модель такого сорта показана на рис. 8.7, где на каждой дуге указана ее максимальная пропускная способность. Задача состоит в получении максимально возможного потока от источника (узел 1) к стоку (узел 6). Направления потоков, указанные стрелками, не могут быть изменены; расходы на перевозку также не учитываются. Это частный случай сетевой задачи, называемый *задачей о максимальном потоке*.

Неизвестными величинами являются потоки  $x_{ij}$  из одного узла  $i$  в другой узел  $j$ . Мы пользуемся двухиндексным обозначением для удобства, несмотря на то что  $x$  является вектором. В рассмотрении участвуют только пары вида  $x_{12}, x_{24}, \dots, x_{56}$ , соответствующие дугам, изображенным на рисунке. Если верхние

<sup>1)</sup> H. R. Lewis, C. H. Papadimitriou, The efficiency of algorithms, 302 (1978), 96—109.

пределы (пропускные способности) этих дуг равняются  $u_{ij}$ , то ограничения на отдельные потоки имеют вид

$$0 \leq x_{ij} \leq u_{ij}^1).$$

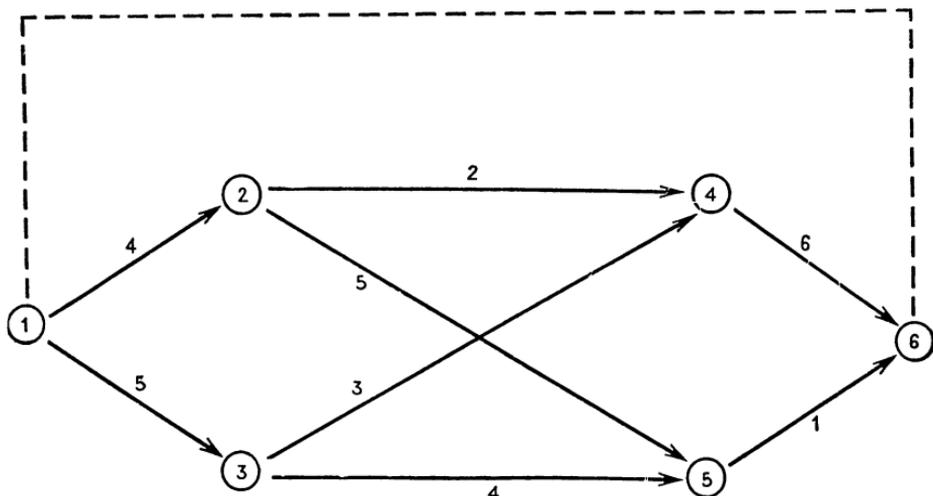


Рис. 8.7. Система с ограниченной пропускной способностью.

В дополнение к этому существует также ограничение на каждый промежуточный узел, состоящее в равенстве входящего и выходящего потоков. Это справедливо также для источника и стока, если мы введем в рассмотрение фиктивную трубу с бесконечно большой пропускной способностью (на рисунке она обозначена пунктиром) и будем возвращать все назад от стока к источнику. Суммарный поток, поступающий в узел с номером  $j$ , равен  $\sum x_{ij}$  по всем  $i$ , а суммарный поток из  $j$ -го узла равен  $\sum x_{jk}$  по всем  $k$ . Поэтому ограничение по равновесию имеет вид

$$\sum_i x_{ij} - \sum_k x_{jk} = 0, \quad j = 1, 2, \dots, 6.$$

Матрица коэффициентов для этого уравнения имеет специальную структуру и называется **матрицей инцидентности**. Каждому из шести узлов соответствует строка, и каждой из девяти дуг соответствует столбец матрицы; числа  $+1$  и  $-1$  полностью опи-

<sup>1)</sup> Верхняя и нижняя границы на каждую из компонент вектора  $x$  называются «**ящичными**» ограничениями, что вызвано формой допустимого множества; они очень просты и хорошо исследованы.

связают связи между узлами и дугами:

$$A = \begin{bmatrix} 1 & 1 & & & & & & & -1 \\ -1 & & 1 & 1 & & & & & \\ & -1 & & & 1 & 1 & & & \\ & & -1 & & -1 & & 1 & & \\ & & & -1 & & -1 & & 1 & \\ & & & & & & -1 & -1 & 1 \end{bmatrix}.$$

$x_{12}$     $x_{13}$     $x_{24}$     $x_{25}$     $x_{34}$     $x_{35}$     $x_{46}$     $x_{56}$     $x_{61}$

В задаче о максимальном потоке мы стремимся сделать величину  $x_{61}$  как можно большей, так что эта задача представляет собой обычную линейную программу (упр. 8.4.3) и можно пользоваться симплекс-методом. Но мы хотим воспользоваться укороченным вариантом. Читатель может определить максимальный поток, не прибегая к помощи ЭВМ, так что мы сразу перейдем к рассмотрению условия, показывающего, что поток действительно является максимальным и не может быть увеличен. Прежде всего, введем понятия *разреза* сети, означающего разделение всех узлов на две группы  $S$  и  $S'$  с источником в группе  $S$  и стоком в  $S'$ . *Пропускной способностью разреза* назовем сумму пропускных способностей по всем дугам, идущим из  $S$  в  $S'$ .

Например, если  $S$  содержит первый и третий узлы, то пропускная способность этого разреза равна  $4 + 3 + 4 = 11$ . Поток, больший чем 11, невозможен, поскольку он не сможет пройти через разрез.

**8К. Теорема о максимальном потоке и минимальном разрезе.** Максимальный поток в сети равняется пропускной способности минимального разреза.

**Доказательство.** Неравенство в одну сторону очевидно: для произвольного потока и произвольного разреза поток не превосходит пропускной способности разреза. Никакой поток не в состоянии сделать больше, чем просто насытить все дуги, пересекающие разрез, и их общая пропускная способность является верхней гранью для потока. Эта ситуация аналогична «слабой двойственности», т. е. простому неравенству  $yb \leq cx$ , которое выполняется для любых допустимых векторов  $x$  и  $y$ . Более трудная задача, как здесь, так и в теореме двойственности, состоит в доказательстве достижимости равенства.

Пусть некоторый поток является максимальным. Рассмотрим все дуги, которые не достигли предела пропускной способности, и пусть  $S$  содержит источник, а также все узлы, которых можно

от него достичь по этим дугам. Остальные узлы образуют группу  $S'$ . Тогда сток будет находиться в  $S'$ , поскольку в противном случае поток не был бы максимальным. Более того, каждая дуга, соединяющая  $S$  с  $S'$ , является заполненной, в противном случае узел, в который она входит, принадлежал бы группе  $S$ , а не  $S'$ . Следовательно, поток действительно наполняет разрез и равенство достигается.

Это дает нам возможность проверять, что заданный поток является максимальным; нужно лишь найти соответствующий разрез. В нашем примере максимальный поток равен 6, и соответствующий разрез изображен на рис. 8.8.

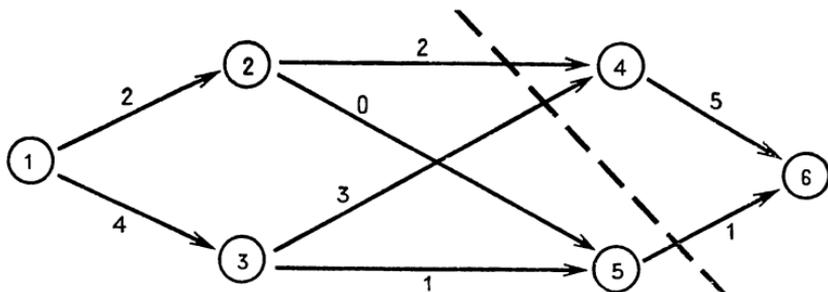


Рис 8.8. Максимальный поток и минимальный разрез.

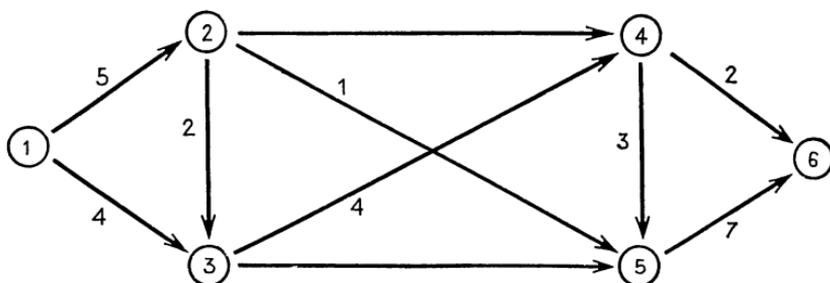
Эта теорема дает также и алгоритм решения: для любого потока нужно вычислять неиспользованную пропускную способность каждой дуги, и если сток может быть достигнут по какому-то недозаполненному пути, то нужно увеличить поток по нему до максимально возможного и вычислить результат. (Заметим, что если происходит пересылка из узла  $A$  в узел  $B$ , то почти все может вернуться назад, не обращая при этом суммарный поток. Другими словами, в узел  $A$  можно попасть из  $B$ .) Постепенно должен быть достигнут максимальный поток, и в случае целочисленных пропускных способностей целочисленным будет и поток, т. е. эта задача относится к *целочисленному программированию*.

**Упражнение 8.4.1.** Если мы можем увеличить пропускную способность одной из труб в сети, то какую именно нужно избрать, чтобы получить наибольшее увеличение максимального потока?

**Упражнение 8.4.2.** Найти максимальный поток и минимальный разрез для сети, изображенной на стр. 392

**Упражнение 8.4.3.** В задаче о максимальном потоке, вводя переменные невязки  $w_{ij} = u_{ij} - x_{ij}$ , максимизировать  $x_{n+1}$  при условиях

$$\begin{bmatrix} A & 0 \\ I & I \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ u \end{bmatrix}, \quad \begin{bmatrix} x \\ w \end{bmatrix} \geq 0.$$



Двойственная к этой задаче содержит шесть неизвестных  $y_i$  и девять неизвестных  $z_{ij}$ :

$$\text{минимизировать } [y \ z] \begin{bmatrix} 0 \\ u \end{bmatrix} = zu \text{ при } [y \ z] \begin{bmatrix} A & 0 \\ I & I \end{bmatrix} \geq [0 \ \dots \ 1 \ 0]$$

Показать, что выбор  $z_{ij}=1$  для дуг, пересекающих минимальный разрез, и  $z_{ij}=0$  для остальных с  $y_i=0$  для узлов из  $S$  и  $y_i=1$  для узлов из  $S'$  является одновременно допустимым и оптимальным

#### ЗАДАЧА О ПРОСТОМ НАЗНАЧЕНИИ

Предположим, что у нас имеется  $m$  людей и  $n$  работ. Каждый из людей может выполнять лишь определенные работы. Можно ли назначить всех на работу по специальности, не давая при этом одну и ту же работу разным людям?

Нетрудно найти необходимые условия, при которых такое назначение будет возможно. Во-первых, каждый из людей должен иметь по крайней мере одну специальность. Во-вторых, каждая пара людей должна иметь две различные специальности. И каждая группа из трех (или  $k$ ) людей должна быть в состоянии выполнять по меньшей мере три (или  $k$ ) различные работы. В противном случае цель, очевидно, не может быть достигнута. Задача состоит в том, чтобы показать, что если каждая группа удовлетворяет перечисленным выше необходимым условиям, то распределение возможно.

**8L.** *Назначение возможно тогда и только тогда, когда каждая группа в состоянии выполнять достаточное число работ; если группа состоит из  $k$  людей, то вместе они должны быть в состоянии выполнять по меньшей мере  $k$  работ.*

На первый взгляд это условие выглядит чересчур слабым, поскольку один очень квалифицированный человек в состоянии помочь всей группе выполнить его. Но ведь это условие относится к *каждой* группе, включая и те, в которых этот человек не состоит. Если распределение работы невозможно, то всегда можно найти группу, для которой это условие не выполняется.

Доказательство начинается с конструирования сети, имеющей пропускную способность  $m$  между каждым человеком и работами, которые он может выполнять. Затем люди подключаются к источнику, а работы к стоку при помощи дуг с пропускной способностью 1 (рис. 8.9). Если в результате суммарный поток будет

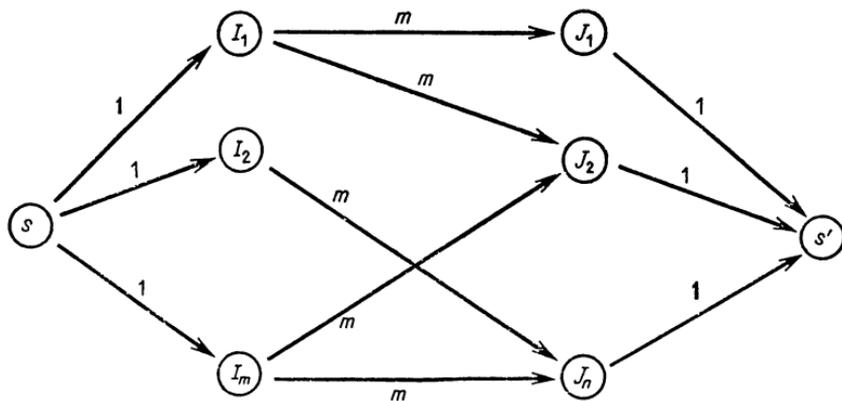


Рис. 8.9. Сеть работ.

равен  $m$ , то все дуги от источника будут заполнены и все будут иметь работу.

Предположим, что максимальный поток оказался меньше, чем  $m$ . Тогда должен существовать разрез с пропускной способностью, меньше  $m$ , например, такой, что в группе  $S$  содержатся источник и узлы  $I_1, \dots, I_k, J_1, \dots, J_l$ . Ни один из этих  $k$  людей не в состоянии выполнять другие работы, поскольку в противном случае нашлась бы дуга с пропускной способностью  $m$ , пересекающая разрез. Следовательно, пропускная способность разреза равняется  $m - k$  (от источника к оставшимся людям) плюс  $l$  (от этих работ к стоку), что дает в сумме  $m - k + l$ ; эта величина будет меньше  $m$  при  $l < k$ . Следовательно, мы нашли группу из  $k$  людей, которая может выполнять лишь  $l < k$  работ. Если назначение оказывается невозможным — поток оказался меньше, чем  $m$ , — то это означает, что какая-то группа людей сдерживает его.

**Упражнение 8.4.4.** Предположим, что каждый человек может выполнять по две работы и не существует трех людей с одной и той же специальностью. Доказать, что в этом случае назначение возможно.

**Упражнение 8.4.5.** Предположим, что имеется бесконечно много людей и работ. Первый человек может выполнить любую работу, а для последующих  $(i+1)$ -й человек имеет только  $i$ -ю специальность. Показать, что каждая группа из  $k$  людей может выполнять по меньшей мере  $k$  работ, но распределить всех на работу невозможно.

Задача о простом назначении называется также задачей о **бракосочетании**. Каждая девушка выбирает только определенных юношей. Чтобы каждая девушка вышла замуж, каждая группа из  $k$  девушек ( $k=1, 2, \dots$ ) должна любить по меньшей мере  $k$  юношей — и этого достаточно. Я не знаю, что произойдет, если юноши также захотят выбирать. По крайней мере будет более реалистично, если девушки будут иметь шкалу предпочтений вместо простого «да» или «нет» и если у нанимаемых на работу людей также будут градации, когда  $i$ -й человек имеет коэффициент  $a_{ij}$  для  $j$ -й работы, причем  $a_{ij}=0$  в случае полного отсутствия соответствующей квалификации. Это приводит к задаче об *оптимальном назначении*: максимизировать сумму коэффициентов пригодности при данных работах. Эта задача также является линейной программой: максимизировать  $\sum \sum a_{ij}x_{ij}$ , когда на одной работе занят один человек и имеется по одной работе на человека.

Другой известный пример дает *задача о коммивояжере*. Начиная от своего дома, он должен посетить  $n-1$  городов и вернуться домой. Если расстояние (или стоимость проезда) от одного города  $i$  до другого города  $j$  равняется  $a_{ij}$ , то коммивояжер должен минимизировать  $\sum \sum a_{ij}x_{ij}$  при условиях, что он должен посетить каждый город по одному разу:

$$\sum_i x_{ij} = 1, \quad \sum_j x_{ij} = 1, \quad x_{ij} \geq 0.$$

Такая постановка не вполне корректна, поскольку решение задачи может представлять собой отдельные группы и быть замкнутым внутри них, так что, выехав из дома, коммивояжер никогда не попадает во все требуемые города. Поэтому необходимо ввести дополнительные ограничения, и их оказывается так много, что эта на первый взгляд простая задача становится на самом деле исключительно трудной.

Соответствующая сетевая задача состоит в отыскании кратчайшего пути от источника к стоку. На каждой дуге задается время перехода  $t_{ij}$  и рассматривается двойственная задача: каждому узлу соответствует минимальное время  $y_i$ , и мы максимизируем  $y$  для стока при условиях

$$y_j - y_i \leq t_{ij} \quad \text{и} \quad y = 0 \quad \text{у источника.}$$

Эта задача может быть решена методами **динамического программирования**. Начиная от источника, мы находим минимальное время, требуемое для достижения соседнего узла, и постепенно движемся вперед; сделать это легко, если циклы в сети запрещены. Существенным принципом в динамическом программировании является принцип рекурсивной оптимизации: оптимальный путь от источника к стоку одновременно является

оптимальным путем от источника к любому узлу на этом пути и от этого узла к стоку. Каково бы ни было текущее состояние, оставшиеся решения должны быть оптимальными. В результате задача разбивается на простые части.

Отметим также **задачу о перевозках**, т. е. о сетевом потоке от производителей к рынкам без промежуточных точек. Производство  $i$  дает продукцию в количестве  $s_i$ , а потребность на рынке  $j$  равняется  $d_j$ ; перевозка одной единицы продукции между ними стоит  $c_{ij}$ . Следовательно, посылаемые количества  $x_{ij}$  должны удовлетворять неравенствам

$$\sum_j x_{ij} \leq s_i, \quad \sum_i x_{ij} \geq d_j, \quad x_{ij} \geq 0.$$

Для оптимальности перевозок мы будем минимизировать величину  $\sum \sum c_{ij} x_{ij}$ .

**Упражнение 8.4.6.** Если числа в упр. 8.4.2 представляют собой время на перевозки, то как будет выглядеть кратчайший путь от источника к стоку? Найти минимальные значения времени  $y_i$  для всех других узлов (частный случай динамического программирования) и объяснить, почему на каждой дуге выполняется неравенство  $y_j - y_i \leq t_{ij}$ .

**Упражнение 8.4.7.** Решить задачу о перевозках, когда имеются два производства с объемами  $s_i=6$ , три рынка с потребностями  $d_j=4$  и стоимости перевозок  $a_{ij}$  образуют матрицу

$$A = \begin{bmatrix} 1 & 3 & 6 \\ 2 & 4 & 8 \end{bmatrix}.$$

**Упражнение 8.4.8.** Какие условия на производство и потребление являются допустимыми условиями задачи о перевозках?

## § 8.5. ТЕОРИЯ ИГР И ТЕОРЕМА О МИНИМАКСЕ

Матричную игру лучше всего объяснить на примере.

**Пример 1.** В игре участвуют двое и правила остаются неизменными:

Игрок  $X$  поднимает вверх одну или обе руки, и независимо от него то же самое делает и игрок  $Y$ . Если их решения совпали, то  $Y$  получает 10 долларов. Если они приняли разные решения, то выигрывает  $X$ , причем он получает 10 долларов, если он поднял вверх одну руку, и 20 долларов, если обе. Суммарный выигрыш игрока  $X$  легко можно записать в виде матрицы

$$A = \begin{bmatrix} -10 & 20 \\ 10 & -10 \end{bmatrix} \begin{array}{l} \text{одна рука у } Y \\ \text{обе руки у } Y \end{array}.$$

одна обе  
рука руки  
у  $X$  у  $X$

Если немного подумать, то примерная стратегия станет очевидной. Ясно, что игрок  $X$  не будет играть одинаково все время, иначе  $Y$  будет копировать его и выиграет все. Аналогично игрок  $Y$  не может придерживаться одной стратегии, иначе  $X$  будет поступать наоборот. Оба игрока должны использовать *смешанную стратегию*, и, более того, выбор в каждой партии должен быть абсолютно независимым от предыдущих партий. В противном случае, если другой игрок заметит некоторую систему, он получит преимущество. Даже стратегия типа «придерживаться одного и того же выбора до тех пор, пока идет выигрыш, и менять стратегию после первого проигрыша» является, очевидно, бесполезной. После достаточного числа партий партнер будет знать точно, что вы будете делать. Таким образом, игроки оказываются в следующей ситуации: игрок  $X$  решает, что он будет поднимать одну руку с частотой  $x_1$  (тогда он поднимает обе руки с частотой  $x_2 = 1 - x_1$ , и в каждой партии это решение является случайным), а  $Y$  выбирает *свои* вероятности  $y_1$  и  $y_2 = 1 - y_1$ . Как мы видели, ни одна из этих вероятностей не должна равняться 0 или 1; в противном случае соперник подстроится и начнет выигрывать. В то же время сомнительно, что все вероятности должны быть равны  $1/2$ , поскольку игрок  $Y$  будет проигрывать по 20 долларов слишком часто. (Он будет терять 20 долларов в  $1/4$  всех партий, 10 долларов еще в  $1/4$  и выигрывать 10 долларов в половине партий, т. е. в среднем на одной партии будет терять 2,5 доллара.) Но чем сильнее игрок  $Y$  будет тяготеть к стратегии «двух рук», тем настойчивее игрок  $X$  будет придерживаться стратегии «одной руки».

Вся задача состоит в *отыскании равновесия*. Существует ли такая смешанная стратегия  $y_1$  и  $y_2$ , что, будучи надлежащим образом использованной игроком  $Y$ , она лишит  $X$  всяких преимуществ? Аналогичным образом может ли игрок  $X$  выбрать такие вероятности  $x_1$  и  $x_2$ , что для  $Y$  не будет смысла избирать свою стратегию? В такой точке равновесия, если она вообще существует, средний выигрыш игрока  $X$  достигает *седловой точки*: он является максимальным для  $X$  и минимальным для  $Y$ . Отыскать такую точку — значит «решить» игру.

Точку зрения игрока  $X$  можно изложить следующим образом: он формирует новый столбец из соответствующих столбцов матрицы, взятых с весами  $x_1$  и  $1 - x_1$ . Пусть, например, он выбрал веса  $3/5$  и  $2/5$ . Тогда столбец будет иметь вид

$$\frac{3}{5} \begin{bmatrix} -10 \\ 10 \end{bmatrix} + \frac{2}{5} \begin{bmatrix} 20 \\ -10 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}.$$

**При таком выборе игрок  $Y$  будет терять 2 доллара независимо от своей стратегии.** Плата по-прежнему равняется 10 и 20 долларам. Но если игрок  $Y$  будет упорно подни-

мать одну руку, то он будет выигрывать по 10 долларов в течение  $3/5$  всего времени и проигрывать по 20 долларов в течение  $2/5$  всего времени, что дает средний проигрыш в размере 2 долларов на партию. Если  $Y$  будет поднимать обе руки или выберет смешанную стратегию в любой пропорции, он все равно будет терять по 2 доллара.

Но это не означает, что для игрока  $Y$  все стратегии являются оптимальными! Если он ленив и будет продолжать поднимать одну руку, то  $X$  изменит стратегию и начнет выигрывать по 20 долларов. Тогда  $Y$  изменит стратегию, и то же самое делает  $X$ . В конце концов, если оба игрока являются достаточно опытными, то как  $Y$ , так и  $X$  придут к оптимальной смешанной стратегии. Это означает, что игрок  $Y$  будет комбинировать строки матрицы с весами  $y_1$  и  $1 - y_1$ , пытаясь получить новую строку, которая была бы как можно меньшей:

$$y_1[-10 \ 20] + (1 - y_1)[10 \ -10] = [10 - 20y_1 \ -10 + 30y_1].$$

В данном случае наилучшая стратегия состоит в том, чтобы сделать обе компоненты равными:  $10 - 20y_1 = -10 + 30y_1$ , откуда  $y_1 = 2/5$ . При таком выборе обе компоненты равняются 2 и новая строка имеет вид  $[2 \ 2]$ . Поэтому *при такой стратегии игрок  $Y$  не может терять больше, чем 2 доллара*. Игрок  $Y$  минимизировал свой максимальный проигрыш, и этот минимакс (2 доллара) совпадает с максимумом, найденным независимо от него игроком  $X$ . *Цена игры* и равняется этому минимаксу или максимуму, т. е. 2 долларам.

Эта седловая точка замечательна тем, что в ней игрок  $X$  пользуется своей второй стратегией лишь на протяжении  $2/5$  всего времени, хотя она дает ему возможность выигрывать 20 долларов. В то же время игрок  $Y$  вынужден принять проигрышную стратегию — он должен был бы повторять действия игрока  $X$ , но вместо этого он играет с противоположными вероятностями  $2/5$  и  $3/5$ . Можно проверить, что  $X$  выигрывает 10 долларов с частотой  $3/5 \cdot 3/5 = 9/25$ , 20 долларов с частотой  $2/5 \cdot 2/5 = 4/25$  и проигрывает 10 долларов с оставшейся частотой  $12/25$ . Как и следовало ожидать, это дает ему средний выигрыш в размере 2 долларов.

Укажем на одну ошибку, которую легко можно сделать. Оптимальная комбинация строк не всегда дает строку с равными компонентами. Пусть, например, игрок  $X$  может, кроме того, поднимать вверх три руки и выигрывать при этом 60 долларов, если  $Y$  поднимает одну руку, и 80 долларов, если  $Y$  поднимает две руки. Платежная матрица при этом будет иметь вид

$$A = \begin{bmatrix} -10 & 20 & 60 \\ 10 & -10 & 80 \end{bmatrix}.$$

Игрок  $X$  все время выбирает эту новую стратегию, т. е. веса при столбцах фиксированы и равны  $x_1=0$ ,  $x_2=0$ ,  $x_3=1$ , и его минимальный выигрыш при этом составляет 60 долларов. Игрок  $Y$  подбирает такую комбинацию строк, которая дает строку с минимальными компонентами, т. е. выбирает первую строку и его максимальные потери равны 60 долларам. Поэтому мы также получим равенство минимакса и максимина, но седловая точка при этом будет сдвинута в угол.

Правило состоит в том, что при оптимальной комбинации строк для игрока  $Y$  цена игры оказывается (как было для двух и для шестидесяти долларов) расположенной только в столбцах, которые фактически используются игроком  $X$ . Аналогичным образом при оптимальной комбинации столбцов для игрока  $X$  то же самое значение находится в тех строках, которые включаются в наилучшую стратегию игрока  $Y$ , — другие строки дают более высокое значение и  $Y$  избегает их. Это правило в точности соответствует условию совместности невязок в линейном программировании.

**Упражнение 8.5.1.** Как изменятся оптимальные стратегии, если ставка будет увеличена с 20 до 70 долларов, и чему равняется новая цена игры (средний выигрыш игрока  $X$ )?

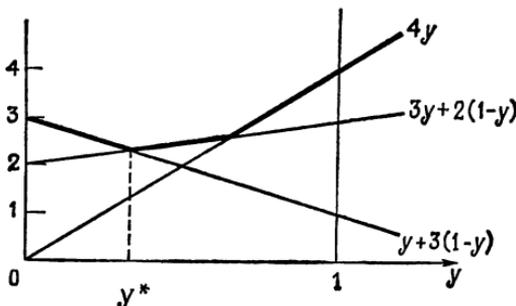
**Упражнение 8.5.2.** Пусть платежная матрица имеет вид  $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ . Объяснить, как игрок  $X$  подсчитывает свой максимин и как  $Y$  подсчитывает свой минимакс. Каковы оптимальные стратегии  $x^*$  и  $y^*$ ?

**Упражнение 8.5.3.** Пусть некоторый  $a_{ij}$  является наибольшим элементом в строке и наименьшим в столбце. Объяснить, почему игрок  $X$  всегда выбирает столбец  $j$ , а игрок  $Y$  всегда выбирает строку  $i$  (независимо от того, каковы остальные элементы матрицы)? Показать, что в предыдущем упражнении матрица содержит такой элемент, и построить матрицу, не содержащую его.

**Упражнение 8.5.4.** Найти наилучшую стратегию для игрока  $Y$ , если

$$A = \begin{bmatrix} 3 & 4 & 1 \\ 2 & 0 & 3 \end{bmatrix},$$

путем комбинирования строк с весами  $y$  и  $1-y$  и нарисовать все три компоненты.



Игрок  $X$  будет придерживаться наибольшей компоненты (жирная линия), а  $Y$  должен минимизировать этот максимум. Чему равен  $y^*$  и какова высота этого минимакса (пунктирная линия)?

**Упражнение 8.5.5.** При той же самой матрице  $A$  найти наилучшую стратегию для игрока  $X$  и показать, что он использует только два столбца (первый и третий) и соответствующие прямые пересекаются в точке минимакса.

**Упражнение 8.5.6.** Найти обе оптимальные стратегии и цену игры, если

$$A = \begin{bmatrix} 1 & 0 & -1 \\ -2 & -1 & 2 \end{bmatrix}.$$

**Упражнение 8.5.7.** Пусть  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ . Какие веса  $x_1$  и  $1 - x_1$  дадут столбец вида  $\begin{bmatrix} u \\ u \end{bmatrix}$  и какие веса  $y_1$  и  $1 - y_1$  для двух строк дадут новую строку  $[v \ v]$ ? Показать, что  $u = v$ .

#### ТЕОРЕМА О МИНИМАКСЕ

Наиболее общая матричная игра выглядит в точности так же, как и наш простой пример, за исключением одного важного пункта: игрок  $X$  имеет выбор из  $n$  возможностей, а  $Y$  — из  $m$ . Поэтому матрица платежа  $A$ , остающаяся неизменной в каждой партии, может иметь  $m$  строк и  $n$  столбцов. Каждый ее элемент  $a_{ij}$  представляет собой выигрыш игрока  $X$  при выборе им  $j$ -й стратегии, в то время как игрок  $Y$  избрал  $i$ -ю стратегию. Отрицательный элемент просто означает отрицательную плату, т. е. выигрыш для  $Y$ . Результат все равно остается *игрой двух лиц, имеющей нулевую сумму*: проигрыш одного игрока означает выигрыш другого. Но теперь существование равновесной седловой точки далеко не очевидно.

Как и в приведенном примере, игрок  $X$  может выбрать любую смешанную стратегию  $x = (x_1, \dots, x_n)$ . Эта стратегия всегда является вектором вероятности: величины  $x_i$  неотрицательны и их сумма равна 1. Эти компоненты дают частоты для  $n$  различных чистых стратегий, и в каждой партии игрок  $X$  выбирает их при помощи некоторого рандомизатора — устройства, генерирующего  $i$ -ю стратегию с частотой  $x_i$ . Одновременно с этим игрок  $Y$  должен принять аналогичное решение: он выбирает вектор-строку  $y = (y_1, \dots, y_m)$ , удовлетворяющий условиям  $y_i \geq 0$  и  $\sum y_i = 1$ , который дает частоты его собственной смешанной стратегии.

Предсказать результат одной партии невозможно, поскольку он является случайным. В среднем, однако, комбинация  $j$ -й стратегии для игрока  $X$  с  $i$ -й стратегией для  $Y$  будет иметь вероятность  $x_j y_i$ , равную произведению двух отдельных вероятностей. Если она реализуется, то выигрыш будет равен  $a_{ij}$ . Следовательно,

ожидаемый от этой комбинации выигрыш для игрока  $X$  равен  $a_{ij}x_jy_i$  и *ожидаемый общий выигрыш партии равен*  $\sum \sum a_{ij}x_jy_i$ . Подчеркнем еще раз, что любой (в том числе и все) элемент  $a_{ij}$  может быть отрицательным, правила одинаковы для обоих игроков и выигрыш зависит от элементов  $a_{ij}$ .

Ожидаемый выигрыш имеет очень простую матричную запись, поскольку двойная сумма  $\sum \sum a_{ij}x_jy_i$  равна  $yAx$  в силу правила умножения матриц:

$$yAx = [y_1 \dots y_m] \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = a_{11}x_1y_1 + \dots + a_{mn}x_ny_m.$$

Именно эту величину  $yAx$  игрок  $X$  пытается максимизировать, а игрок  $Y$  — минимизировать.

**Пример 2.** Пусть  $A$  есть единичная матрица порядка  $n$ ,  $A = I$ . Тогда ожидаемый выигрыш равен  $xIy = x_1y_1 + \dots + x_ny_n$  и идею игры легко объяснить: игрок  $X$  пытается сделать тот же выбор, что и  $Y$ , поскольку в этом случае он получает  $a_{ii} = 1$  доллар. В то же самое время игрок  $Y$  пытается сделать другой выбор, чтобы не дать игроку  $X$  выиграть и не платить: вне диагонали единичной матрицы стоят нули, и если  $X$  выбирает  $i$ -й столбец, то  $Y$  выбирает другую строку  $j$ . Если игрок  $X$  выбирает одну из стратегий чаще, чем любую другую, то  $Y$  чаще может уйти от поражения. Поэтому оптимальной стратегией будет  $x^* = (1/n, 1/n, \dots, 1/n)$ . Аналогично игрок  $Y$  не может делать упор на какую-либо определенную стратегию, иначе  $X$  получит преимущество, и поэтому его оптимальный выбор также состоит в выборе равных вероятностей, сумма которых равняется 1,  $y^* = (1/n, 1/n, \dots, 1/n)$ . Вероятность того, что оба игрока изберут  $i$ -ю стратегию, равна  $(1/n)^2$ , и сумма по всем таким комбинациям дает ожидаемый выигрыш для  $X$ . Общая цена игры есть  $n \times (1/n)^2$ , или  $1/n$ , что подтверждается вычислением:

$$\begin{aligned} y^*Ax^* &= [1/n \dots 1/n] \begin{bmatrix} 1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ & & & & 1 \end{bmatrix} \begin{bmatrix} 1/n \\ \cdot \\ \cdot \\ \cdot \\ 1/n \end{bmatrix} = \\ &= \left(\frac{1}{n}\right)^2 + \dots + \left(\frac{1}{n}\right)^2 = \frac{1}{n}. \end{aligned}$$

Когда число  $n$  возрастает, игрок  $Y$  получает больший шанс на «спасение».

Заметим, что симметричность матрицы не гарантирует справедливости игры для обоих игроков. На самом деле верно прямо противоположное: игра будет совершенно справедливой, когда матрица является *кососимметрической*, т. е.  $A^T = -A$ . При такой матрице оба игрока находятся в равных условиях, поскольку выбор игроком  $X$   $j$ -й стратегии при выборе  $i$ -й стратегии игроком  $Y$  дает выигрыш в размере  $a_{ij}$  игроку  $X$ , а выбор стратегии  $j$  игроком  $Y$  и стратегии  $i$  игроком  $X$  дает тот же выигрыш  $a_{ij}$  игроку  $Y$  (поскольку  $a_{ji} = -a_{ij}$ ). Оптимальные стратегии  $x^*$  и  $y^*$  должны совпадать, и ожидаемый выигрыш должен равняться  $y^*Ax^* = 0$ . Цена игры при  $A^T = -A$  равняется нулю.

### Пример 3.

$$A = \begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix}.$$

Игра состоит в том, что оба игрока  $X$  и  $Y$  выбирают число между 1 и 3, и тот, кто выбрал меньшее, выигрывает 1 доллар. (Если игрок  $X$  выбрал число 2 и  $Y$  выбрал 3, то выигрыш равен  $a_{23} = 1$  доллару. Если же они выбирают одно и то же число, то мы будем находиться на главной диагонали матрицы и ни один из игроков не выигрывает.) Очевидно, что ни один из игроков не выберет число 2 или 3, поскольку другой игрок может выбрать меньшее. Следовательно, чистые стратегии  $x^* = y^* = (1, 0, 0)$  являются оптимальными — оба игрока каждый раз выбирают единицу — и цена игры равна  $y^*Ax^* = a_{11} = 0$ .

Следует отметить, что матрица, оставляющая все решения неизменными, является не единичной матрицей, а матрицей  $E$ , у которой *каждый* элемент  $e_{ij}$  равен единице. Добавление кратного матрицы  $E$  к платежной матрице,  $A \rightarrow A + \alpha E$ , просто означает, что игрок  $X$  будет выигрывать дополнительную сумму  $\alpha$  в каждой партии. Цена игры увеличивается на  $\alpha$ , но  $x^*$  и  $y^*$  остаются прежними.

Теперь мы возвращаемся к общей теории, приняв для начала точку зрения игрока  $X$ . Пусть он выбирает смешанную стратегию  $x = (x_1, \dots, x_n)$ . Тогда игрок  $Y$  постепенно поймет эту стратегию и выберет свою стратегию таким образом, чтобы минимизировать потери  $yAx$ . В результате игрок  $X$  получит  $\min yAx$ .

Опытный игрок  $X$  выберет такой вектор  $x^*$  (быть может, не единственный), который *максимизирует этот минимум*. Сделав этот

выбор, игрок  $X$  может быть уверен в том, что его выигрыш окажется не меньше, чем

$$\min_y yAx^* = \max_x \min_y yAx. \quad (14)$$

На большее он рассчитывать не может.

Игрок  $Y$  поступает противоположным образом. При любой из своих собственных стратегий  $y$  он должен ожидать, что игрок  $X$  найдет вектор, максимизирующий  $yAx$ <sup>1)</sup>. Поэтому  $Y$  должен выбрать смешанную стратегию  $y^*$ , минимизирующую этот максимум и гарантирующую, что он потеряет не больше, чем

$$\max_x y^*Ax = \min_y \max_x yAx. \quad (15)$$

Ничего лучшего игроку  $Y$  не придумать.

Надеюсь вы уже догадались, каков будет результат, если все это вообще верно. Мы хотим, чтобы величина (14) гарантированного выигрыша игрока  $X$  совпадала с величиной потерь (15) игрока  $Y$ . Тогда смешанные стратегии  $x^*$  и  $y^*$  дадут равновесную седловую точку и игра будет решена: при любых стратегиях, не совпадающих с  $x^*$  и  $y^*$ , оба игрока будут лишь ухудшать свое положение. Существование этой седловой точки было доказано фон Нейманом и носит название *теоремы о минимаксе*:

**8N.** Для произвольной матрицы  $A$  размера  $m \times n$  минимакс по всем смешанным стратегиям совпадает с максимином:

$$\max_x \min_y yAx = \min_y \max_x yAx. \quad (16)$$

Эта величина является ценой игры. Если максимум в левой части достигается в точке  $x^*$ , а минимум в правой части — в точке  $y^*$ , то эти стратегии являются оптимальными и они дают седловую точку, покидать которую не имеет смысла:

$$y^*Ax \leq y^*Ax^* \leq yAx^* \quad \text{для всех } x \text{ и } y. \quad (17)$$

В седловой точке вектор  $x^*$  во всяком случае «не хуже» любого другого вектора  $x$  (поскольку  $y^*Ax \leq y^*Ax^*$ ). Аналогично второй игрок будет лишь проигрывать больше, если он отойдет от стратегии  $y^*$ .

Так же как и в теории двойственности, мы можем начать с одностороннего неравенства  $\max \min \leq \min \max$ . Это просто

<sup>1)</sup> Этот вектор может не совпадать с  $x^*$ . Если игрок  $Y$  выбирает нерациональную стратегию, то  $X$  может получить больше, чем ему гарантирует формула (14). В теории игр мы предполагаем, что оба игрока являются профессионалами.

комбинация определений (14) для  $x^*$  и (15) для  $y^*$ :

$$\max_x \min_y yAx = \min_y \max_x yAx^* \leq y^* Ax^* \leq \max_x y^* Ax = \min_y \max_x yAx. \quad (18)$$

Это означает, что если игрок  $X$  с гарантией выигрывает не меньше, чем  $\alpha$ , а  $Y$  с гарантией теряет не больше, чем  $\beta$ , то обязательно  $\alpha \leq \beta$ . Заслугой фон Неймана было то, что он доказал равенство  $\alpha = \beta$ . Это и есть теорема о минимаксе. Она означает, что в (18) везде должны стоять равенства, а свойство седловой точки (17) выводится из этого в упр. 8.4.6.

Самое замечательное здесь то, что при доказательстве используется в точности такая же техника, как и в теории линейного программирования. Интуитивно это почти очевидно: игроки  $X$  и  $Y$  играют «двойственные» роли и оба выбирают стратегии из «допустимого множества» векторов вероятности  $x_i \geq 0$ ,  $\sum x_i = 1$ ,  $y_i \geq 0$ ,  $\sum y_i = 1$ . Удивительно, что даже фон Нейман не сразу заметил, что обе теории совпадают. Он доказал теорему о минимаксе в 1928 году, линейное программирование появилось около 1947 года, а первое доказательство теоремы двойственности было дано Гейлом, Куном и Такером в 1951 году — и оно опиралось на работы фон Неймана! Их доказательство было опубликовано в той же книге, в которой Данциг показал эквивалентность линейных программ и матричных игр, так что на самом деле мы обращаем реальную историческую последовательность событий, выводя теорему о минимаксе из двойственности.

В общих чертах теорему о минимаксе можно доказать следующим образом. Пусть  $b = (1, \dots, 1)^T$  — вектор-столбец длины  $m$  и  $c = (1, \dots, 1)$  — вектор-строка длины  $n$ . Рассмотрим двойственные линейные программы

$$(P) \quad \begin{aligned} &\text{минимизировать } cx \\ &\text{при условиях } Ax \geq b, \\ &x \geq 0 \end{aligned}$$

$$(D) \quad \begin{aligned} &\text{максимизировать } yb \\ &\text{при условиях } yA \leq c, \\ &y \geq 0. \end{aligned}$$

Для применения двойственности мы должны быть уверены, что обе задачи являются допустимыми, так что при необходимости мы добавляем ко всем элементам матрицы  $A$  одно и то же большое число  $\alpha$ . Это не может изменить оптимальные стратегии в игре, поскольку каждая выплата увеличивается на  $\alpha$  и то же самое происходит с величинами  $\min \max$  и  $\max \min$ . Для измененной матрицы, которую мы по-прежнему обозначаем через  $A$ , вектор  $y = 0$  является допустимым вектором двойственной задачи и любой достаточно большой вектор  $x$  будет допустимым для исходной задачи.

Теперь из теоремы двойственности для линейного программирования следует, что существуют векторы  $x^*$  и  $y^*$ , удовлетворяющие равенству  $cx^* = y^*b$ . Поскольку  $b$  и  $c$  состоят из единиц,

получаем  $\sum x_i^* = \sum y_i^*$ . Если эти суммы равняются  $\theta$ , то деление на  $\theta$  делает эти суммы равными 1, и в результате смешанные стратегии  $x^*/\theta$  и  $y^*/\theta$  являются оптимальными в теории игр. Для любых других смешанных стратегий  $x$  и  $y$

$$Ax^* \geq b \quad \text{влечет за собой} \quad yAx^* \geq yb = 1$$

и

$$y^*A \leq c \quad \text{влечет за собой} \quad y^*Ax \leq cy = 1.$$

Отметим, что  $y^*Ax \leq 1 \leq yAx^*$ . Это означает, если мы разделим неравенство на  $\theta$ , что игрок  $X$  не может выиграть больше, чем  $1/\theta$ , против стратегии  $y^*/\theta$  и что игрок  $Y$  не может проиграть меньше, чем  $1/\theta$ , против стратегии  $x^*/\theta$ . Следовательно, именно эти стратегии и нужно выбрать, так что  $\min \max = \max \min = 1/\theta$ .

Это завершает теорию. Но мы не дали ответа на самый естественный вопрос: какие игры в обычном смысле слова эквивалентны нашим «матричным играм»? Например, подпадают ли под теорию фон Неймана шахматы, бридж, покер?

На мой взгляд, игра в шахматы не подходит по двум причинам. Во-первых, стратегия белых фигур состоит не просто из движения вперед. Она должна включать решение вопроса о том, как ответить на первый ход черных фигур, а затем — как ответить на их второй ход и т. д. до конца игры. На каждом шаге количество возможностей таково, что игрок  $X$  имеет миллиарды чистых стратегий и то же самое справедливо для его противника. Поэтому числа  $m$  и  $n$  являются недопустимо большими. Более того, случай здесь практически не играет роли. Если белые смогут найти выигрышную стратегию или если черные смогут найти стратегию защиты — что до сих пор не было сделано, — то игре в шахматы придет конец. Разумеется, в нее будут продолжать играть, как в крестики-нолики, но интерес сильно поубавится.

В отличие от шахмат, в бридже содержится некоторый элемент хитрости, заключающийся в некоторых специальных ухищрениях. Поэтому его можно рассматривать как матричную игру. К сожалению, числа  $m$  и  $n$  вновь оказываются фантастически большими, но, возможно, отдельный элемент игры и может быть проанализирован на оптимальную стратегию.

С другой стороны, «Черный Джек»<sup>1)</sup> не является матричной игрой (по крайней мере вариант, употребляемый в казино), поскольку в заведении действуют определенные правила. Когда мой приятель Эд Торп нашел и опубликовал выигрышную стратегию (что привело к изменению правил игры в Лас-Вегасе), его метод был основан на прослеживании перемещения карт, уже открытых в процессе игры. Элемент случайности полностью отсутствовал, и поэтому никакой смешанной стратегии  $x^*$  не существовало.

<sup>1)</sup> Карточная игра. — Прим. перев.

Существует также так называемая **дилемма заключенного**, состоящая в следующем. Двое сообщников арестованы, и перед каждым из них стоит дилемма: если он сознается, то его освободят, но при условии, что другой не сознается (и получит при этом 10 лет). Если сознаются оба, то они получают по 6 лет каждый. Но если ни один из них не сознается, то им можно дать лишь по 2 года, поскольку доказанное преступление не является тяжелым. Что делать в этой ситуации? Желание сознаться очень велико, но если они могут положиться друг на друга, то они выкарабкаются. Эта игра имеет ненулевую сумму: потерять могут оба.

Отличным примером матричной игры является *покер*. Блефование является существенной частью хорошего покера, и для эффективности оно должно быть непредсказуемым. Это означает, что решение блефовать должно приниматься совершенно случайным образом, если мы принимаем фундаментальное предположение теории игр о том, что противник является опытным, иначе он обнаружит систему и будет выигрывать. Вероятности за и против блефования будут зависеть от известных карт и от размеров ставок — как от сделанных, так и от тех, которые предстоит сделать. Фактически число вариантов вновь оказывается таким, что отыскание наилучшей возможной стратегии  $x^*$  совершенно нереально. Тем не менее хороший игрок в покер должен играть очень близко к  $x^*$ . Эта стратегия может быть вычислена точно, если мы предельно упростим игру:

Игрок  $X$  получает при сдаче валета или короля с равными вероятностями, в то время как игрок  $Y$  всегда получает даму. Посмотрев на свою карту, игрок  $X$  может либо бросить ее и проиграть поставленный доллар, либо поставить еще два доллара. Если  $X$  делает ставку, то  $Y$  может либо бросить свою карту и проиграть поставленный доллар, либо ответить двумя долларами и увидеть, открыв карты, блефовал игрок  $X$  или нет. В итоге старшая карта выиграет 3 доллара, поставленные противником.

Даже в этой простой игре возможные «чистые стратегии» не так уж легко найти, но сделать это будет полезно. Игрок  $Y$  имеет только две возможности, так как он лишь отвечает на действия игрока  $X$ :

- 1) если  $X$  ставит, то  $Y$  бросает свою карту,
- 2) если  $X$  ставит, то  $Y$  отвечает и идет на 3 доллара.

Игрок  $X$  может использовать четыре стратегии, некоторые из них разумные, а некоторые — нет:

- 1) он ставит 2 доллара, имея на руках короля, и бросает карту, имея валета,
- 2) он ставит в любом случае (блефует),

3) он бросает карту в любом случае и проигрывает 1 доллар,

4) он бросает карту, имея на руках короля, и ставит, имея валаета.

Составление платежной матрицы требует некоторого терпения:

$a_{11} = 0$ , так как игрок  $X$  теряет один доллар в половине случаев на валете и выигрывает на короле ( $Y$  бросает карту).

$a_{21} = 1$ , так как игрок  $X$  теряет один доллар в половине случаев и выигрывает 3 доллара также в половине случаев (поскольку  $Y$  пытается обыграть его, даже если  $X$  имеет короля).

$a_{12} = 1$ , так как игрок  $X$  ставит, а  $Y$  бросает карту (блеф удаётся).

$a_{22} = 0$ , так как игрок  $X$  выигрывает 3 доллара на короле и проигрывает 3 доллара на валете (блеф срывается).

При третьей стратегии всегда теряется доллар, четвертая также не приводит к успеху. В результате матрица  $A$  будет иметь вид

$$A = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 1 & 0 & -1 & -2 \end{bmatrix}.$$

Оптимальной стратегией для игрока  $X$  в этой игре будет блефование в половине случаев,  $x^* = (1/2, 1/2, 0, 0)$ , а подчинившийся игрок  $Y$  должен выбрать  $y^* = (1/2, 1/2)$ . Цена игры равна 50 центам.

Странно заканчивать книгу таким образом, обучая упрощенному варианту игры в покер, но я думаю, что и покер должен иметь свое место в линейной алгебре и ее приложениях.

**Упражнение 8.5.8.** Найти  $x^*$ ,  $y^*$  и цену игры  $v$  для матрицы

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

**Упражнение 8.5.9.** Вычислить  $\min_{y_i \geq 0} \max_{x_i \geq 0} (x_1 y_1 + x_2 y_2)$ .  
 $y_1 + y_2 = 1$   $x_1 + x_2 = 1$

**Упражнение 8.5.10.** Объяснить каждое из неравенств в (18). Затем, после того как теорема о минимаксе превращает их в равенства, объяснить соотношения для седловой точки (17).

**Упражнение 8.5.11.** Показать, что  $x^* = (1/2, 1/2, 0, 0)$  и  $y^* = (1/2, 1/2)$  являются оптимальными стратегиями при игре в покер. Для этого вычислить  $y A x^*$  и  $y^* A x$  и проверить выполнение условий для седловой точки (17).

**Упражнение 8.5.12.** Было ли доказано, что в шахматах не может существовать стратегии, всегда приносящей выигрыш черным? Я думаю, что это известно только для случая, когда игроки могут делать по два хода подряд. Если бы черные имели выигрышную стратегию, то, поскольку белые могли бы двигать коня вперед и назад, мы пришли бы к тому, что выигрывают оба.

## Приложение А

# ЛИНЕЙНЫЕ ПРЕОБРАЗОВАНИЯ, МАТРИЦЫ И ЗАМЕНЫ БАЗИСОВ

Пусть задана матрица  $A$  размера  $m \times n$  и  $n$ -мерный вектор  $x$ . Вектор  $x$  лежит в пространстве  $\mathbf{R}^n$ , а произведение  $Ax$  — в  $\mathbf{R}^m$ . Отображение  $x$  в  $Ax$  задает преобразование пространства  $\mathbf{R}^n$  в  $\mathbf{R}^m$ , и правила матричного умножения немедленно приводят к его наиболее важному свойству: для любых векторов  $x$  и  $y$  и скалярных величин  $b$  и  $c$  справедливо соотношение

$$A(bx + cy) = b(Ax) + c(Ay). \quad (1)$$

Преобразование, обладающее свойством (1), называется *линейным*, и, следовательно, любая матрица  $A$  задает линейное преобразование.

Целью этого дополнения является изучение других примеров линейных преобразований с тем, чтобы выработать общий взгляд на них. Эта цель лежит также в основе более абстрактного подхода к предмету линейной алгебры, когда мы берем в качестве исходного свойство (1) и выводим следствия из него; мы же предпочли начать изложение с матриц. Но теперь пора отойти от исследования линейных алгебраических систем  $Ax = b$  и их матриц и рассмотреть другие примеры линейных преобразований.

Прежде всего мы не должны забывать о существовании векторных пространств, отличных от  $\mathbf{R}^n$  и  $\mathbf{C}^n$ . Для того чтобы некоторое множество было векторным пространством, требуется лишь подходящий способ образования комбинаций вида  $bx + cy$  и уверенность в том, что любая подобная комбинация лежит в этом пространстве.

**Пример 1.** Пусть  $P_n$  есть пространство многочленов степени  $n$ ,  $p = a_0 + a_1t + \dots + a_nt^n$ . Эти многочлены являются «векторами», и сумма двух таких векторов, безусловно, принадлежит  $P_n$ . Многочлены специального вида,  $p_0 = 1$ ,  $p_1 = t$ ,  $\dots$ ,  $p_n = t^n$ , образуют базис в этом пространстве, и размерность его равна  $n + 1$ .

**Пример 2.** Если задано линейное дифференциальное уравнение порядка  $n$  с постоянными коэффициентами

$$\frac{d^n u}{dt^n} + c_1 \frac{d^{n-1} u}{dt^{n-1}} + \dots + c_n u = 0, \quad (2)$$

то его решения  $u(x)$  образуют векторное пространство  $S_n$ . Для любых двух решений  $u$  и  $v$  из закона суперпозиции следует, что  $bu + cv$  также является решением. Другими словами, решения можно складывать и умножать на скалярные величины. Простейшие решения являются чисто экспоненциальными,  $u = e^{\lambda t}$ ; возможны также решения вида  $te^{\lambda t}$ , если  $\lambda$  является корнем кратности два характеристического уравнения, вида  $t^2 e^{\lambda t}$ , если  $\lambda$  является корнем кратности три, и т. д. В целом имеется  $n$  решений такого вида и они образуют базис пространства решений<sup>1)</sup>.

**Пример 3.** Пусть  $H$  есть гильбертово пространство, введенное на стр. 159. Оно состоит из всех векторов  $v = (v_1, v_2, v_3, \dots)$ , имеющих бесконечное число компонент, но конечную длину:

$$\|v\|^2 = v_1^2 + v_2^2 + v_3^2 + \dots < \infty.$$

Разумеется, в этом случае конечный базис отсутствует ( $H$  является бесконечномерным), поскольку для определения  $v$  требуется бесконечное количество «элементов информации».

Используя эти примеры векторных пространств, нам будет легче проиллюстрировать идею линейного преобразования. Напомним, что основным свойством является линейность; преобразование  $A$  может отображать одно векторное пространство в любое другое (или в себя), но оно должно удовлетворять соотношению  $A(bx + cy) = b(Ax) + c(Ay)$ .

**Пример 4.** Пусть  $A$  есть оператор дифференцирования на пространстве многочленов степени  $n$ :  $A = d/dt$ . Тогда для любого многочлена  $p$  справедливо соотношение

$$Ap = A(a_0 + a_1 t + \dots + a_n t^n) = a_1 + 2a_2 t + \dots + na_n t^{n-1}. \quad (3)$$

Мы можем рассматривать  $A$  либо как линейное преобразование пространства  $P_n$  в пространство  $P_{n-1}$  меньшей размерности, либо как преобразование из  $P_n$  в  $P_n$ . В первом случае это будет преобразование на пространство  $P_{n-1}$  и каждый многочлен из  $P_{n-1}$  может быть получен путем дифференцирования некоторого многочлена из  $P_n$ . Во втором случае это не так, поскольку в пространстве  $P_n$  существуют многочлены (например,  $t^n$ ), которые не могут присутствовать в правой части (3); если найти «матрицу дифференцирования», то она не будет обратимой.

<sup>1)</sup> Если коэффициенты  $c_i$  зависят от  $t$ , то  $S_n$  по-прежнему будет  $n$ -мерным векторным пространством. Но отыскать его базис или даже отдельный элемент значительно труднее.

**Пример 5.** Пусть оператор  $A$ , действующий на элементы пространства  $P_{n-1}$ , есть оператор *интегрирования*:

$$A(a_0 + a_1 t + \dots + a_{n-1} t^{n-1}) = a_0 t + a_1 \frac{t^2}{2} + \dots + a_{n-1} \frac{t^n}{n}. \quad (4)$$

Эта формула определяет преобразование из  $P_{n-1}$  в более широкое пространство  $P_n$ , и оно вновь является линейным:

$$\int_0^t [bp(t) + cq(t)] dt = b \int_0^t p(t) dt + c \int_0^t q(t) dt.$$

Можно ли при помощи интегрирования получить все многочлены пространства  $P_n$ , т. е. является ли процедура интегрирования отображением «на»? Нет, хотя это отображение является взаимно однозначным, т. е. различные многочлены имеют различные интегралы.

**Пример 6.** Выше мы обозначили через  $S_n$  пространство всех решений дифференциального уравнения (2). Пусть  $x$  есть произвольный вектор из  $\mathbf{R}^n$  и его компоненты являются начальными значениями решения  $u$ , т. е. при  $t=0$  имеем

$$u(0) = x_1, \quad \frac{du}{dt}(0) = x_2, \quad \dots, \quad \frac{d^{n-1}u}{dt^{n-1}}(0) = x_n.$$

Задавая эти начальные значения, мы выбираем одно определенное решение дифференциального уравнения. Такое соответствие между вектором  $x$  из  $\mathbf{R}^n$  и функцией  $u$  из  $S_n$  является линейным преобразованием между двумя пространствами: решение линейного уравнения линейно зависит от начальных значений. Это преобразование является одновременно взаимно однозначным и преобразованием «на»: различные начальные значения должны давать различные решения и любое решение может быть получено при надлежащем выборе начальных значений. Следовательно, матрица порядка  $n$ , описывающая это преобразование, будет обратима.

**Пример 7.** Пусть  $A$ —оператор *левого сдвига* на гильбертовом пространстве  $H$ :

$$Av = A(v_1, v_2, v_3, \dots) = (v_2, v_3, v_4, \dots).$$

Это отображение является линейным и «на»; при надлежащем выборе входного вектора на выходе может быть получен любой вектор. Поскольку мы можем выбирать любой элемент  $v_1$ , оператор не будет взаимно однозначным;  $v_1$  не присутствует на выходе. Он выбрасывается, и в результате оператор  $A$  имеет одномерное нуль-пространство:  $A(v_1, 0, 0, \dots) = (0, 0, \dots)$ .

В этих примерах проверка свойства линейности тривиальна; если линейность имеется, то не заметить ее практически невоз-

можно. Тем не менее данное свойство является наиболее полезным свойством преобразования<sup>1)</sup>, и оно естественным образом возвращает нас к матрицам. Фактически главная цель этого приложения состоит в том, чтобы показать, что *каждое линейное преобразование может быть представлено в виде матрицы*. Эта матрица будет зависеть от базисов, выбираемых в векторных пространствах, и строится по следующему правилу:

Пусть в пространстве  $V$  задан базис  $v_1, \dots, v_n$ , а в пространстве  $W$  — базис  $w_1, \dots, w_m$ . Тогда любое линейное преобразование  $A$  из  $V$  в  $W$  представляется матрицей  $[A]$  размера  $m \times n$ . Ее элементы  $a_{ij}$  определяются путем применения преобразования  $A$  к каждому из элементов  $v_j$  и представления результата в виде комбинации элементов  $w$ :

$$Av_j = \sum_{i=1}^m a_{ij} w_i, \quad j = 1, \dots, n. \quad (5)$$

Другими словами,  $j$ -й столбец матрицы определяется тем, как изменяется базисный вектор  $v_j$  под действием преобразования  $A$ .

**Пример 8.** Пусть  $A$  есть оператор интегрирования, а в качестве базиса выбраны многочлены  $1, t, t^2, \dots$ : элементы  $v_j$  и  $w_j$  равны  $t^{j-1}$ . Для вычисления матрицы  $[A]$  нужно посмотреть, как  $A$  действует на  $v_j$ :

$$Av_j = \int_0^t t^{j-1} dt = \frac{t^j}{j} = \frac{1}{j} w_{j+1}.$$

Единственным ненулевым элементом в  $j$ -м столбце матрицы является  $a_{ij}$ , равный  $1/j$ . Если пространство  $V$  имеет размерность  $n=4$ , то  $W$  будет иметь размерность  $m=5$  и искомая матрица размера  $5 \times 4$  есть

$$[A] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{bmatrix}.$$

Пусть мы хотим проинтегрировать функцию  $v(t) = 2t + 8t^3$ . Тогда мы записываем  $v$  в виде

$$2t + 8t^3 = 0v_1 + 2v_2 + 0v_3 + 8v_4$$

<sup>1)</sup> На втором месте стоит, пожалуй, обратимость.

и умножаем соответствующие коэффициенты на  $A$ :

$$[A] \begin{bmatrix} 0 \\ 2 \\ 0 \\ 8 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 2 \end{bmatrix}, \quad \text{т. е.} \quad \int 2t + 8t^3 = w_3 + 2w_5 = t^2 + 2t^4.$$

**Пример 9.** Пусть  $A$  есть оператор дифференцирования, переводящий  $P_n$  в пространство  $P_{n-1}$  меньшей размерности. При том же выборе базиса действие  $A$  на  $v_j$  имеет вид

$$Av_j = \frac{d}{dt} t^{j-1} = (j-1)t^{j-2} = (j-1)w_{j-1}.$$

Поэтому  $j$ -й столбец матрицы  $[A]$  вновь содержит лишь один ненулевой элемент, равный  $j-1$  и находящийся в строке с номером  $i=j-1$ . Для размерностей 5 и 4 матрица имеет вид

$$[A] = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}.$$

*Замечание.* Мы рассматриваем дифференцирование и интегрирование как взаимно обратные операции. По крайней мере дифференцирование, следующее за интегрированием, приводит к исходной функции, т. е. вторая матрица является *левой обратной* к первой. Произведение двух матриц, приведенных выше, есть единичная матрица 4-го порядка. Но прямоугольная матрица не может иметь двухстороннюю обратную и произведение в другом порядке не является единичной матрицей:

$$[A]_{\text{integ}} [A]_{\text{diff}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ c & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Нулевой элемент в левом верхнем углу соответствует тому, что для константы операция дифференцирования с последующим интегрированием дает нуль.

**Пример 10.** Матрица, представляющая левый сдвиг в гильбертовом пространстве, является бесконечно большой, но ничем дру-

гим не примечательна:

$$[A] = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdot \\ 0 & 0 & 1 & 0 & \cdot \\ 0 & 0 & 0 & 1 & \cdot \\ 0 & 0 & 0 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \quad \text{и} \quad [A] \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ \vdots \end{bmatrix} = \begin{bmatrix} v_2 \\ v_3 \\ v_4 \\ v_5 \\ \vdots \end{bmatrix}.$$

Нам остается лишь проверить выполнение общего правила (5) при составлении матрицы и показать, что  $[A]$  содержит полную информацию о линейном преобразовании  $A$ , представляемом ей. Другими словами, зная базисы в пространствах  $V$  и  $W$  и зная матрицу  $[A]$ , мы должны получать результат воздействия преобразования  $A$  на произвольный вектор  $v = \sum_1^n x_j v_j$  пространства  $V$ .

*Если вектор  $x$  состоит из коэффициентов разложения  $v$  по базису  $v_1, \dots, v_n$ , то вектор  $y = [A]x$  дает коэффициенты разложения  $Av$  по базису  $w_1, \dots, w_m$ . Следовательно, результат воздействия преобразования  $A$  на любой элемент  $v$  восстанавливается путем умножения матриц:*

$$Av = \sum_{i=1}^m y_i w_i = \sum_{i,j} a_{ij} x_j w_i. \quad (6)$$

Доказательство основывается на линейности преобразования  $A$ : для произвольного  $v = \sum x_j v_j$  имеем

$$Av = A \left( \sum_1^n x_j v_j \right) = \sum_1^n x_j Av_j.$$

Подставляя выражение для  $Av_j$  из (5), получаем (6). Если задано действие преобразования  $A$  на каждый элемент базиса, то свойства линейности достаточно для определения действия  $A$  на любой другой вектор пространства.

Укажем на еще одну связь между линейными преобразованиями и матрицами. Если  $A$  и  $B$  являются линейными преобразованиями из  $V$  в  $W$  и из  $W$  в  $Z$  соответственно, то их произведение есть преобразование из  $V$  в  $Z$ ; оно начинается с элемента  $v$  из  $V$ , который переходит в  $Av$  из  $W$  и затем в  $BAv = B(Av)$  из  $Z$ . Последнее соотношение в точности совпадает с правилом (7) стр. 25, которое служило основой умножения матриц. Поэтому вполне естественно, что матричное умножение правильно отражает это произведение линейных преобразований:

Если матрицы  $[A]$  и  $[B]$  представляют линейные преобразования  $A$  и  $B$  в базисах  $\{v_i\}$  в  $V$ ,  $\{w_i\}$  в  $W$  и  $\{z_i\}$  в  $Z$ , то произведение этих двух матриц представляет преобразование  $BA$ :

$$[BA] = [B][A]. \quad (7)$$

Едва ли следует говорить о том, что  $BA$  вновь является линейным преобразованием, хотя оно и выглядит как квадратичное. Может показаться, что при  $A(x) = x$  и  $B(x) = x$  произведение  $BA$  будет связано с  $x^2$ , но на самом деле это, разумеется, не так и квадрат тождественного преобразования вновь является тождественным преобразованием:  $B(A(x))$  равняется  $x$ .

**Упражнение А.1.** Найти матрицу размера  $3 \times 5$ , представляющую оператор двукратного дифференцирования  $d^2/dt^2$ , переводящий пространство  $P_4$  в  $P_2$  (многочлены четвертого порядка в многочлены второго порядка).

**Упражнение А.2.** Для оператора левого сдвига из примера 10 показать, что оператор правого сдвига  $[A]^T$  является лишь односторонним обратным. Строки матрицы  $[A]$  являются ортонормированными, но столбцы таковыми не являются — ситуация невозможная для матриц размера  $n \times n$ .

**Упражнение А.3.** Пусть  $V = W$  есть пространства матриц размера  $2 \times 2$  с базисом

$$\begin{aligned} v_1 = w_1 &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, & v_2 = w_2 &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \\ v_3 = w_3 &= \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, & v_4 = w_4 &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

Найти преобразование  $A$ , переводящее каждую матрицу в транспонированную к ней, и, пользуясь правилом (5), найти матрицу четвертого порядка  $[A]$ . Почему  $[A]^2 = I$ ?

**Упражнение А.4.** В конце второй главы рассматривалось преобразование, состоящее в умножении слева на фиксированную матрицу  $\begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}$  каждого элемента указанного выше пространства  $V$ . Найти матрицу порядка 4, представляющую это преобразование на  $V$ .

**Упражнение А.5.** Пусть дифференциальное уравнение (2) имеет вид  $d^n u/dt^n = 0$ . Найдя базис, описать пространство  $S_n$  всех его решений.

**Упражнение А.6.** Пусть на плоскости  $xu$  преобразование  $A$  переводит каждый вектор в его зеркальное отражение относительно оси  $x$ , а преобразование  $B$  переводит каждый вектор в вектор той же длины, но направленный в противоположную сторону от начала координат. Найти матрицы  $[A]$ ,  $[B]$  и  $[BA]$ , пользуясь стандартными базисными векторами  $e_1$  и  $e_2$ , и описать, как изменится вектор после действия на него преобразования  $A$  и затем  $B$ .

## ЗАМЕНА БАЗИСА

Пусть  $v_1, \dots, v_n$  и  $w_1, \dots, w_n$  являются базисами в одном и том же векторном пространстве, и пусть некоторый вектор записывается сначала в одном базисе, а затем в другом

$$v = \sum_1^n x_j v_j = \sum_1^n y_i w_i. \quad (8)$$

Как связаны между собой эти наборы коэффициентов?

*Если два базиса связаны друг с другом соотношениями  $v_j = \sum_{i=1}^n s_{ij} w_i$ , то коэффициенты  $x$  и  $y$  связаны иным образом, но с теми же коэффициентами  $s_{ij}$ :*

$$y_i = \sum_{j=1}^n s_{ij} x_j. \quad (9)$$

Легче всего установить это, проверив, что новые коэффициенты  $y_i$  удовлетворяют условию  $\sum x_j v_j = \sum y_i w_i$ , так что они действительно являются коэффициентами в разложении вектора  $v$ :

$$\sum_{j=1}^n x_j v_j = \sum_{j=1}^n \sum_{i=1}^n x_j s_{ij} w_i \quad \text{равняется} \quad \sum_{i=1}^n y_i w_i = \sum_{i=1}^n \sum_{j=1}^n s_{ij} x_j w_i.$$

Замена базиса также относится к теории матриц, хотя и несколько менее очевидным образом. (Часто это представляют как замену переменных, что с математической точки зрения приводит к тому же результату.) Пусть мы определили преобразование  $A$  векторного пространства в себя и выбрали базис  $w_1, \dots, w_n$ , один и тот же для  $W$  и для  $V=W$ . Тогда уравнение (5) дает соответствующую матрицу  $[A]_w$  (с элементами  $a_{ij}$ ) по правилу  $A w_j = \sum a_{ij} w_i$ , и мы хотим узнать, как будет выглядеть матрица того же самого преобразования  $A$  при выборе другого базиса  $v_1, \dots, v_n$  в исходном пространстве (т.е. в пространствах  $V$  и  $W$ ). Займемся отысканием нового представления преобразования  $A$ . Пусть  $S$  есть матрица, описывающая изменение базиса, т.е.  $v_j = \sum s_{ij} w_i$ .

*Новая матрица  $[A]_v$ , представляющая рассматриваемое преобразование в базисе  $v$ , связана с исходной матрицей  $[A]_w$ , построенной по базису  $w$ , соотношением*

$$[A]_v = S^{-1} [A]_w S. \quad (10)$$

*Другими словами, замена базиса порождает преобразование подобия матрицы, представляющей линейное преобразование.*

Для произвольного вектора  $v = \sum x_j v_j$  вектор  $Sx$  дает его коэффициенты разложения по базису  $w$ , затем  $[A]_w Sx$  дает коэффициенты разложения  $Av$  по базису  $w$  и, наконец,  $S^{-1}[A]_w Sx$  дает коэффициенты разложения  $Av$  по исходному базису  $v$ . Поскольку это в точности совпадает со свойством, определяющим  $[A]_v$ , обе части в (10) должны совпадать друг с другом.

**Пример 11.** Пусть  $A$  есть преобразование, переводящее любой вектор  $(x, y)$  в его зеркальное отражение  $(y, x)$  относительно биссектрисы первого квадранта. Для стандартного базиса  $w_1 = (1, 0)$  и  $w_2 = (0, 1)$  соответствующая матрица легко вычисляется; поскольку преобразование  $A$  переводит  $w_1$  в  $w_2$  и  $w_2$  в  $w_1$ , это будет матрица перестановки

$$[A]_w = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Пусть мы поворачиваем плоскость, так что один из векторов нового базиса образует угол в  $45^\circ$  с осью  $x$ ,  $v_1 = (1, 1)$ , а другой перпендикулярен ему,  $v_2 = (-1, 1)$ . Тогда, записывая  $v_1$  в виде  $w_1 + w_2$  и  $v_2$  в виде  $-w_1 + w_2$ , получаем

$$S = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Для нового базиса преобразование  $A$  принимает особенно простой вид. Поскольку вектор  $v_1$  лежит на биссектрисе первого квадранта, он совпадает со своим зеркальным отражением, так что  $Av_1 = v_1$ . Другой базисный вектор  $v_2 = (-1, 1)$  переходит в противоположный,  $Av_2 = -v_2$ . Следовательно,

$$[A]_v = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

*Эта матрица является диагональной, поскольку базис  $v_1, v_2$  образован из собственных векторов преобразования  $A$ . Эти собственные векторы являются столбцами «матрицы замены базиса»  $S$ , так что она приводит  $A$  к диагональному виду, как это было в гл. 5:*

$$\Lambda = [A]_v = S^{-1}[A]_w S.$$

**Упражнение А.7.** Если вектор  $v_1$  ортогонален к  $v_2$  и  $A$  представляет собой проекцию на направление  $v_2$ , то как будет выглядеть матрица  $[A]_v$ ? Чему равны собственные значения этой (или любой другой) матрицы проектирования?

**Упражнение А.8.** Найти матрицу, собственные векторы которой суть  $v_1 = (1, 3)$  с собственным значением 4 и  $v_2 = (2, 0)$  с собственным значением 0.

## Приложение В

### ЖОРДАНОВА ФОРМА МАТРИЦЫ

Пусть задана квадратная матрица  $A$ , и мы ищем такую матрицу  $M$ , чтобы матрица  $M^{-1}AM$  была как можно ближе к диагональной. В простейшем случае матрица  $A$  обладает полным набором собственных векторов, которые и располагаются в столбцах матрицы  $M$ , или, в старых обозначениях, матрицы  $S$ . Жорданова форма в этом случае есть  $J = M^{-1}AM = \Lambda$ , и она состоит целиком из одномерных блоков  $J_i = \lambda_i$ , так что цель получения диагональной матрицы достигнута. В более общем и более сложном случае некоторые собственные векторы отсутствуют и получить диагональную форму матрицы невозможно. Именно этот случай нас сейчас и интересует.

Будем доказывать следующую теорему, которая уже была нами сформулирована в гл. 5:

**5S.** Если матрица имеет  $s$  линейно независимых собственных векторов, то она подобна матрице, имеющей специальную жорданову форму,

$$J = M^{-1}AM = \begin{bmatrix} J_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & J_s \end{bmatrix}$$

с блоками

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & & \cdot & \\ & & & 1 \\ & & & & \lambda_i \end{bmatrix}.$$

Примером такой жордановой матрицы может служить матрица

$$J = \begin{bmatrix} 8 & 1 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 8 & 1 \\ 0 & 8 \end{bmatrix} & & & & \\ & \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} & & & \\ & & & & \\ & & & & \\ & & & & [0] \end{bmatrix} = \begin{bmatrix} J_1 & & & & \\ & J_2 & & & \\ & & & & \\ & & & & \\ & & & & J_3 \end{bmatrix}.$$

Собственному значению  $\lambda=8$  кратности 2 здесь соответствует единственный собственный вектор, направленный вдоль первой координатной оси  $e_1=(1, 0, 0, 0, 0)^T$ ; поэтому число  $\lambda=8$  присутствует лишь в одном блоке  $J_1$ . Трехкратному собственному значению  $\lambda=0$  отвечают два собственных вектора,  $e_3$  и  $e_5$ , соответствующие жордановым клеткам  $J_2$  и  $J_3$ .

Основной вопрос состоит в следующем: если  $A$  является произвольной матрицей размера  $5 \times 5$ , то при каких условиях ее жорданова форма будет совпадать с  $J$ ? Когда найдется матрица  $M$ , такая, что  $M^{-1}AM=J$ ? Первое требование заключается в том, что каждая подобная матрица  $A$  должна иметь те же самые собственные значения 8, 8, 0, 0, 0. Но этого недостаточно, ведь диагональная матрица с такими собственными значениями не подобна  $J$ . В рассмотрении должны участвовать собственные векторы матрицы.

Для ответа на этот вопрос перепишем соотношение  $M^{-1}AM=J$  в более простом виде  $AM=MJ$ :

$$A \begin{bmatrix} x_1 & x_2 & \dots & x_5 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_5 \end{bmatrix} \begin{bmatrix} 8 & 1 & & & \\ 0 & 8 & & & \\ & & 0 & 1 & \\ & & 0 & 0 & \\ & & & & 0 \end{bmatrix}.$$

Перемножая матрицы, получаем

$$Ax_1=8x_1 \quad \text{и} \quad Ax_2=8x_2+x_1, \tag{1}$$

$$Ax_3=0x_3 \quad \text{и} \quad Ax_4=0x_4+x_3; \quad Ax_5=0x_5. \tag{2}$$

Теперь легко увидеть условия на матрицу  $A$ . Она должна иметь три истинных собственных вектора, как и матрица  $J$ . Собственный вектор, отвечающий собственному значению  $\lambda=8$ , будет находиться в первом столбце матрицы  $M$  точно так же, как это было бы для матрицы  $S$ :  $Ax_1=8x_1$ . Другие два вектора, обозначаемые через  $x_3$  и  $x_5$ , размещаются в третьем и пятом столбцах матрицы  $M$ :  $Ax_3=Ax_5=0$ . Кроме того, должны существовать еще два специальных вектора  $x_2$  и  $x_4$ , которые называются «обобщенными собственными векторами». Говорят, что вектор  $x_2$

принадлежит **цепочке векторов**, начинающейся с  $x_1$  и описываемой формулами (1). Фактически  $x_2$  является единственным, не считая  $x_1$ , вектором в цепочке<sup>1)</sup> и соответствующий блок  $J_1$  имеет второй порядок. Соотношения (2) описывают две различные цепочки, в одной из которых кроме  $x_3$  имеется еще вектор  $x_4$ , а другая состоит из одного вектора  $x_5$ ; блоки  $J_2$  и  $J_3$  имеют размеры  $2 \times 2$  и  $1 \times 1$  соответственно.

*Отыскание жордановой формы матрицы  $A$  сводится к отысканию этих цепочек векторов, каждая из которых возглавляется собственным вектором: для каждого  $i$  имеем*

$$\text{либо } Ax_i = \lambda_i x_i, \quad \text{либо } Ax_i = \lambda_i x_i + x_{i-1}. \quad (3)$$

Векторы  $x_i$  образуют столбцы матрицы  $M$ , и каждая отдельная цепочка дает один блок в матрице  $J$ . Мы должны показать, как строятся эти цепочки для произвольной матрицы  $A$ . Тогда если эти цепочки удовлетворяют соотношениям (1), (2), то матрица  $J$  будет жордановой формой для  $A$ .

Я думаю, что наиболее ясным и простым является метод, предложенный Филипповым и описанный в т. 26 «Вестника Московского университета». Построение осуществляется по индукции. Очевидно, что матрица первого порядка совпадает со своей жордановой формой. Предполагая, что построение выполнено для матриц порядка, меньшего, чем  $n$  (это и является предположением индукции), осуществим это построение для матрицы порядка  $n$ . Мы хотим объединить общее объяснение для матрицы  $A$  с частным для рассматриваемой матрицы  $J$  и поэтому заранее предполагаем, что обе эти матрицы имеют одинаковые собственные значения.

Метод Филиппова состоит из трех шагов.

*Шаг 1.* Поскольку одно из собственных значений есть  $\lambda = 0$ , матрица  $A$  вырождена и пространство ее столбцов имеет размерность  $r < n$ . Для этого пространства меньшей размерности предположение индукции гарантирует существование жордановой формы — в пространстве столбцов должны существовать  $r$  независимых векторов  $w_i$ , таких, что

$$Aw_i = \lambda_i w_i \quad \text{или} \quad Aw_i = \lambda_i w_i + w_{i-1}. \quad (4)$$

В нашем примере с  $A = J$  ранг матрицы равен  $r = 3$  и пространство столбцов натянуто на векторы  $e_1, e_2, e_3$ ; они будут играть роль векторов  $w_i$ , так что

$$Jw_1 = 8w_1, \quad Jw_2 = 8w_2 + w_1, \quad Jw_3 = 0w_3. \quad (5)$$

<sup>1)</sup> В отечественной литературе эти векторы называются присоединенными. — *Прим. перев.*

*Шаг 2.* Пусть нуль-пространство и пространство столбцов матрицы  $A$  имеют пересечение размерности  $p$ . Разумеется, каждый вектор из нуль-пространства является собственным вектором, соответствующим  $\lambda=0$ . Поэтому шаг 1 должен давать  $p$  цепочек, соответствующих этому собственному значению, и нас интересуют векторы  $w_i$ , находящиеся в *конце* цепочек. Каждый из этих  $p$  векторов принадлежит пространству столбцов, так что они являются комбинациями вектор-столбцов матрицы  $A$ :  $w_i = Ay_i$  для некоторого  $y_i$ .

В нашем примере с  $A=J$  нуль-пространство содержит векторы  $e_3$  и  $e_5$  и его пересечение с пространством столбцов натянуто на вектор  $e_3$ ; следовательно,  $p=1$ . В (5) имеется только одна цепочка, соответствующая  $\lambda=0$ , вектор  $e_3$  расположен на ее конце (или в начале) и  $e_3 = Je_4$ . Значит,  $y = e_4$ .

*Шаг 3.* Нуль-пространство имеет размерность  $n-r$ . Следовательно, помимо своего  $p$ -мерного пересечения с пространством столбцов оно должно содержать  $n-r-p$  дополнительных базисных векторов  $z_i$ , лежащих *вне* этого пересечения. В нашем примере  $n-r-p=5-3-1=1$  и нуль-вектор  $z=e_5$  является независимым от пространства столбцов: именно этот вектор  $z$  и дает одномерный блок  $J_3=(0)$ .

Объединяя эти шаги, мы и приходим к теореме Жордана:

*$r$  векторов  $w_i$ ,  $p$  векторов  $y_i$  и  $n-r-p$  векторов  $z_i$  образуют жордановы цепочки матрицы  $A$  и эти векторы являются линейно независимыми. Они составляют столбцы матрицы  $M$  и  $J = M^{-1}AM$  есть искомая жорданова форма.*

Если мы хотим перенумеровать эти векторы в виде  $x_1, \dots, x_n$  и согласовать их с условиями (3), то каждый вектор  $y_i$  должен располагаться сразу за порождающим его вектором  $w_i$ ; он будет завершать цепочку, соответствующую  $\lambda_i=0$ . Векторы  $z$  располагаются в конце, каждый в своей собственной цепочке, и вновь собственное значение равняется нулю, поскольку эти векторы принадлежат нуль-пространству. Блоки, соответствующие ненулевым собственным значениям, обработаны на первом шаге, блоки с нулевыми собственными значениями расширяются на одну строку и столбец на втором шаге, и третий шаг добавляет блоки  $J_i=(0)$  первого порядка.

Единственный, притом технический, вопрос заключается в проверке независимости всего набора векторов  $w_i$ ,  $y_i$  и  $z_i$ . Для этого предположим, что некоторая комбинация векторов обращается в нуль:

$$\sum c_i w_i + \sum d_i y_i + \sum g_i z_i = 0. \quad (6)$$

Умножая на  $A$  и учитывая соотношения (4) для векторов  $w_i$ , получаем

$$\sum c_i \begin{bmatrix} \lambda_i w_i \\ \text{либо} \\ \lambda_i w_i + w_{i-1} \end{bmatrix} + \sum d_i A y_i = 0. \quad (7)$$

Векторы  $A y_i = w_i$  располагаются в концах цепочек, соответствующих собственным значениям  $\lambda_i = 0$ , так что они не могут присутствовать в первой сумме. Поскольку (7) является некоторой комбинацией векторов  $w_i$ , которые по предположению индукции являются независимыми (они образуют жорданову форму в пространстве столбцов), коэффициенты  $d_i$  должны обращаться в нуль. Возвращаясь к (6), получаем  $\sum c_i w_i = -\sum g_i z_i$ , где левая часть принадлежит пространству столбцов. Поскольку векторы  $z$  не лежат в этом пространстве, все коэффициенты  $g_i$  должны равняться нулю. В результате имеем  $\sum c_i w_i = 0$ , так что независимость векторов  $w_i$  влечет за собой равенства  $c_i = 0$ .

Если исходная матрица  $A$  была невырожденной, то эти три шага должны применяться к матрице  $A' = A - cI$ . (Постоянная  $c$  выбирается таким образом, чтобы матрица  $A'$  была вырожденной; она может равняться любому собственному значению матрицы  $A$ .) Алгоритм приводит матрицу  $A'$  к ее жордановой форме  $M^{-1}A'M = J'$ , образуя цепочки  $x_i$  из векторов  $w_i$ ,  $y_i$  и  $z_i$ . Тогда в жордановой форме для матрицы  $A$  будут использованы те же самые цепочки и матрица  $M$ :

$$M^{-1}AM = M^{-1}A'M + M^{-1}cM = J' + cI = J.$$

Этим завершается доказательство того, что каждая матрица  $A$  подобна некоторой жордановой матрице  $J$ . С точностью до перестановки блоков **она подобна лишь одной такой матрице  $J$** , т. е. жорданова форма матрицы  $A$  является единственной. Поэтому множество матриц распадается на подмножества со следующим свойством: *Все матрицы одного и того же подмножества имеют одинаковую жорданову форму и все они связаны друг с другом преобразованиями подобия, но никакие две матрицы из разных подмножеств не являются подобными.* Этой классификацией и завершается наша книга.

**Пример.**

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Эта матрица имеет собственное значение  $\lambda = 0$  кратности три. Она взята нами со стр. 276 и имеет ранг  $r = 2$  и лишь один

собственный вектор. Пространство столбцов содержит лишь одну цепочку  $\omega_1, \omega_2$ , которая совпадает с последними двумя столбцами матрицы  $A$ :

$$A \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 0 \quad \text{и} \quad A \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

или

$$A\omega_1 = 0 \quad \text{и} \quad A\omega_2 = 0\omega_2 + \omega_1.$$

Нуль-пространство лежит в пространстве столбцов, и оно натянуто на вектор  $\omega_1$ . Следовательно, на втором шаге мы получаем  $p=1$ , так что вектор  $y$  находится из уравнения

$$Ay = \omega_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \quad \text{и он равен} \quad y = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Таким образом, цепочка  $\omega_1, \omega_2, y$  образует матрицу  $M$ :

$$M = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{и} \quad M^{-1}AM = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = J.$$

*Применение к решению уравнения  $du/dt = Au$ .* Как обычно, мы упрощаем задачу путем разделения переменных. Это разделение будет полным при наличии полного набора собственных векторов и  $u = Sv$ ; в настоящем же случае наилучшей заменой будет  $u = Mv$ . Это дает новое уравнение  $Mdv/dt = AMv$ , или  $dv/dt = Jv$ , которое является наипростейшим для данной матрицы. Его неизвестные связаны только внедиагональными единицами в каждой жордановой клетке. Для приведенной выше матрицы, подобной одному блоку, имеем

$$\frac{dv}{dt} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} v, \quad \text{или} \quad \begin{aligned} da/dt &= b, & a &= a_0 + b_0 t + c_0 t^2/2, \\ db/dt &= c, & b &= b_0 + c_0 t, \\ dc/dt &= 0, & c &= c_0. \end{aligned}$$

Система решается с последнего уравнения и на каждом шаге добавляется степень  $t$ . (Для блока порядка  $l$  в решении будут степени до  $t^{l-1}$ .) Экспоненты от матрицы  $J$  для этого случая и предыдущего случая матрицы порядка 5 имеют вид

$$e^{Jt} = \begin{bmatrix} 1 & t & t^2/2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix} \quad \text{и} \quad e^{Jt} = \begin{bmatrix} e^{st} & te^{st} & 0 & 0 & 0 \\ 0 & e^{st} & 0 & 0 & 0 \\ 0 & 0 & 1 & t & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Легко видеть, каким образом коэффициенты  $a$ ,  $b$  и  $c$  участвуют в первой из экспонент. И для второго примера также легко увидеть все пять «частных решений» уравнения  $du/dt = Au$ . Три из них являются чистыми экспонентами  $u_1 = e^{8t}x_1$ ,  $u_3 = e^{0t}x_3$ ,  $u_5 = e^{0t}x_5$ , полученный точно так же, как и раньше, из трех собственных векторов матрицы  $A$ . Оставшиеся два решения содержат обобщенные собственные векторы

$$u_2 = e^{8t}(tx_1 + x_2) \quad \text{и} \quad u_4 = e^{0t}(tx_3 + x_4). \quad (8)$$

Наиболее общее решение уравнения  $du/dt = Au$  является комбинацией  $c_1u_1 + \dots + c_5u_5$ , и комбинация, равняющаяся  $u_0$  при  $t=0$ , вновь имеет вид

$$u_0 = c_1x_1 + \dots + c_5x_5, \quad \text{или} \quad u_0 = Mc, \quad \text{или} \quad c = M^{-1}u_0.$$

Это означает лишь, что  $u = Me^{Jt}M^{-1}u_0$  и что матрицы  $S$  и  $\Lambda$  в старой формуле  $Se^{\Lambda t}S^{-1}u_0$  заменяются на  $M$  и  $J$ .

**Упражнение В.1.** Найти жорданову форму (в три шага) для матриц

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{и} \quad B = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

**Упражнение В.2.** Показать, что частное решение  $u_2$  в (8) удовлетворяет уравнению  $du/dt = Au$  именно благодаря наличию цепочки  $Ax_1 = 8x_1$ ,  $Ax_2 = 8x_2 + x_1$ .

**Упражнение В.3.** Для приведенной выше матрицы  $B$  вычислить, используя формулу  $Me^{Jt}M^{-1}$ , экспоненту  $e^{Bt}$  и сравнить ее с рядом  $I + Bt + \frac{1}{2}(Bt)^2 + \dots$ .

## Список литературы

### *Абстрактная линейная алгебра*

- Гантмахер Ф. Р. Теория матриц. — М.: Наука, 1967.  
Halmos P. R. Finite-Dimensional Vector Spaces, Van Nostrand-Reinhold, Princeton, New Jersey, 1958.  
Hoffman K., Kunze R. Linear Algebra, 2nd ed., Prentice-Hall, Englewood Cliffs, New Jersey, 1971.  
Muir T. Determinants, Dover, New York, 1960.

### *Прикладная линейная алгебра*

- Беллман Р. Введение в теорию матриц. — М.: Наука, 1969.  
Ben-Israel A., Greville T. N. E. Generalized Inverses. Theory and Applications, Wiley, New York, 1974.  
Bradley S. P., Нax A. C., Magnanti T. L. Applied Mathematical Programming, Addison-Wesley, Reading, Massachusetts, 1977.  
Gale D. The Theory of Linear Economic Models, McGraw-Hill, New York, 1960.  
Luenberger D. G. Introduction to Linear and Nonlinear Programming, Addison-Wesley, Reading, Massachusetts, 1973.  
Noble B. Applied Linear Algebra. Prentice-Hall, Englewood Cliffs, New Jersey, 1969.

### *Вычислительная линейная алгебра*

- Форсайт Дж., Молер К. Численное решение систем линейных уравнений. — М.: Мир, 1969.  
Lawson C. L., Hanson R. J., Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, New Jersey, 1974.  
Stewart G. W. Introduction to Matrix Computations, Academic Press, New York, 1973.  
Varga R. S. Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, New Jersey, 1962.  
Wilkinson J. H. Rounding Errors in Algebraic Processes, Prentice-Hall, Englewood Cliffs, New Jersey, 1963.  
Wilkinson J. H. The Algebraic Eigenvalue Problem, Oxford Univ. Press, London and New York, 1965. [Русский перевод: Уилкинсон Дж. Алгебраическая проблема собственных значений. — М.: Наука, 1970.]  
Wilkinson J. H., Reinsch C. Eds. Handbook for Automatic Computation II, Linear Algebra, Springer, Berlin and New York, 1971. [Русский перевод: Уилкинсон, Райнш. Справочник алгоритмов на языке АЛГОЛ. — М.: Машиностроение, 1976.]  
Young D. M. Iterative Solution of Large Linear Systems, Academic Press, New York, 1971.

# РЕШЕНИЯ

## Глава 1

1.2.1.  $u=3, v=1, w=0$ , с ведущими элементами 2, 1,  $-5$ .

1.2.2.  $u=1, v=2, w=3, z=4$ .

1.2.3. Коэффициент  $+1$  делает систему вырожденной.

1.2.4. При исключении вычисляется множитель  $l=c/a$ , а затем  $d-lb$ .

1.2.5. Стоимость 3 600 000 операций на ЭВМ составляет 1 доллар, так что теоретически можно решить систему порядка  $n=10\,800\,000^{1/3}=221$ ; на практике же нужно учитывать и накладные расходы. При увеличении затрат в 1000 раз число  $n$  увеличится лишь в 10 раз.

1.2.6. Второй член  $bc+ad$  равен  $(a+b)(c+d)-ac-bd$ .

1.2.7.  $u=1, v=3, w=2$ .

$$1.3.1. Ax = \begin{bmatrix} 17 \\ 4 \\ 17 \end{bmatrix}.$$

$$1.3.2. Ax = \begin{bmatrix} -1 \\ 6 \\ -4 \\ -1 \end{bmatrix} \quad \text{и} \quad [26].$$

$$1.3.3. \begin{cases} u_{\text{кон}} = 0.8u_{\text{нач}} + 0.1v_{\text{нач}} \\ v_{\text{кон}} = 0.2u_{\text{нач}} + 0.9v_{\text{нач}} \end{cases}, \quad A = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix}$$

$$(i) \begin{bmatrix} u_{\text{к}} \\ v_{\text{к}} \end{bmatrix} = A \begin{bmatrix} 30 \\ 200 \end{bmatrix} = \begin{bmatrix} 44 \\ 186 \end{bmatrix},$$

$$(ii) \begin{bmatrix} 30 \\ 200 \end{bmatrix} = A \begin{bmatrix} u_{\text{н}} \\ v_{\text{н}} \end{bmatrix}, \quad \begin{bmatrix} u_{\text{н}} \\ v_{\text{н}} \end{bmatrix} = \begin{bmatrix} 10 \\ 220 \end{bmatrix},$$

$$(iii) \begin{bmatrix} u \\ v \end{bmatrix} = A \begin{bmatrix} u \\ v \end{bmatrix}, \quad \begin{bmatrix} 0.2 & -0.1 \\ -0.2 & 0.1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = 0, \quad v = 2u.$$

1.3.4. Углы параллелограмма суть точки (2, 1), (0, 3), (2, 4) и (0, 0).

1.3.5. Результатом первого и второго шагов исключения является матрица

$$EA = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 3 & 2 \end{bmatrix}.$$

1.3.6. Результатом первого, второго и третьего шагов исключения является матрица

$$EA = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \end{bmatrix}.$$

1.3.7.  $EA = [0]$ ; произведение состоит из  $ln$  элементов; на получение каждого из них затрачивается  $m$  умножений.

$$1.3.8. EA = \begin{bmatrix} 13 \\ -8 \\ 2 \end{bmatrix}.$$

1.3.9. Матрица  $ABC$  является результатом первого и второго шагов исключения.

$$1.3.10. \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \text{ откуда следует } b=c=0,$$

$$\text{далее } \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}, \text{ откуда } a=d.$$

$$1.3.11. A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, D=A, E=F = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}.$$

1.3.12. Верно; неверно; верно; неверно.

1.3.13. Весами являются элементы первой строки матрицы  $A$ .

$$1.4.2. A = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} = LU; A=L, U=I.$$

$$1.4.3. LU = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 3 \\ 0 & 5 & 7 \\ 0 & 0 & -1 \end{bmatrix}, c = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}.$$

$$1.4.4. \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ c/a & 1 \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & (ad-bc)/a \end{bmatrix} \begin{bmatrix} 1 & b/a \\ 0 & 1 \end{bmatrix}.$$

$$1.4.5. L = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{bmatrix}, D = \begin{bmatrix} 2 \\ \frac{3}{2} \\ \frac{4}{3} \end{bmatrix}, U = \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 1 \end{bmatrix},$$

$$c = \begin{bmatrix} 6 \\ 3 \\ -4 \end{bmatrix}, x = \begin{bmatrix} 3 \\ 0 \\ -3 \end{bmatrix}.$$

1.4.6. Отношение  $(150)^3/3$  к  $(150)^2$  равно 50.

$$1.4.7. c = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}, x = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix}.$$

1.5.1.  $u=2, v=-3, w=4$ ; переставить строки 2 и 3.

1.5.2. Второе уравнение есть  $0u+0v=b_2-3b_1$ , так что решение существует только при  $b_2=3b_1$ .

$$1.5.3. P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, L=I, D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, U = \begin{bmatrix} 1 & \frac{3}{2} \\ 0 & 1 \end{bmatrix}.$$

$$1.5.4. P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

1.5.5. Обе системы являются вырожденными; первая не имеет решения, а вторая имеет решение  $u=v=w=7$  (или любое другое число).

$$1.5.6. C^{-1}BA^{-1}.$$

$$1.5.7. A^{-1}=A, B^{-1} \text{— транспонированная к } B = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}.$$

1.5.9. Третья строка умножается на 8 и прибавляется к первой строке; в матрице  $E^{-1}$  вместо числа 8 будет стоять  $-8$ .

$$1.5.10. A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \quad A^{-1} = \frac{1}{4} \begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

$$1.5.11. A^{-1} = \begin{bmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}.$$

1.5.12. Ведущие элементы равняются 0.001 и 1000; после перестановки они становятся равными 1 и  $-1$ .

1.5.13. Если ведущий элемент является большим, то он умножается на число, меньшее единицы, при исключении каждого из расположенных ниже элементов. Предельный случай, когда все множители равняются плюс или минус единице, — это

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 1 \\ -\frac{1}{2} & 0 & 1 \\ -\frac{1}{2} & -1 & 1 \end{bmatrix}.$$

$$1.6.1. L = \begin{bmatrix} 1 & & & & & \\ & -1 & & & & \\ & & \cdot & \cdot & & \\ & & & \cdot & \cdot & \\ & & & & \cdot & \cdot \\ & & & & & -1 & 1 \end{bmatrix}, \quad D=I, \quad U=L^T.$$

1.6.2. Нижняя матрица второго порядка становится симметрической:

$$\begin{bmatrix} d-b^2/a & e-bc/a \\ e-bc/a & f-c^2/a \end{bmatrix}.$$

$$1.6.3. A = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}.$$

1.6.4.  $(u_1, u_2, u_3) = (\pi^2/8, 0, -\pi^2/8)$  вместо истинных значений  $(1, 0, -1)$ .

## Глава 2

2.1.1. Точки с целочисленными координатами; крест, образованный осями.

2.1.2. Да, нет, нет, да, да, да.

2.1.3. Если  $Ax=0$ ,  $Ay=0$ , то  $A(cx+dy)=0$ . Если  $b \neq 0$ , то решения не содержат точку  $x=0$  и не могут образовывать подпространство.

$$2.2.1. \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \text{ — не ступенчатая форма.}$$

$$2.2.2. x+y+z=1, \quad x+y+z=0.$$

$$2.2.3. LU = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Первые два столбца матрицы  $U$  имеют ведущие элементы, так что  $r=2$ . Две другие переменные,  $w$  и  $y$ , являются свободными, и общее решение записывается в виде

$$x = \begin{bmatrix} 2w-y \\ -w \\ w \\ y \end{bmatrix} = w \begin{bmatrix} 2 \\ -1 \\ 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

2.2.4. Вторая переменная является базисной (ведущий элемент не равен нулю), а переменные  $u$ ,  $w$  и  $y$  являются свободными. Ранг равен единице, и общее решение имеет вид  $x=(u, -4w, w, y)^T$ . При  $b \neq 0$  система является совместной только в случае  $b_2 - 2b_1 = 0$ , и тогда

$$x = \begin{bmatrix} u \\ b_1 - 4w \\ w \\ y \end{bmatrix} = u \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + w \begin{bmatrix} 0 \\ -4 \\ 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ b_1 \\ 0 \\ 0 \end{bmatrix}.$$

2.2.5. Переменная  $u$  является базисной, а  $v$  — свободной,  $r=1$ ,  $x=v \begin{bmatrix} -2 \\ 1 \end{bmatrix}$ .

В общем случае должно быть  $b_1=0$ ,  $b_3=4b_2$ ,  $b_4=0$ , так что  $x=v \begin{bmatrix} -2 \\ 1 \end{bmatrix} + \begin{bmatrix} b_2 \\ 0 \end{bmatrix}$ .

2.2.6. Процесс исключения дает  $u+2v+2w=1$ ,  $w=2$ , так что переменная  $u$  является свободной:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} -2v-3 \\ v \\ 2 \end{bmatrix} = v \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix}.$$

2.2.7. Из третьего уравнения имеем условие  $0=b_3-2b_1-3b_2$ ; ранг равен двум.

2.2.8.  $c=7$ .

2.3.1.  $v_1-v_2+v_3-v_4=0$ .

2.3.2. Строки являются независимыми.

2.3.4. Если  $c_1(v_1+v_2)+c_2(v_1+v_3)+c_3(v_2+v_3)=0$ , то  $(c_1+c_2)v_1+(c_1+c_3)v_2+(c_2+c_3)v_3=0$ . В силу независимости  $v$  отсюда получаем  $c_1+$

$+c_2=0, c_1+c_3=0, c_2+c_3=0$ , так что все коэффициенты  $c$  равны нулю. Следовательно,  $w$  также являются независимыми.

2.3.5. Пространство столбцов порождено вектором  $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ ; пространство строк также представляет собой прямую, но оно порождается вектором  $(1, 2)$ .

2.3.6. Столбцы 2 и 3 образуют базис в пространстве столбцов; последний столбец равен 7 (столбец 2) — (столбец 3). Строки 1 и 2 образуют базис в пространстве строк.

2.3.7. Один из базисов имеет вид  $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ .

Ступенчатые матрицы порождают подпространство треугольных матриц.

2.3.8.  $v_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$  и  $v_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ , или  $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$  и  $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ .

2.3.9. Пусть  $v_1=(1, 0, 0, 0), \dots, v_4=(0, 0, 0, 1)$  являются координатными векторами. Если прямая  $W$  порождена вектором  $(1, 2, 3, 4)$ , то ни один из векторов  $v$  не принадлежит этой прямой.

2.3.10. Нулевая размерность при  $x=y=z=0$ ; единичная размерность при  $x=(1, 0, 0), y=(0, 1, 0), z=(-1, -1, 0)$ ; поскольку векторы зависимы, они не могут образовать базис всего пространства.

2.3.12. (i) Если бы эти векторы не образовывали базис, то мы могли бы добавить независимые векторы, увеличив тем самым заданную размерность  $k$ , что приводит к противоречию. (ii) Если бы эти векторы не образовывали базис, то мы могли бы убрать некоторые из них, уменьшив тем самым размерность  $k$ , что приводит к противоречию.

2.3.13. Размерность равняется 6.

2.3.14. Если векторы  $v_1, v_2, v_3$  образуют базис пространства  $V$ , а векторы  $w_1, w_2, w_3$  образуют базис пространства  $W$ , то эти шесть векторов не могут быть независимыми и некоторая их линейная комбинация дает нуль:

$\sum c_i v_i + \sum d_i w_i = 0$ , т. е.  $\sum c_i v_i = -\sum d_i w_i$  — вектор, принадлежащий обоим подпространствам.

2.4.1. Неверно; размерности совпадают.

2.4.2. Второй столбец является базисом пространства  $\mathfrak{R}(A)$ , а первый столбец — базисом для  $\mathfrak{R}(A^T)$ . Вектор  $(-2, 1)^T$  порождает левое нуль-пространство, а векторы  $(1, 0, 0, 0)^T, (0, -4, 1, 0)^T, (0, 0, 0, 1)^T$  образуют базис пространства  $\mathfrak{N}(A)$ .

2.4.3. Первый и второй столбцы образуют базис пространства  $\mathfrak{R}(A)$ , а первая и вторая строки — базис пространства  $\mathfrak{R}(A^T)$ . Вектор  $(1, 0, -1)^T$  порождает левое нуль-пространство, а векторы  $(-1, 0, 0, 1)^T, (2, -1, 1, 0)^T$  являются базисом пространства  $\mathfrak{N}(A)$ .

2.4.4. В пространстве столбцов все четвертые компоненты равняются нулю, а в пространстве строк равны нулю все первые компоненты. Нуль-пространство порождено вектором  $(1, 0, 0, 0)^T$ , а левое нуль-пространство — вектором  $(0, 0, 0, 1)^T$ .

2.4.5. Если  $AB=0$ , то матрица  $A$  переводит каждый из столбцов матрицы  $B$  в нуль; следовательно, то же самое она делает и с любой комбинацией столбцов, и все эти комбинации должны лежать в пространстве  $\mathfrak{N}(A)$ .

2.4.6. Если размерность не увеличивается при добавлении столбца  $b$ , то вектор  $b$  должен быть комбинацией столбцов матрицы  $A$ , так что система  $Ax=b$  будет разрешима.

$$2.4.7. d = bc/a; \quad A = \begin{bmatrix} a \\ c \end{bmatrix} [1 \quad b/a].$$

$$2.4.8. A = \begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix} [1 \quad -1], \quad B = \begin{bmatrix} 1 \\ 3 \end{bmatrix} [1 \quad 1 \quad 2],$$

$$AB = u(v^T w) z^T = \begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix} (-2) [1 \quad 1 \quad 2] = \begin{bmatrix} -4 & -4 & -8 \\ -8 & -8 & -16 \\ 0 & 0 & 0 \end{bmatrix}.$$

2.4.10. Отображение  $y = e^x$  является взаимно однозначным, но никогда не принимает нулевое значение; функция  $z = \operatorname{tg} x$  принимает все значения, но более чем по одному разу.

2.4.11. Если условие существования решения для матрицы  $A$  выполняется, то  $r = m$  и  $m$  столбцов матрицы  $A^T$  должны быть независимыми. Если для матрицы  $A$  выполняется условие единственности, то  $r = n$  и столбцы матрицы  $A^T$  должны образовывать базис всего пространства  $R^n$ .

2.4.12. Матрица  $A$  не имеет левой обратной; все правые обратные имеют вид

$$C = \begin{bmatrix} 1-u \\ u \\ v \end{bmatrix}.$$

$$2.5.1. \|x\|^2 = 21, \|y\|^2 = 18, x^T y = 0.$$

2.5.2. Векторы  $(1, 0)$  и  $(1, 1)$  независимы, но не являются ортогональными; векторы  $(0, 0)$  и  $(1, 1)$  ортогональны, но не являются независимыми.

$$2.5.3. \text{Если } (x_2/x_1) (y_2/y_1) = -1, \text{ то } x_2 y_2 = -x_1 y_1, \text{ или } x_1 y_1 + x_2 y_2 = 0.$$

2.5.4. Если  $i \neq j$ , то элемент с индексами  $i, j$  матрицы  $BB^{-1} = I$  равен нулю.

$$2.5.5. v_1^T v_3 = 0, v_2^T v_3 = 0.$$

2.5.6. Любой вектор, кратный решению  $(1, 1, -2)$  системы  $u + v + w = 0$ ,  $u - v = 0$ , ортогонален к двум заданным векторам. Ортонормированная система имеет вид  $(1, 1, 1)/\sqrt{3}$ ,  $(1, -1, 0)/\sqrt{2}$ ,  $(1, 1, -2)/\sqrt{6}$ .

$$2.5.7. w_2 = (0, 0, 0, 1); v_3 = (0, 0, 5, -4).$$

2.5.8. Любой вектор из  $V \cap W$  должен быть ортогонален самому себе.

2.5.9. Решая систему  $u + v + 2w = 0$ ,  $u + 2v + 3w = 0$ , получаем, что ортогональным дополнением является пространство, порожденное вектором  $(1, 1, -1)$ .

$$2.5.10. \text{Система имеет вид } x + y - z = 0.$$

2.5.11. Выбрать в качестве  $y$  компоненту вектора  $b$ , лежащую в левом нуль-пространстве.

$$2.5.12. \text{Вектор } (2, 2, -1) \text{ является базисом нуль-пространства; } (3, 3, 3) = (1, 1, 4) + (2, 2, -1) = x_r + x_n$$

2.5.13. Вектор  $x$  ортогонален к  $V + W$  тогда и только тогда, когда он ортогонален к каждому вектору из  $V$  и  $W$ ; он принадлежит пересечению пространств  $V^\perp$  и  $W^\perp$ .

2.5.15. Падение напряжения равняется 5 между 2 и 3, между 3 и 1, между 4 и 1 и между 2 и 5. Между 3 и 4 нет падения напряжения, поскольку  $p_3 = p_4 = 5$ . Между 1 и 2 скачок напряжения равен 10, так что  $p_2 = 10$ .

2.5.16. Ток от 1 к 3 равен  $I = 3$ ; половина тока возвращается через 2, другая половина — через 4, а от 2 к 4 ток не течет.

2.6.1. Пересечение  $V \cap W$  равно пространству двухдиагональных матриц, имеющих нулевые элементы всюду, кроме главной диагонали и первой поддиагонали; размерность равна 7.  $V+W$  равно совокупности матриц с нулевыми элементами ниже первой поддиагонали; размерность равна 13. Исходные размерности равны  $\dim V=10$  и  $\dim W=10$ .

2.6.2. Если  $x=v+w$  и  $x=v'+w'$ , то  $v+w=v'+w'$ , или  $v-v'=w'-w$ . Последний вектор принадлежит как  $V$ , так и  $W$  и должен равняться нулю.

2.6.3. Одна из возможностей: пространство  $W$  порождается векторами  $(1, 0, 0, 0)$  и  $(0, 0, 1, 0)$ . Все четыре вектора образуют независимую систему.

2.6.4. Векторы  $v_1, v_2$  и  $w_1$  образуют базис пространства  $V+W$ ; пространство  $V \cap W$  является одномерным, и оно порождается векторами  $v_1-v_2 (=w_1-w_2)$ .

$$2.6.6. A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad AB = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

$$2.6.7. \text{ Пусть } A = [0 \ 0], \quad B = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad AB = [0], \quad v(AB) = 1 < v(A) = 2.$$

$$2.6.8. A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} [0 \ 1 \ 4 \ 0] = \bar{L}\bar{U}.$$

$$2.6.9. A = P^{-1}LU = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 4 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 4 \\ 0 \end{bmatrix} [1 \ 2] = \bar{L}\bar{U}.$$

$$2.6.10. A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} [1 \ 3 \ 3 \ 2] + \\ + \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} [0 \ 0 \ 3 \ 1].$$

2.6.11. Ранг равен единице; любая подматрица этого размера обратима.

2.6.12.  $\text{ранг}(AB) \leq \text{ранг}(B) \leq n < m$ , но матрица  $AB$  имеет размер  $m \times m$ .

2.6.13. Если векторы  $v_1, \dots, v_r$  образуют базис в пространстве столбцов матрицы  $A$ , а векторы  $w_1, \dots, w_s$  образуют базис в пространстве столбцов матрицы  $B$ , то любой столбец матрицы  $A+B$  является линейной комбинацией векторов  $v$  и  $w$ ; они порождают пространство столбцов матрицы  $A+B$ , так что его размерность не может превышать  $r+s$ .

2.6.14. Пусть вектор  $y$  принадлежит пространству  $\mathfrak{N}(B) \cap \mathfrak{N}(A)$ , так что  $y=Bx$  для некоторого  $x$  и  $Ay=0$ . Тогда  $ABx=0$ , так что вектор  $x$  лежит в пространстве  $\mathfrak{N}(AB)$ . Но оно совпадает с  $\mathfrak{N}(B)$ , так что  $Bx=0$ . Следовательно, единственным вектором из  $\mathfrak{N}(B) \cap \mathfrak{N}(A)$  является  $y=Bx=0$ .

2.6.16. (i) Любая матрица  $B = \begin{bmatrix} 0 & 0 \\ c & d \end{bmatrix}$  дает  $AB=0$ , так что нуль-пространство является двумерным. (ii) Если  $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ , то  $AB = \begin{bmatrix} a & b \\ 2a & 2b \end{bmatrix}$ , что является двумерным подпространством.

## Глава 3

3.1.1. (a)  $a^T b = 2 \sqrt{xy} \leq \|a\| \|b\| = x + y$ . (b)  $\|x + y\|^2 \leq (\|x\| + \|y\|)^2$  — это то же самое, что  $x^T x + 2x^T y + y^T y \leq x^T x + 2\|x\| \|y\| + y^T y$ , что представляет собой неравенство Шварца  $x^T y \leq \|x\| \|y\|$ .

3.1.2.  $Op^2 = (a^T b / a^T a)^2 a^T a$ ; добавляя это к уравнению (5), получаем  $b^T b$ , т. е. квадрат гипотенузы.

3.1.3.  $p = (a^T b / a^T a) a = \frac{10}{3} (1, 1, 1)$ ; другая проекция есть  $(b^T a / b^T b) b$ .

3.1.4. Оно является равенством в обоих случаях; угол равен 0 или  $\pi$ , косинус равен 1 или  $-1$ , так что ошибка  $bp$  равняется нулю, поскольку вектор  $b$  совпадает со своей проекцией  $p$ .

3.1.5.  $\theta$  есть угол, косинус которого равен  $1/\sqrt{n}$ .

3.1.6. Для каждого  $j$   $|a_j|^2 + |b_j|^2 - 2|a_j||b_j|$  равняется  $(|a_j| - |b_j|)^2$  и не может быть отрицательным.

3.1.7. Транспонированная к матрице  $A^{-1}$  равняется, как было показано,  $(A^T)^{-1}$ , что совпадает с  $A^{-1}$  при  $A = A^T$ .

3.1.8.  $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = A^T$ ,  $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = B^T$ ,  $AB = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \neq (AB)^T$ .

3.1.9. Пусть матрица  $A$  имеет больше строк, чем столбцов, например  $A = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ .

3.1.10. Лучи к вершинам образуются векторами  $v = \left(-\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right)$  и  $w = \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)$ , так что  $\cos \theta = v^T w / \|v\| \|w\| = -\frac{1}{3}$ .

3.2.1.  $\bar{x} = 151$ .

3.2.2.  $\bar{x} = 2$ .

3.2.3.  $\bar{u} = \bar{v} = \frac{1}{3}$ ;  $\bar{x} = 3$ .

3.2.4.  $E^2 = (u-1)^2 + (v-3)^2 + (u+v-4)^2$ ;  $dE^2/du = 2(u-1) + 2(u+v-4)$ ,  $dE^2/dv = 2(v-3) + 2(u+v-4)$ ,  $\bar{u} = 1$ ,  $\bar{v} = 3$ ;  $p = A \begin{bmatrix} 1 \\ 3 \end{bmatrix} = b$ , поскольку вектор  $b$  сам лежит в пространстве столбцов.

3.2.5.  $P = \frac{1}{3} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix}$ ,  $Pb = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ ,  $(I-P)b = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$ ,  $\bar{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ .

3.2.6. (a)  $(I-P)^2 = I - 2P + P^2 = I - P$ ;  $(I-P)^T = I - P^T = I - P$ .

(b)  $(P_1 + P_2)^2 = P_1^2 + P_1 P_2 + P_2 P_1 + P_2^2 = P_1 + P_2$ ;  $(P_1 + P_2)^T = P_1^T + P_2^T = P_1 + P_2$ .

3.2.7.  $(2P - I)^2 = 4P^2 - 4P + I = I$ .

3.2.8.  $P^2 = uu^T uu^T = uu^T = P$ , поскольку  $u^T u = 1$ ; матрица  $uu^T$  является симметрической.

3.2.9.  $P = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ ; тогда  $P \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}$ .

3.2.11.  $y = -\frac{3}{10} - \frac{12}{5}t$ .

$$3.2.12. \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -3 \\ -5 \end{bmatrix}.$$

$$3.3.1. \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} -4 \\ -3 \\ -1 \\ 0 \end{bmatrix}.$$

$c=2C$ ,  $d=\sqrt{10D}$ ;  $y=-2+t$  является оптимальным и действительно проходит через все четыре точки;  $E^2=0$  и  $b=p$ , поскольку  $b$  принадлежит пространству столбцов.

3.3.2. Отдельными проекциями являются  $2a_1$  и  $2a_2$ ;  $p=2a_1+2a_2$ .

3.3.3. Эта проекция равна  $-a_3$ ; сумма  $2a_1+2a_2-a_3=(0, 3, 0)$  совпадает с вектором  $b$ ; оператор проектирования на все пространство является тождественным.

$$3.3.4. (Q_1Q_2)^T(Q_1Q_2)=Q_2^TQ_1^TQ_1Q_2=Q_2^T I Q_2=I.$$

3.3.5.  $Q^TQ=(I-2uu^T)(I-2uu^T)=I-4uu^T+4uu^Tuu^T=I$ , поскольку  $u^Tu=1$ .

3.3.6. Третий столбец есть  $\pm(1, -2, 1)/\sqrt{6}$ .

$$3.3.7. \|v\|^2=(x_1q_1+\dots+x_nq_n)^T(x_1q_1+\dots+x_nq_n)=x_1^2q_1^Tq_1+\dots+x_n^2q_n^Tq_n=x_1^2+\dots+x_n^2.$$

3.3.8. Матрица  $Q$  является ортогональной.

$$3.3.9. q_1=a_1, q_2=a_2-a_1, q_3=a_3-a_2,$$

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = QR.$$

$$3.3.10. \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 5 & 4 \\ 0 & 3 \end{bmatrix}.$$

$$3.3.11. \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & & 0 \\ & \frac{2}{3} & 1/\sqrt{2} \\ & \frac{2}{3} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 3 & 3 \\ 0 & \sqrt{2} \end{bmatrix};$$

матрицы  $A$  и  $Q$  имеют размер  $m \times n$ , а матрица  $R$  — размер  $n \times n$ .

$$3.3.12. \bar{x} = R^{-1}Q^Tb = \begin{bmatrix} 5 \\ 9 \\ 0 \end{bmatrix}.$$

3.3.13. Пространства столбцов матриц  $Q$  и  $A$  совпадают, так что  $P = Q(Q^TQ)^{-1}Q^T = QQ^T$ .

$$3.3.14. v_3 = \left( c - \frac{v_1^T c}{v_1^T v_1} v_1 \right) - \frac{v_2^T c}{v_2^T v_2} v_2, \text{ поскольку } v_2^T v_1 = 0.$$

$$3.3.15. \|v\|^2 = \frac{1}{2} + \frac{1}{4} + \dots = 1, \|e^x\|^2 = \int e^{2x} dx = (e^2 - 1)/2, \int e^x e^{-x} dx = 1.$$

3.3.16.  $b_1 = \int y \sin x / \int \sin^2 x$ ; если  $y = \cos x$ , то коэффициент Фурье  $b_1$  равен нулю.

$$3.3.17. a_0 = 1, a_1 = 0, b_1 = 2/\pi.$$

3.3.18. Функция  $x^3$  уже ортогональна четным функциям 1 и  $x^2 - \frac{1}{3}$ ;

$$\int_{-1}^1 x^3 x dx = \frac{2}{5}, \text{ так что следующий многочлен Лежандра имеет вид } x^3 - \frac{2}{5} x.$$

3.3.19. Горизонтальная прямая  $y = 1/3$ .

3.4.1. Если  $A = 0$ , то пространство строк есть 0,  $\bar{x} = 0$  и  $A^+ = 0$ .

$$3.4.2. A = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} [1 \ 1 \ 1] = \bar{L}\bar{U}; \quad A^+ = \frac{1}{6} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix},$$

$$\bar{x} = A^+ b = \frac{1}{6} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad A\bar{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix} = p.$$

3.4.3.  $A^+ = A^T$ , поскольку решением системы  $Ax = b$  по методу наименьших квадратов является вектор  $\bar{x} = A^T b$ .

$$3.4.4. \bar{u} = \bar{v} = \frac{3}{2}.$$

$$3.5.1. \text{Из (59) имеем } \bar{x}_W = (4b_1 + b_2)/(4 + 1) = \frac{17}{5}.$$

3.5.2. Вектор  $c$  является  $W$ -перпендикулярным к  $b$ , если  $(c_1, c_2) W^T W b = 0$ , или  $4c_1 + c_2 = 0$ ;  $\|b\|_W = \|Wb\| = \sqrt{5}$ .

$$3.5.3. \bar{x}_W = \frac{8}{13}, \text{ обратное к изображенному на рис. 3.9.}$$

3.5.4. Любая ортогональная матрица.

## Глава 4

$$4.2.1. \det 2A = 2^n \det A; \det(-A) = (-1)^n \det A; \det A^2 = (\det A)^2.$$

4.2.2. Правила (5) и (1) влекут за собой (2).

$$4.2.3. \det A = 20, \det A = 5.$$

4.2.4.  $\det A^T = \det A$ ,  $\det(-A) = (-1)^3 \det A$ , следовательно,  $\det A = -\det A$  или  $\det A = 0$ .

$$4.2.5. (a) 0; (b) 16; (c) 16; (d) \frac{1}{16}; (e) 16.$$

4.2.7.  $\det Q^T \det Q = 1$ , или  $(\det Q)^2 = 1$ ; единичный куб.

4.3.1. Столбцы 2, 1, 4, 3; перестановка четная, так что  $|A| = 1$ .

$$4.3.4. |A_4| = -3, |A_3| = 2, |A_2| = -1, |A_n| = (-1)^n (1 - n).$$

$$4.3.5. |A| = 20, B = \begin{bmatrix} 20 & -10 & -12 \\ 0 & 5 & 0 \\ 0 & 0 & 4 \end{bmatrix}, A^{-1} = B/20.$$

$$4.4.1. A^{-1} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

$$4.4.2. x=3, y=-1, z=-2.$$

$$4.4.3. J=r^2 \cos \varphi \quad (\varphi=0 \text{ есть экватор!}).$$

4.4.4. Треугольник  $A'B'C'$  имеет такую же площадь, как и  $ABC$ , он просто сдвинут в начало координат.

4.4.5. Каждая из трех строк увеличивается в три раза.

4.4.6. Ведущие элементы равны  $2, \frac{6}{2}, \frac{0}{6}$ ; нет, матрица  $A$  вырожденная.

4.4.7. Начиная с  $n=2$ , перестановки являются нечетной, нечетной, четной, четной, нечетной, нечетной и т. д. Задача для  $n+1$  подобна задаче для  $n$ , за исключением того, что число  $n+1$  необходимо будет передвинуть из одного конца в другой, для чего требуется  $n$  дополнительных перестановок. Поэтому четность сохраняется, когда  $n$  четное, и изменяется при нечетном  $n$ .

4.4.8. Четное.

4.4.9. Зависимы, поскольку составленный из них определитель равен нулю.

## Глава 5

5.1.1.  $\lambda=2, \lambda=3$ , след равен 5, определитель равен 6.

$$5.1.2. u = c_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} e^{2t} + c_2 \begin{bmatrix} 1 \\ -2 \end{bmatrix} e^{3t} = 6 \begin{bmatrix} 1 \\ -1 \end{bmatrix} e^{2t} - 6 \begin{bmatrix} 1 \\ -2 \end{bmatrix} e^{3t}.$$

5.1.3.  $\lambda=-5, \lambda=-4$ ; оба уменьшаются на 7 при неизменных собственных векторах.

$$5.1.4. c_1 x_1 + c_2 x_2 + c_3 x_3 = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}, \text{ или } c_1 = \frac{5}{3}, c_2 = 0, c_3 = -\frac{2}{3}.$$

$$5.1.5. \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}; \lambda=3, 1, 2; \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

5.1.7. Если  $Ax = \lambda x$ , то  $(A-7I)x = (\lambda-7)x$ ; если  $Ax = \lambda x$ , то  $x = \lambda A^{-1}x$  или  $A^{-1}x = \lambda^{-1}x$ .

5.1.8. Взять  $\lambda=0$ .

$$5.1.11. \det(A - \lambda I) = \det(A^T - \lambda I).$$

$$5.1.12. \lambda_1=5, \lambda_2=-5, x_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, x_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

$$5.1.13. \lambda_1=2, \lambda_2=0, \lambda_3=-2.$$

5.2.1. Столбцы матрицы  $S$  суть  $x_1, x_2, x_3$ ;  $S^{-1}AS = \text{diag}(0, 1, 3)$ .

$$5.2.2. A = \begin{bmatrix} -5 & 18 \\ -3 & 10 \end{bmatrix}.$$

5.2.3.  $\lambda=0, 0, 3$ ; третий столбец матрицы  $S$  является кратным вектору  $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ , а другие столбцы ортогональны к нему.

5.2.4.  $S^T A^T (S^{-1})^T = \Lambda^T$ , так что  $(S^{-1})^T$  является матрицей собственных векторов для  $A^T$ .

$$5.2.5. AB = SA_1 S^{-1} SA_2 S^{-1} = SA_1 A_2 S^{-1} = SA_2 A_1 S^{-1} = SA_2 S^{-1} SA_1 S^{-1} = \\ = BA, \quad B = \begin{bmatrix} \alpha & \beta \\ \beta & \alpha \end{bmatrix}.$$

$$5.2.6. R = (\text{вращение на } 45^\circ) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix}.$$

5.2.7. Матрица  $\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$  имеет только один собственный вектор.

$$5.3.1. u_0 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad c_1 = \frac{3 - \lambda_2}{\lambda_1 - \lambda_2}, \quad c_2 = \frac{-3 + \lambda_1}{\lambda_1 - \lambda_2}, \quad F_k = c_1 \lambda_1^k + c_2 \lambda_2^k;$$

$$F_{k+1}/F_k \rightarrow \lambda_1.$$

$$5.3.2. \begin{bmatrix} G_{k+2} \\ G_{k+1} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} G_{k+1} \\ G_k \end{bmatrix}, \quad \lambda_1 = 1, \quad \lambda_2 = -\frac{1}{2},$$

$$\begin{bmatrix} G_{k+1} \\ G_k \end{bmatrix} = S \Lambda^k S^{-1} \begin{bmatrix} G_1 \\ G_0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 1^k & \\ & \left(-\frac{1}{2}\right)^k \end{bmatrix} \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix},$$

$$G_k = \frac{1}{3} \left[ 1 - \left(-\frac{1}{2}\right)^k \right] \rightarrow \frac{1}{3}.$$

$$5.3.3. \begin{bmatrix} 3000 \\ 3000 \\ 3000 \end{bmatrix}, \quad \begin{bmatrix} 18\,000 \\ 1500 \\ 1000 \end{bmatrix}, \quad \begin{bmatrix} 6000 \\ 9000 \\ 500 \end{bmatrix}, \quad \begin{bmatrix} 3000 \\ 3000 \\ 3000 \end{bmatrix}, \quad \begin{bmatrix} 18\,000 \\ 1500 \\ 1000 \end{bmatrix}, \quad \begin{bmatrix} 6000 \\ 9000 \\ 500 \end{bmatrix},$$

$$\begin{bmatrix} 3000 \\ 3000 \\ 3000 \end{bmatrix}.$$

$$5.3.4. A = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \begin{array}{l} \text{Бостон} \\ \text{Лос-Анджелес,} \\ \text{Чикаго} \end{array}, \quad u_\infty = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}.$$

5.3.5. В стационарном состоянии все являются мертвыми:  $d_\infty = 1$ ,  $s_\infty = 0$ ,  $w_\infty = 0$ .

$$5.3.6. A = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{4} & \frac{1}{2} \end{bmatrix}, \quad u_\infty = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \\ \frac{1}{4} \end{bmatrix}.$$

$$5.3.7. A^T A - I = \begin{bmatrix} -1 & 0 \\ 0 & a_{12}^2 - \frac{3}{4} \end{bmatrix};$$

собственные значения лежат на диагонали, и оба будут отрицательными при  $a_{12}^2 < \frac{3}{4}$ .

5.3.8. Система неустойчива при  $|\alpha| > \frac{1}{2}$  и устойчива при  $|\alpha| < \frac{1}{2}$ .

5.3.9. Увеличение матрицы  $A$  означает рост производственных расходов, что приводит к замедлению темпов развития.

$$5.3.10. A^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A^3 = 0, \quad (I - A)^{-1} = I + A + A^2 = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

$$5.4.1. e^{At} = Se^{\Lambda t}S^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} e^{-t} & \\ & e^{-3t} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} = \\ = \frac{1}{2} \begin{bmatrix} e^{-t} + e^{-3t} & e^{-t} - e^{-3t} \\ e^{-t} - e^{-3t} & e^{-t} + e^{-3t} \end{bmatrix} > 0.$$

5.4.2. Собственные значения равны  $\lambda = 1$ ,  $\lambda = 3$ ;  $u = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} e^t + \frac{1}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} e^{3t} \rightarrow \infty$ .

$$5.4.3. \lambda_1 = 0, \lambda_2 = -2, \quad x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad u(t) = c_1 x_1 + c_2 x_2 e^{-2t}; \\ u_0 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \text{ дает } u = \begin{bmatrix} 2 + e^{-2t} \\ 2 - e^{-2t} \end{bmatrix} \rightarrow \begin{bmatrix} 2 \\ 2 \end{bmatrix} \text{ при } t \rightarrow \infty.$$

5.4.4. Замена дает  $dv/dt = \Lambda v$ ,  $v = e^{\Lambda t} v_0 = e^{\Lambda t} S^{-1} u_0$ , т. е.  $u = Se^{\Lambda t} S^{-1} u_0$ , как и ранее.

$$5.4.5. e^t A e^s A = (Se^t \Lambda S^{-1})(Se^s \Lambda S^{-1}) = Se^t \Lambda e^s \Lambda S^{-1} = Se^{(t+s)} \Lambda S^{-1} = e^{(t+s)} A.$$

$$5.4.6. \frac{d}{dt} (x_1 y_1 + \dots) = \frac{dx_1}{dt} y_1 + x_1 \frac{dy_1}{dt} + \dots = \left( \frac{dx}{dt} \right)^T y + x^T \left( \frac{dy}{dt} \right).$$

5.4.7. Матрица  $A$  имеет мнимые собственные значения, так что система нейтрально устойчива.

5.4.8. Собственные значения матрицы  $A$  равны  $\lambda_1 = \lambda_2 = -1$ , что дает убывание при  $t \rightarrow \infty$ ; собственные значения матрицы  $A^T + A$  равны  $\lambda_1 = 1$ ,  $\lambda_2 = -5$ , так что при  $t=0$  может не быть немедленного убывания решения.

5.4.9. Система является неустойчивой.

$$5.4.10. u = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \cos t + \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \frac{\sin \sqrt{3}t}{\sqrt{3}}.$$

$$5.4.11. u = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cos \sqrt{6}t + \frac{1}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \cos 2t.$$

$$5.4.12. \rho^2 e^{\rho t} x = -\rho F e^{\rho t} x + A e^{\rho t} x \text{ или } (A - \rho F - \rho^2 I) x = 0.$$

$$5.4.14. u = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \sin t & 0 \\ 0 & (\sin \sqrt{3}t)/\sqrt{3} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \\ = \frac{1}{2} \begin{bmatrix} \sin t + (\sin \sqrt{3}t)/\sqrt{3} \\ \sin t - (\sin \sqrt{3}t)/\sqrt{3} \end{bmatrix}.$$

5.5.1. Все числа лежат вне единичного круга.

5.5.2. Сумма вещественна; также лежит на единичной окружности; также лежит на единичной окружности; лежит внутри или на границе круга радиуса два

$$5.5.4. \|x\|=6=\|y\|, x^H y = -12.$$

$$5.5.5. C = \begin{bmatrix} 1 & -i \\ -i & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & i & 0 \\ i & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & i & -i \\ -i & 1 & 0 \\ i & 0 & 1 \end{bmatrix}, \quad C = C^H, \quad \text{поскольку}$$

$$(A^H A)^H = A^H A^{HH} = A^H A.$$

5.5.6.  $\begin{bmatrix} 1 & i & 0 \\ i & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & i & 0 \\ 0 & 1 & 1 \end{bmatrix} = U$ ;  $Ax=0$ , если  $x$  является кратным вектору  $\begin{bmatrix} i \\ -1 \\ 1 \end{bmatrix}$ ; этот вектор ортогонален к столбцам матрицы  $A^H$  (но не к столбцам матрицы  $A^T$ , т. е. строкам матрицы  $A$ ).

$$5.5.7. A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad x = \begin{bmatrix} 1 \\ i \end{bmatrix}, \quad x^H A x = i.$$

5.5.8.  $\lambda_1=1, \lambda_2=-1, x_1 = \begin{bmatrix} 1 \\ -i \end{bmatrix}, x_2 = \begin{bmatrix} 1 \\ i \end{bmatrix}, A = x_1 x_1^H - x_2 x_2^H$ ; произведение  $x_1^H x_2$  равно нулю, поскольку собственные векторы ортогональны (свойство 3).

$$5.5.9. U = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}, \quad A = 0 + \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} +$$

$$+ 3 \begin{bmatrix} \frac{1}{6} & -\frac{2}{6} & \frac{1}{6} \\ -\frac{2}{6} & \frac{4}{6} & -\frac{2}{6} \\ \frac{1}{6} & -\frac{2}{6} & \frac{1}{6} \end{bmatrix}.$$

5.5.10. Определитель равен произведению собственных значений.

5.5.12. Единица не является собственным значением, поскольку она вещественна.

5.5.13.  $K = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, x = \begin{bmatrix} 4 \\ 5 \end{bmatrix}, x^H K x = 0$  (оно должно быть мнимым в силу  $1'$  и вещественным, поскольку  $K$  и  $x$  вещественны; следовательно, оно равняется нулю).

$$5.5.14. K = (Z - Z^H)/2 = -K^H; Z = A + K,$$

$$\begin{bmatrix} 3+i & 4+2i \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 2+i \\ 2-i & 5 \end{bmatrix} + \begin{bmatrix} i & i \\ i & 0 \end{bmatrix}.$$

$$5.5.15. (UV)^H (UV) = V^H U^H U V = V^H V = I.$$

5.5.16. Определитель равен произведению собственных значений, и модуль произведения  $\lambda_1 \dots \lambda_n$  есть  $|\lambda_1| \dots |\lambda_n| = 1$ ; матрица первого порядка имеет вид  $U = [i]$ ;

$$U = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{i\psi} \end{bmatrix} \quad \text{для произвольных углов } \theta \text{ и } \psi.$$

5.5.17. Последний столбец равен  $(1, -2, 1)/\sqrt{6}$ , умноженному на любое число (вещественное или комплексное), по модулю равное единице.

$$5.5.18. \Lambda = \begin{bmatrix} 0 & & \\ & 2i & \\ & & \end{bmatrix}, \quad S = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \quad e^{Kt} = S e^{\Lambda t} S^{-1} = \\ = \frac{1}{2} \begin{bmatrix} 1 + e^{2it} & -1 + e^{2it} \\ -1 + e^{2it} & 1 + e^{2it} \end{bmatrix}; \quad \text{при } t=0 \quad \frac{de^{Kt}}{dt} = \begin{bmatrix} i & i \\ i & i \end{bmatrix} = K.$$

5.5.19. Диагональ матрицы  $A$  состоит из чисел  $+1$  или  $-1$ ; имеется восемь возможностей.

5.6.1.  $C = (M')^{-1} B M' = (M')^{-1} M^{-1} A M M'$  или  $C = (M M')^{-1} A (M M')$ . Единственная матрица, подобная  $I$ , есть  $M^{-1}/M = I$ .

5.6.2. Элемент в позиции  $(3, 1)$  имеет вид  $g \cos \theta + h \sin \theta$  и равен нулю при  $\operatorname{tg} \theta = -g/h$ .

5.6.3. Положить  $M$  равной  $A$ . Тогда  $M^{-1} (AB) M = BA$  подобна матрице  $AB$ . Если собственные значения совпадают, то будут совпадать также их суммы, т. е. след. Если  $AB$  и  $BA$  имеют одинаковый след, то матрица  $AB - BA$  имеет нулевой след и не может равняться  $I$ .

5.6.4.  $\lambda_1 = -1$  с собственным вектором  $(1, 1)^T/\sqrt{2}$  единичной длины; он образует первый столбец матрицы  $U$ , а второй столбец  $(-1, 1)^T/\sqrt{2}$  делает  $U$  унитарной; тогда  $T = U^{-1} A U = \begin{bmatrix} -1 & -7 \\ 0 & 2 \end{bmatrix}$ . Возможны и другие матрицы  $T$ , имеющие элемент  $7$  вместо  $-7$  и (или) с переставленными местами элементами  $2$  и  $-1$ .

$$5.6.5. T = \begin{bmatrix} 0 & \pm 2 \\ 0 & 0 \end{bmatrix}.$$

$$5.6.6. p(T) = \begin{bmatrix} 0 & * & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix} \begin{bmatrix} * & * & * \\ 0 & 0 & * \\ 0 & 0 & * \end{bmatrix} \begin{bmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & 0 \end{bmatrix} = 0, \quad \text{так что } P(A) = 0.$$

$$5.6.7. U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -2/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \end{bmatrix};$$

первые два столбца могут иметь вид  $x_1 = (1, -1, 0)^T/\sqrt{2}$ ,  $x_2 = (1, 1, -2)^T/\sqrt{6}$  (имеется много других возможностей); всегда

$$x_1 x_1^H + x_2 x_2^H = I - x_3 x_3^H = \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

5.6.8. (i)  $T T^H = U^{-1} A U U^H A^H (U^{-1})^H = I$ ; (ii) если матрица  $T$  является треугольной и унитарной, то ее диагональные элементы (собственные значения) по модулю должны равняться единице. Тогда, если столбцы должны быть единичными векторами, все внедиагональные элементы должны равняться нулю.

$$5.6.9. A = \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}.$$

5.6.11. Если  $N = U \Lambda U^{-1}$ , то  $N N^H = (U \Lambda U^{-1}) (U \Lambda U^{-1})^H = U \Lambda \Lambda^H U^H = U \Lambda \Lambda U^H = (U \Lambda U^{-1}) U (U \Lambda U^{-1}) = N^H N$ .

5.6.12.  $B M = M J_2$  выполняется при  $m_{11} = m_{32}$ ,  $m_{21} = m_{31} = m_{33} = 0$ .

5.6.13.  $M^{-1} J_3 M = 0$ , так что последние два неравенства очевидны;  $M J_1 = J_2 M$  означает равенство нулю первого столбца матрицы  $M$ , так что она не может быть обратимой.

5.6.14.  $J_1$  равна клетке размера  $4 \times 4$ ;  $J_2$  состоит из клеток размера  $3 \times 3$  и  $1 \times 1$ ;  $J_3$  — из клеток размера  $2 \times 2$  и  $2 \times 2$ ;  $J_4$  — из клеток размера  $2 \times 2$ ,  $1 \times 1$  и  $1 \times 1$ ;  $J_5$  состоит из четырех одномерных клеток, равных 0. Матрицы  $J_2$  и  $J_3$  имеют по два собственных вектора.

## Глава 6

6.1.1.  $ac - b^2 = 2 - 4 = -2$ ;  $f = (\sqrt{2}x + \sqrt{2}y)^2 - y^2$ .

6.1.2. Нет, нет, нет, да (неопределенная, положительно полуопределенная, неопределенная, положительно определенная).

6.1.4. Матрицы вторых производных имеют вид  $\begin{bmatrix} 4 & -5 \\ -5 & 12 \end{bmatrix}$  (минимум),

$\begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}$  (максимум).

6.2.1. Полуопределенная, полуопределенная, положительно определенная:  $x^T Bx = (u+w)^2 + v^2$ .

6.2.2.  $\det A = 1 + 2\alpha^2\beta - \beta^2 - 2\alpha^2 < 0$ , если  $\alpha = -\beta = -2/3$ .

6.2.3. Если матрица  $A$  имеет положительные собственные значения  $\lambda_i$ , то собственные значения матрицы  $A^2$  равны  $\lambda_i^2$ , а собственные значения матрицы  $A^{-1}$  равны  $1/\lambda_i$  и они также положительны; воспользоваться условием II.

6.2.4. Если  $x^T Ax > 0$  и  $x^T Bx > 0$ , то  $x^T (A+B)x > 0$ ; условие I.

6.2.5.  $A = \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 2 \\ \frac{3}{2} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \times$   
 $\times \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$ , так что  $W = \sqrt{DL}^T = \begin{bmatrix} \sqrt{2} & -1/\sqrt{2} \\ 0 & \sqrt{3}/\sqrt{2} \end{bmatrix}$  или  
 $W = \sqrt{\Lambda} Q^T = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ \sqrt{3}/\sqrt{2} & -\sqrt{3}/\sqrt{2} \end{bmatrix}$ .

6.2.6.  $R = \frac{1}{2} \begin{bmatrix} 1 + \sqrt{3} & 1 - \sqrt{3} \\ 1 - \sqrt{3} & 1 + \sqrt{3} \end{bmatrix}$ ,  $R^T = Q (\sqrt{\Lambda})^T Q^T = R$ ,

$R^2 = Q \sqrt{\Lambda} Q^T Q \sqrt{\Lambda} Q^T = Q \Lambda Q^T = A$  (поскольку  $Q^T Q = I$ ).

6.2.8.  $x^H A x = x^H (\alpha + i\beta) x$ , и если вещественная часть левого произведения положительна, то будет положительна и вещественная часть  $x^H \alpha x$ ; следовательно,  $\alpha > 0$ . С другой стороны, матрица  $A = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$  имеет положительные собственные значения, но матрица  $A^H + A$  не является положительно определенной.

6.2.9. Матрица  $A$  является отрицательно определенной, если  $x^T A x < 0$  для всех ненулевых векторов  $x$ ; все  $\lambda_i < 0$ ; знаки определителей матриц  $A_k$  чередуются (а не  $\det A_k < 0!$ ); отрицательные ведущие элементы;  $A = -W^T W$ .

6.2.10.  $\lambda_1 = 1$ ,  $\lambda_2 = 4$ , полуоси проходят через  $(\pm 1, 0)$  и  $(0, \pm 1/2)$ .

6.2.11.  $A = \begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix}$  с собственными значениями 1 и 4.

6.2.12. Одно нулевое собственное значение превращает эллипсоид в бесконечный цилиндр  $\lambda_1 y_1^2 + \lambda_2 y_2^2 = 1$ , параллельный третьей оси; два нулевых собственных значения дают две плоскости  $y_1 = \pm 1/\sqrt{\lambda_1}$ ; три нулевых собственных значения приводят к равенству  $0 = 1$  (рисунка не существует).

$$6.3.1. W = \sqrt{DL^T} = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & -1 & -1 \\ 0 & \sqrt{3} & -\sqrt{3} \\ 0 & 0 & 0 \end{bmatrix};$$

$$x^T A x = \left( \sqrt{2}u - \frac{\sqrt{2}}{2}v - \frac{\sqrt{2}}{2}w \right)^2 + \left( \frac{\sqrt{3}}{\sqrt{2}}v - \frac{\sqrt{3}}{\sqrt{2}}w \right)^2;$$

поверхность представляет собой цилиндр (см. упр. 6.2.12).

6.3.3. Матрица  $C^T A C = \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}$  имеет, как и сама  $A$ , положительные и нулевые собственные значения;  $C(t) = \begin{bmatrix} 2-t & 0 \\ 0 & 1 \end{bmatrix}$ .

6.3.4. Последний ведущий элемент матрицы  $A$  является отрицательным, но все ведущие элементы матрицы  $A + 2I$  положительны.

$$6.3.5. [\sqrt{3}-1 \ 1] \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1+\sqrt{3} \\ -1 \end{bmatrix} = 0; \quad u_0 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad a_1 = \frac{1}{2},$$

$a_2 = -1/2$ ; меньшая масса смещается до положения  $\sqrt{3}$ , а большая масса никогда не выходит за пределы своего начального смещения 1.

$$6.3.6. \lambda_1 = 54, \quad \lambda_2 = \frac{54}{5}, \quad x_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

$$6.3.7. A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad |A - \lambda B| = -\lambda^2 - 1, \quad \lambda = \pm i.$$

$$6.4.1. P = x_1^2 - x_1 x_2 + x_2^2 - x_2 x_3 + x_3^2 - 4x_1 - 4x_3;$$

$$\partial P / \partial x_1 = 2x_1 - x_2 - 4, \quad \partial P / \partial x_2 = -x_1 + 2x_2 - x_3,$$

$$\partial P / \partial x_3 = -x_2 + 2x_3 - 4.$$

6.4.2.  $\partial P_1 / \partial x = x + y = 0$ ,  $\partial P_1 / \partial y = x + 2y - 3 = 0$ ,  $x = -3$ ,  $y = 3$ .  $P_2$  не имеет минимума (пусть  $y \rightarrow \infty$ ) и соответствует неопределенной матрице  $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ .

6.4.3. Минимизация  $Q$  приводит к нормальным уравнениям  $A^T A x = A^T b$ .

6.4.4.  $R(x) = a_{11}$ , так что  $\lambda_1 = \min R(x) \leq a_{11}$ ;  $x = e_i$  дает  $R(x) = a_{ii} \geq \lambda_1$ .

6.4.5. Минимальное значение есть  $R(x) = 1$  с  $x = (1, 1)$ .

6.4.6. Поскольку  $x^T B x > 0$  для всех ненулевых векторов  $x$ , число  $x^T(A+B)x$  будет больше, чем  $x^T A x$ .

$$6.4.7. \text{Если } (A+B)x = \theta_1 x, \text{ то } \lambda_1 + \mu_1 \leq \frac{x^T A x}{x^T x} + \frac{x^T B x}{x^T x} = \theta_1.$$

6.4.8. Максимум равен  $3/2$ .

6.4.9. Поскольку  $\lambda_1 = \lambda_2 = 1$ ,  $\mu_1$  должно равняться единице; футбол (регби).

6.4.10. Когда две строки и два столбца исключаются,  $\lambda_{\min} \leq$  исходного  $\lambda_3$ .

6.4.11.  $\mu_2 = 3$ ,  $2 < 3 < 2 + \sqrt{2}$ .

6.4.12. Предельное подпространство  $S_j$  натянуто на первые  $j$  собственных векторов.

6.4.13. Максимум берется по подпространству  $S_j$ , ортогональному вектору  $z$ .

6.4.14. Положить  $x = (1, 0, \dots, 0)^T$ , тогда  $\lambda_1 \leq R(x) = a_{11}/b_{11}$ .

$$6.5.1. b_j = h \text{ (площадь под } V_4 \text{ на рис. 6 б); система } Ay = b \text{ принимает вид}$$

$$\frac{-y_{j-1} + 2y_j - y_{j+1}}{h} = \frac{-(x-h) + (x-h)^2 + 2x - 2x^2 - (x+h) + (x+h)^2}{2h} = h.$$

$$6.5.2. P(y) - P(b) = \frac{1}{2} y^T y - y^T b + \frac{1}{2} b^T b = \frac{1}{2} \|y - b\|^2; \text{ минимизируя } P(y)$$

на пробных функциях, мы одновременно минимизируем  $P(y) - P(b)$ , т. е. расходясь до  $b$ .

$$6.5.3. \lambda_1 = 54/5 > \lambda_1 = \pi^2.$$

6.5.4. Из принципа минимакса следует, что  $\lambda_2$  является минимальным значением для  $\lambda_{\max}$  на всех двумерных подпространствах;  $\lambda_2$  увеличивает этот минимум, поскольку соответствующие двумерные подпространства должны лежать внутри пространства пробных функций.

## Глава 7

7.2.1. Если матрица  $Q$  является ортогональной, то  $\|Q\| = \max \|Qx\| / \|x\| = 1$ , так как  $Q$  сохраняет длины:  $\|Qx\| = \|x\|$  для каждого  $x$ . Матрица  $Q^{-1}$  также является ортогональной, и ее норма равна единице, так что  $c(Q) = 1$ .

7.2.2. Неравенство треугольника для векторов дает  $\|Ax + Bx\| \leq \|Ax\| + \|Bx\|$ , и, когда мы разделим на  $\|x\|$  и максимизируем каждый член, то получим неравенство треугольника для матричных норм.

7.2.3 По определению нормы матрицы  $A$  имеем  $\|ABx\| \leq \|A\| \|Bx\|$ , откуда  $\|Bx\| \leq \|B\| \|x\|$ . Деля на  $\|x\|$  и максимизируя, приходим к неравенству  $\|AB\| \leq \|A\| \|B\|$ . То же самое справедливо и для обратных матриц,  $\|B^{-1}A^{-1}\| \leq \|B^{-1}\| \|A^{-1}\|$ . Перемножая эти неравенства, получаем  $c(AB) \leq c(A)c(B)$ .

$$7.2.4. \|A^{-1}\| = 1, \|A\| = 3, c(A) = 3; \text{ взять } b = x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \delta b = x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

7.2.5. По определению нормы имеем  $\|A\| = \max \|Ax\| / \|x\|$ ; если выбрать в качестве  $x$  рассматриваемый собственный вектор, то  $\|Ax\| = |\lambda| \|x\|$  и отношение равно  $|\lambda|$ .

$$7.2.6. A^T A = \begin{bmatrix} 1 & 100 \\ 100 & 10001 \end{bmatrix}, \lambda^2 - 10002\lambda + 1 = 0, \lambda_{\max} = 5001 + (5001^2 - 1)^{1/2}.$$

Норма равняется квадратному корню из этой величины, и она совпадает с  $\|A^{-1}\|$ .

7.2.7. Матрицы  $A^T A$  и  $A A^T$  имеют одинаковые собственные значения (даже в том предельном случае, когда матрица  $A$  является вырожденной), и равенство максимальных собственных значений дает  $\|A\| = \|A^T\|$ .

7.2.8. Поскольку  $A = W^T W$  и  $A^{-1} = W^{-1} (W^T)^{-1}$ , имеем  $\|A\| = \|W\|^2$  и  $\|A^{-1}\| = \|(W^T)^{-1}\|^2 = \|W^{-1}\|^2$ . (Согласно предыдущему упражнению, транспонированная матрица имеет ту же самую норму.) Поэтому  $c(A) = (c(W))^2$ .

$$7.2.9. \begin{bmatrix} 0.01 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 100 & 1 \end{bmatrix} \begin{bmatrix} 0.01 & 1 \\ 0 & -100 \end{bmatrix}; c(A) \text{ близко к } 1, \text{ в то время как } c(L) \text{ и } c(U) \text{ близки к } 10^4.$$

7.2.10. Если  $x = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ , то  $Ax = \begin{bmatrix} -1 \\ 7 \end{bmatrix}$  и  $\|Ax\|_{\infty} / \|x\|_{\infty} = 7$ ; это предельный случай, так что  $\|A\|_{\infty} = 7$ .

$$7.3.1. u_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, u_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, u_2 = \begin{bmatrix} 5 \\ -4 \end{bmatrix}, u_3 = \begin{bmatrix} 14 \\ -13 \end{bmatrix}, u_{\infty} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

$$7.3.2. u_0 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, u_1 = \frac{1}{3} \begin{bmatrix} 10 \\ 11 \end{bmatrix}, u_2 = \frac{1}{9} \begin{bmatrix} 31 \\ 32 \end{bmatrix}, u_3 = \frac{1}{27} \begin{bmatrix} 94 \\ 95 \end{bmatrix}; \text{ со сдвигом}$$

$$\alpha = \frac{26}{25} \text{ и } u_1 = \frac{25}{49} \begin{bmatrix} 24 & 25 \\ 25 & 24 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} \approx \begin{bmatrix} 172 \\ 171 \end{bmatrix}.$$

7.3.3.  $Hx = x - (x-y) \frac{2(x-y)^T x}{(x-y)^T (x-y)} = x - (x-y) = y$ . Для доказательства равенства  $Hu = x$  следует либо поменять местами  $x$  и  $y$ , либо рассмотреть равенство  $H(Hx) = Hu$  и воспользоваться тем, что  $H^2 = I$ .

$$7.3.4. \sigma = 5, \quad v = \begin{bmatrix} 8 \\ 4 \end{bmatrix}, \quad H = \frac{1}{5} \begin{bmatrix} -3 & -4 \\ -4 & 3 \end{bmatrix}.$$

$$7.3.5. U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{3}{5} & -\frac{4}{5} \\ 0 & -\frac{4}{5} & \frac{3}{5} \end{bmatrix} = U^{-1}, \quad U^{-1}AU = \begin{bmatrix} 1 & -5 & 0 \\ -5 & \frac{9}{25} & \frac{12}{25} \\ 0 & \frac{12}{25} & \frac{16}{25} \end{bmatrix}.$$

$$7.3.6. \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = Q_0 R_0 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} 5 & -4 \\ 0 & 3 \end{bmatrix}, \quad A_1 = R_0 Q_0 = \\ = \frac{1}{5} \begin{bmatrix} 14 & -3 \\ -3 & 6 \end{bmatrix}.$$

$$7.3.7. \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & 0 \end{bmatrix} = QR = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & \cos \theta \sin \theta \\ 0 & -\sin^2 \theta \end{bmatrix}; \\ RQ = \begin{bmatrix} \cos \theta (1 + \sin^2 \theta) & -\sin^3 \theta \\ -\sin^3 \theta & -\sin^2 \theta \cos \theta \end{bmatrix}.$$

7.3.8. Матрица  $A$  является ортогональной, так что  $Q = A$ ,  $R = I$  и вновь  $RQ = A$ .

7.3.9. Пусть  $(Q_0 \dots Q_{k-1})(R_{k-1} \dots R_0)$  является  $QR$ -разложением матрицы  $A^k$ , что во всяком случае справедливо, когда  $k=1$ . По построению  $A_{k+1} = R_k Q_k$ , т. е.  $R_k = A_{k+1} Q_k^T = Q_k^T \dots Q_0^T A Q_0 \dots Q_k Q_k^T$ . Умножая справа на  $(R_{k-1} \dots R_0)$  и пользуясь сделанным предположением, получаем  $R_k \dots R_0 = Q_k^T \dots Q_0^T A^{k+1}$ . Переноса матрицы  $Q$  в левую часть, приходим к требуемому результату для матрицы  $A^{k+1}$ .

$$7.4.1. D^{-1}(-L-U) = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 \end{bmatrix}, \quad \mu = 0, \pm \frac{1}{\sqrt{2}};$$

$$(D+L)^{-1}(-U) = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{8} & \frac{1}{4} \end{bmatrix},$$

собственные значения  $0, 0, \frac{1}{2}$ ;  $\omega_{\text{опт}} = 4 - 2\sqrt{2}$  с  $\lambda_{\text{max}} = 3 - 2\sqrt{2} \approx 0.2$ .

7.4.2. Матрица  $J$  имеет числа  $1/2$  на диагонали, прилегающей к главной, и нули на всех остальных местах;  $Jx_1 = \frac{1}{2}(\sin 2\pi h, \sin 3\pi h + \sin \pi h, \dots) = (\cos \pi h)x_1$ .

$$7.4.3. Jx_k = \frac{1}{2} (\sin 2k\pi h, \sin 3k\pi h + \sin k\pi h, \dots) = (\cos k\pi h) x_k.$$

7.4.4. Окружность с центром в точке  $a_{ii}$  не может достигнуть нуля, если ее радиус  $r_i$  меньше, чем  $|a_{ii}|$ ; следовательно, нуль не является собственным значением и матрица с диагональным преобладанием не может быть вырожденной.

$$7.4.5. J = - \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{4} \\ \frac{2}{5} & \frac{2}{5} & 0 \end{bmatrix};$$

радиусы равны  $r_1 = \frac{2}{3}$ ,  $r_2 = \frac{1}{4}$ ,  $r_3 = \frac{4}{5}$ , центры окружностей находятся в нуле, так что все  $|\lambda_i| < 1$ .

## Глава 8<sup>1)</sup>

8.1.1. Углы находятся в точках (0, 6), (2, 2), (6, 0); см. рис. 8.4.

8.1.2.  $x+y$  минимальна в точке (2, 2), и стоимость равна 4;  $3x+y$  минимальна в точке (0, 6), где стоимость равна 6; минимум для  $x-y$  равен  $-\infty$ , когда  $x=0$  и  $y \rightarrow \infty$ .

8.1.3. Ограничения дают  $3(2x+5y)+2(-3x+8y) \leq 9-10$ , т. е.  $31y \leq -1$ , что противоречит условию  $y \geq 0$ .

8.1.4. Взять  $x$  и  $y$  равными и достаточно большими.

8.1.5. Неравенства  $x \geq 0$ ,  $y \geq 0$ ,  $x+y \leq 0$  допускают лишь точку (0, 0).

8.1.6. Допустимое множество является равносторонним треугольником в плоскости  $x+y+z=1$  с углами в точках  $(x, y, z) = (1, 0, 0)$ ,  $(0, 1, 0)$ ,  $(0, 0, 1)$ ; последний угол дает максимальное значение, равное 3.

8.2.3.  $r = [1 \ 1]$ , так что угол является оптимальным.

8.2.4.  $r = [3 \ -1]$ , так что второй столбец матрицы  $F$  включается в базис; этот столбец есть  $u = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$ , и вектор  $B^{-1}u = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$  является отрицательным, так что ребро имеет бесконечную длину и минимальная стоимость равна  $-\infty$ .

8.2.5. В точке  $P$  имеем  $r = [-5 \ 3]$ ; в точке  $Q$  имеем  $r = [5/3 \ -1/3]$ ; в точке  $R$   $r \geq 0$ .

8.2.6. (а) Пара  $x=0$ ,  $w=b$  неотрицательна, удовлетворяет уравнению  $Ax+w=b$  и является допустимой, так как  $x=0$  добавляет  $n$  нулевых компонент. (б) Дополнительная задача состоит в минимизации  $w_1$  при условиях  $x_1 \geq 0$ ,  $x_2 \geq 0$ ,  $w_1 \geq 0$ ,  $x_1 - x_2 + w_1 = 3$ . Вектор первой фазы имеет вид  $x_1 = x_2 = 0$ ,  $w_1 = 3$ ; здесь оптимальным вектором является  $x_1^* = 3$ ,  $x_2^* = w_1^* = 0$ . Угол находится в точке  $x_1 = 3$ ,  $x_2 = 0$ , и допустимое множество представляет собой прямую, проходящую через эту точку и имеющую наклон 1.

8.2.7. Критерий останова принимает вид  $r \leq 0$ ; если он не выполняется и  $i$ -я компонента является максимальной, то этот столбец матрицы  $F$  включается в базис; правило 8С для вектора, исключаемого из базиса, не изменяется.

8.3.1. Максимизировать  $4y_1 + 11y_2$  при условиях  $y_1 \geq 0$ ,  $y_2 \geq 0$ ,  $2y_1 + y_2 \leq 1$ ,  $3y_2 \leq 1$ ;  $x_1^* = 2$ ,  $x_2^* = 3$ ,  $y_1^* = \frac{1}{3}$ ,  $y_2^* = \frac{1}{3}$ , стоимость  $= 5$ .

<sup>1)</sup> В связи с тем, что автором был прислан для перевода на русский язык дополненный вариант гл. 8, решения некоторых упражнений отсутствуют.

8.3.2 Минимизировать  $3x_1$  при условиях  $x_1 \geq 0$ ,  $x_1 \geq 1$ ;  $y_1^* = 0$ ,  $y_2^* = 3$ ,  $x_1^* = 1$ , стоимость = 3.

8.3.3. В двойственной задаче мы максимизируем  $y^T b$  при условии  $ly \geq c^T$ ; векторы  $y = c$  и  $x = b$  являются допустимыми и дают одно и то же значение  $cb$ , так что, согласно 8F, они должны быть оптимальными.

8.3.4.  $A = [-1]$ ,  $b = [1]$ ,  $c = 0$  являются недопустимыми; в двойственной задаче мы максимизируем  $y$  при условиях  $y \geq 0$  и  $-ly \leq 0$ , и эта задача является неограниченной.

$$8.3.5. b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, c = [-1 \ 0].$$

8.3.6. Если вектор  $x$  является очень большим, то  $Ax \geq b$  и  $x \geq 0$ ; если  $y = 0$ , то  $yA \leq c$  и  $y \geq 0$ . Следовательно, обе задачи являются допустимыми.

8.3.7. Поскольку  $cx = 3 = y^T b$ ,  $x$  и  $y$  являются оптимальными в силу 8F.

8.3.8.  $Ax = [1 \ 1 \ 3 \ 1]^T \geq b = [1 \ 1 \ 1 \ 1]^T$  со строгим неравенством для третьей компоненты; следовательно, третья компонента должна быть нулем. Аналогично  $yA = [1 \ 1 \ 1 \ 1] \leq c = [1 \ 1 \ 1 \ 3]$  и строгое неравенство приводит к  $x_4 = 0$ .

8.3.9.  $x^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = y^*$  с  $(y^*)b = 1 = cx^*$ ; вторые компоненты  $Ax^* \geq b$  и  $y^*A \leq c$  удовлетворяют строгим неравенствам, давая нулевые вторые компоненты в  $y^*$  и  $x^*$ .

8.3.10.  $Ax = b$  приводит к  $yAx = yb$  независимо от того, удовлетворяется ли неравенство  $y \geq 0$ .  $yA \leq c$  приводит, как и выше, к  $yAx \leq cx$ , но только в силу неотрицательности  $x \geq 0$ . Сравнивая, имеем  $yb \leq cx$ .

8.3.13. Столбцы образуют конус между положительной частью оси  $x$  и прямой  $x = y$ . В первом случае  $x = (1, 2)^T$ ;  $y = (1, 1)^T$  удовлетворяет выбранной возможности.

8.3.14. Столбцы матрицы

$$\begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}.$$

8.3.15. Положить  $y = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ ; тогда  $A^T y = 0$ ,  $y^1 b \neq 0$ .

8.3.16. Положить  $y = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ ; тогда  $A^T y \geq 0$ ,  $y^T b < 0$ .

8.3.17.  $A^T y \geq 0$  приводит к  $y^T A \geq 0$ , откуда  $y^T Ax \geq 0$ ; с другой стороны,  $Ax \geq b$  приводит к  $y^T Ax \leq y^T b < 0$ .

## Приложение А

$$A.1 \ [A] = \begin{bmatrix} 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 12 \end{bmatrix}.$$

A.2.  $[A][A]^T = I$ , но элемент в позиции (1,1) матрицы  $[A]^T[A]$  равен нулю. Первая компонента вектора  $v$  аннулируется левым сдвигом подобно тому, как это происходит с константами при дифференцировании.

А.3. Транспонирование оставляет  $v_1$  и  $v_4$  и меняет местами  $v_2$  и  $v_3$ :

$$[A] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

$[A]^2 = I$ , поскольку транспонированная к транспонированной совпадает с исходной матрицей.

$$\text{А.4. } [A] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix}.$$

А.5. Любое решение является многочленом степени  $n-1$ , так что  $S_n = P_{n-1}$  с базисом  $1, t, \dots, t^{n-1}$ .

А.6.  $[A] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ ,  $[B] = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ ,  $[BA] = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$  равно зеркальному отражению относительно оси  $y$ .

А.7.  $[A] = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ ; все собственные значения матрицы проектирования равны 0 или 1.

$$\text{А.8. } [A] = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & \frac{4}{3} \\ 0 & 4 \end{bmatrix}.$$

## Приложение В

$$\text{В.1. } J = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}, \quad J = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

$$\text{В.2. } \frac{du_2}{dt} = 8e^{st}(tx_1 + x_2) + e^{st}x_1,$$

$$Au_2 = 8e^{st}(tAx_1 + Ax_2) = e^{st}(8tx_1 + 8x_2 + x_1).$$

$$\text{В.3. } e^{Bt} = \begin{bmatrix} 1 & t & 2t \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = I + Bt, \quad \text{поскольку } B^2 = 0.$$

## Указатель

- Алгебраическая кратность 223  
 Алгебраическое дополнение 196, 197  
 Алгоритм Кроута 37  
 Альтернатива Фредгольма 111  
 Альтернативная теорема 384, 386  
 Анализ основных компонент 261, 292
- Базис** 82  
 Базисное решение 364  
 Базисные переменные 71, 364  
 Бесконечномерное пространство 159, 408  
 Блочное умножение матриц 205  
 Блочный степенной метод 333  
 Бридж 404  
 Булева алгебра 186  
 Буняковский 130
- Вандермонд** 99, 191  
 Ведущий элемент 13, 34, 50, 56, 184, 289  
 — — в симплекс-методе 366  
 — — формула 204, 205
- Вектор** 18  
 — невязки 136  
 Векторное пространство 62—64  
 Вес 63, 139, 175  
 Вещественное пространство 65  
 Взаимно однозначное преобразование 97, 409  
 Взвешенная длина 176, 178  
 Взвешенное скалярное произведение 178  
 — среднее значение 175  
 Взвешенные наименьшие квадраты 175  
 Волновое уравнение 250
- Выбор портфеля ценных бумаг 359  
 Вырожденный угол 366
- Галуа** 218  
 Гармоники (гармонические составляющие) 216, 230, 243, 249  
 Гаусс 11, 13, 31, 43, 48, 74, 206  
 Гейзенберг 269  
 Гейл 386, 403  
 Гены 235  
 Геометрическая кратность 223  
 Гершгорин 351  
 Гессе 285  
 Гивенс 301  
 Гильберт 51, 56, 159, 161, 269, 408, 409  
 Гильбертово пространство 159, 269, 408, 409  
 Главная подматрица 296  
 Голуб 350  
 Грузовые перевозки 234  
 Гурвиц 246
- Данциг** 361  
 Двойственная задача 357, 375  
 Двухдиагональная матрица 57  
 Двухточечная краевая задача 53  
 Дефект 89, 120  
 Дефектная матрица 216, 223, 275  
 Диагонализация 221, 258, 272, 303, 314, 415  
 Диагональная матрица 37, 170  
 Диагональное преобладание 352  
 Дилемма заключенного 405  
 Динамическое программирование 394  
 Дисперсия 146

- Дифференциальное уравнение 52, 211, 239, 421  
 Диффузия 242, 244, 245  
 Длина 101, 159, 254, 265  
 — взвешенная 76, 178  
 Допустимое множество 355, 360, 376,  
 Допустимый вектор 355, 377, 379  
 Достижимый вектор 384  
 Дуги сети 388
- Единичная матрица 29  
 Единичный вектор 103  
 — — в направлении оси координат 78  
 Единственность решения 95
- Жордан 45, 75, 275, 416, 419  
 Жорданова форма 275, 416, 419
- Задача об оптимальном назначении 394  
 — о бракосочетании 394  
 — — диете 357, 376  
 — — коммивояжере 394  
 — — максимальном потоке 388  
 — — перевозках 359, 395  
 — — простом назначении 392  
 Закон ассоциативности 27  
 — дистрибутивности 28  
 — инерции 298  
 — Кирхгофа 111  
 — Ньютона 247, 301  
 — Ома 112  
 — Сильвестра 298  
 Замена базиса 414  
 — переменных 244, 268  
 Заикливание 366  
 Зейдель 345  
 Значение программы 355
- Идемпотентная матрица 140  
 Изоморфизм 117  
 Исключение с частичным выбором ведущего элемента 51  
 Исправленный симплекс-метод 372  
 Исходная задача 357, 374  
 Итеративное уточнение 352  
 Итерационные методы 322, 343
- Калифорния 22, 232, 333  
 Квадратичная форма 281, 283, 285  
 Квадратный корень 227, 291  
 Квантовая механика 225, 269
- Кирхгоф 111  
 Коммутирование матриц 29, 133, 225  
 — — контрпример 28  
 Комплексное число 216, 251  
 Комплексно-сопряженное число 252  
 Комплексный вектор 253, 267  
 Конечно-разностная матрица 59, 191, 198, 319  
 Консервативная система 246  
 Контур 111  
 Конус 384, 387  
 Корреляционная матрица 261  
 Косинус 128, 130, 161  
 Кососимметрическая матрица 190, 401  
 Косозермитова матрица 263, 264  
 Коши 130  
 Коэффициент корреляции 177  
 — регрессии 175  
 Коэффициенты влияния 184  
 Краевая задача 53  
 Крамер 185, 201  
 Критерий останова 368, 381  
 Кроут 37  
 Кун 379, 403
- Лагранж 378  
 Левая обратная матрица 44, 94, 138, 411  
 Левое нуль-пространство 92  
 Левый нуль-вектор 92  
 Лежандр 162, 164  
 Лемма Шура 270  
 Ленбергер 373  
 Ленточная матрица 52, 58  
 Леонтьев 238, 239  
 Линейная зависимость 77  
 — комбинация 22, 63, 77  
 — независимость 77, 90, 224  
 — оболочка 81  
 Линейное преобразование 97, 407, 410  
 — программирование 353  
 Линейность 23, 185, 407  
 Линейные неравенства 353  
 Ляпунов 246
- Максимум 283  
 Марковский процесс 232, 235, 244, 245  
 Массы 249, 301, 315  
 Масштабирование 51  
 Матрица 19  
 — Вандермонда 99  
 — весов 178

- вырожденная 40, 187
- Гессе 285
- Гильберта 51, 56, 161
- дефектная 216, 223, 275
- единичная 29
- Жордана 275, 417
- «затраты — выпуск» Леонтьева 238
- идемпотентная 140
- инцидентности 112, 389
- квадратный корень 227, 291
- ковариации 177, 179
- конечных разностей 59, 191, 198, 319
- корреляционная 251
- кососимметрическая 190, 401
- косоэрмитова 263, 264
- коэффициентов 19
- ленточная 52, 58
- Мура — Пенроуза 165
- невырожденная 41, 98
- неопределенная 295
- неотрицательная 232, 237, 238
- нижняя треугольная 31
- нормальная 32, 44, 94
- обратимая 44, 188
- обратная 32, 44, 94
- ортогональная 151, 258, 329
- ортонормированная 151
- отражения 141
- отрицательно определенная 236
- перестановки 40, 43, 70, 152
- плохо обусловленная 48, 324
- положительно определенная 178, 285, 287, 331
- полуопределенная 295
- потребление 237
- присоединенная 201
- проектирования 139, 415
- псевдообратная 110, 165
- ранга один 93
- симметрическая 54, 132
- с кратными собственными значениями 260, 268, 272, 417
- — различными собственными значениями 222, 224, 260, 272
- ступенчатая форма 69, 76
- транспонированная 55, 131
- трапециевидная 65
- треугольная 31, 187, 270
- трехдиагональная 54, 336, 341
- унитарная 258, 264, 265
- Хессенберга 118, 336, 341
- хорошо обусловленная 48, 324
- элементарная 31
- эрмитова 256, 272
- эрмитовая к  $A$  255
- эрмитово сопряженная 255
- Метан 134
- Метод взвешенных наименьших квадратов 176, 178
- Гаусса — Зейделя 345, 351
- Гивенса 301
- исключения Гаусса 11, 13, 31, 43, 74, 206
- — — с полным выбором ведущего элемента 50
- — — частичным выбором ведущего элемента 50, 51, 324
- — Гаусса — Жордана 45, 75
- конечных разностей 53
- элементов 309, 315, 316
- наименьших квадратов 126, 135, 137, 166, 174, 306
- переменных направлений 350
- последовательной верхней релаксации 346
- сопряженных градиентов 351
- Якоби 336, 344, 351
- Минимум 279, 282
- Миnor 196
- Мнимое число 226, 251, 263
- Многомерный анализ 280
- Многочлены 98, 407
- Лежандра 162, 164
- Множители 33
- Лагранжа 378
- Множитель, характеризующий сходимость 333, 334
- Модуль 253
- Моленар 48
- Мур 165, 182
- Начальная задача 211
- Невырожденность 68, 98, 122
- Недоопределенная система 68
- Нейман 48, 236, 269, 402, 403
- Нейтральная устойчивость 211, 235, 245
- Нелинейная задача о наименьших квадратах 145
- Нелинейное убывание 236
- Ненулевое решение 72
- Необратимость во времени 244
- Неопределенная квадратичная форма 283
- матрица 295
- Неотрицательная матрица 232, 237, 238
- Неотрицательное решение 384
- Неравенство Коши—Шварца—Бунаковского 130
- треугольника 133

- Шварца 130, 133, 159  
 Несовместная система 68, 126, 138  
 Нетривиальная комбинация 77  
 Неустойчивость 211, 235, 245  
 Нижняя треугольная матрица 31  
 Нобл 48  
 Норма 324, 328, 330  
 Нормальная матрица 274, 277  
 Нормальные уравнения 137, 330  
 Нулевая длина 178  
 — строка 186  
 Нулевой ведущий элемент 14, 39, 98, 206  
 — вектор 64, 65  
 — определитель 182  
 Нуль-пространство 66, 72, 88, 105, 419
- Обобщенная задача на собственные значения 301, 304  
 — обратная матрица 165  
 Обобщенный собственный вектор 241, 417  
 Образ 89  
 Обратимая матрица 44, 188  
 Обратимое отображение 97  
 Обратная матрица 32, 44, 94  
 — — к произведению 44  
 — — — транспонированной 133  
 — — Мура — Пенроуза 165  
 — — формула 20  
 — подстановка 14, 35  
 Обратный степенной метод 333, 334  
 Обусловленность 48  
 Общее решение 74  
 Объединение 115  
 Ограничения 360, 363, 379  
 Однородная система 72  
 Определитель 182  
 — Вандермонда 191  
 — матрицы перестановки 189, 191, 194, 206  
 — свойства 185  
 — формула 191, 194, 197, 218  
 — Якоби 183, 207  
 Оптимальный вектор 355, 377, 379  
 Ортогонализация Грамма — Шмидта 127, 154, 162, 203, 309, 338  
 Ортогональная матрица 151, 258, 329  
 Ортогональное дополнение 107  
 Ортогональные векторы 101, 102  
 — подпространства 104  
 — собственные векторы 257, 263, 265, 274  
 Ортогональный базис 146  
 Ортонормированная матрица 151
- Ортонормированные векторы 103, 147, 155  
 — собственные векторы 258, 274  
 Осн эллипсоида 293, 311  
 Основная теорема линейной алгебры 92, 108  
 Основные подпространства 86, 119, 256  
 Отношение Релея 307, 309, 312, 320, 330, 334  
 Отображение «на» 97, 409  
 Отрицательно определенная квадратичная форма 283  
 — — матрица 236  
 — полуопределенная квадратичная форма 283  
 Ошибка 134, 324, 325  
 — округления 12, 47, 162, 227, 229
- Параболоид 305  
 Пенроуз 165, 174  
 Переменная невязки 357, 363  
 Переопределенная система 126, 143  
 Пересечение 114  
 Перестановка 193, 209  
 — строк 26, 36, 56, 186  
 Пифагор 100, 153  
 Плавающая точка 32  
 Планирование производства 359  
 Плохо обусловленная матрица 48, 324  
 Площадь 207  
 Погрешность (ошибка) 324, 325  
 Подгонка данных 143, 163  
 Подматрица 121  
 Подобные матрицы 268  
 Подпространство 64  
 Подсчет числа действий 15, 36, 46, 58, 341  
 Покер 405  
 Полный выбор ведущего элемента 50  
 — квадрат 282, 289  
 Положительно определенная квадратичная форма 281, 285  
 — — матрица 178, 285, 287, 331  
 — полуопределенная квадратичная форма 283  
 Полуопределенная матрица 295  
 Полупространство 353  
 Последовательность Фибоначчи 128, 135  
 Почти периодическое движение 249  
 Правая обратная матрица 44, 94  
 Правило Крамера 185, 201  
 Предельная стоимость 382  
 Представление в виде матрицы 410  
 Преобразование конгруэнтности 293

- подобия 268, 277, 414, 420  
 — Хаусхолдера 153, 336, 339  
 Принцип максимина 311  
 — минимума 304  
 — неопределенности Гейзенберга 269  
 — Релея 307  
 — Релея — Ритца 315  
 Присоединенная матрица 201  
 Присоединенный вектор 418  
 Проекция 109, 126, 136, 137, 148, 273  
 Произведение ведущих элементов 57, 184, 191  
 — матриц 25, 119, 413  
 — определителей 188  
 — псевдообратных матриц 176  
 Пропускная способность разреза 390  
 Пространство 64  
 — столбцов 64, 89  
 — строк 81, 87, 105  
 Псевдообратная матрица 110, 165  
 — — произведения 176  
 — — формула 170, 173  
  
 Разбиение на блоки 205  
 Разделяющая гиперплоскость 385, 386  
 Разложение определителя на алгебраические дополнения 196  
 — Холецкого 291, 331  
 —  $LDU$  37, 38, 55  
 —  $LU$  12, 34, 70  
 —  $\bar{L}\bar{U}$  120, 172  
 —  $QR$  157, 338  
 —  $Q_1\Sigma Q_2^T$  170, 339  
 Размерность 84  
 — основных подпространств 92  
 Разностное уравнение 53, 227  
 Разрез сети 390  
 Разрешимость системы 63  
 Райнш 323  
 Ранг 75, 77  
 — подматрицы 121  
 — произведения 120  
 Расстояние 129  
 Раус 246  
 Ребро 361, 362  
 Регрессионный анализ 126, 260  
 Релаксационный множитель 346  
 Релей 307, 315  
 Ритц 315  
 Ряды Фурье 160, 164  
  
 Сверхубывание 243  
 Свободные переменные 71, 88, 364  
 — члены 18  
 Свойства определителя 185  
  
 Сдвиг 334, 340  
 Седловая точка 283, 309, 397  
 Сетевые задачи 388  
 — модели 388  
 Сильвестр 298  
 Симметрическая матрица 54, 132  
 Симметрическое исключение 297  
 Симплекс-метод 360, 361, 372, 379  
 Сингулярное разложение 170, 330, 339  
 Сингулярные числа 170, 330  
 Скалярное произведение 20, 25, 102  
 — — взвешенное 178  
 — — в комплексном случае 254, 265  
 — — функций 160  
 Слабая двойственность 377  
 След 218,  
 Смешанная стратегия 396, 399  
 Собственное подпространство 214  
 Собственные значения 182, 213, 263, 267, 268  
 — — кратные 260, 272, 417  
 — — различные 222, 224, 260, 272  
 — функции 243  
 Собственный вектор 213, 230, 257  
 — — обобщенный 241, 417  
 Сопрягающий член 149  
 Сопряженно транспонированная матрица 255  
 Спектральная теорема 258, 271  
 Спектральный радиус 344  
 Среднее арифметическое 133, 150  
 — геометрическое 133  
 — значение 146  
 Средняя ошибка 134, 142, 143  
 Стандартный базис 147  
 Стационарная точка 280  
 Стационарное состояние 233, 244  
 Степенные методы 323, 332, 334  
 Ступенчатая форма матрицы 63, 76  
 Стюарт 342  
 Сумма подпространств 115, 118  
 Существование решения 95  
  
 Таблица для симплекс-метода 367  
 Такер 379  
 Теневые цены 382  
 Теорема двойственности 376  
 — Кэли — Гамильтона 271  
 — о кругах Гершгорина 351  
 — — максимальной точке и минимальном разрезе 390  
 — — минимаксе 402  
 — — равновесии 379  
 — — разделяющей гиперплоскости 386  
 Теория игр 395

- Горп 404  
 Транспонированная матрица 55, 131  
 — — определитель 189  
 — — к обратной 133  
 — — произведению 131  
 Трапецидальная матрица 69  
 Треугольная матрица 31, 187, 270  
 Трехдиагональная матрица 54, 336, 341
- Угол 127, 265, 356, 361  
 Узлы сети 388  
 Уилкинсон 48, 323, 329, 342  
 Умножение вектора на матрицу 19, 20  
 — матриц 25  
 Унитарная матрица 258, 264, 265  
 Уравнение теплопроводности 243, 244  
 Условие оптимальности 368  
 Условия Куна — Такера 379  
 — совместности невязок 379, 382, 398  
 Устойчивость 212, 235, 245, 271
- Факторный анализ 260, 261  
 Фибоначчи 228, 235  
 Фикс 320  
 Филиппов 276, 418  
 Формула для ведущего элемента 204, 205  
 — — обратной матрицы 20  
 — — определителя 191, 194, 197, 218  
 — — псевдообратной матрицы 170, 173  
 Форсайт 48  
 Фредгольм 111  
 Функциональное пространство 158, 159  
 Функция стоимости 355, 360, 368  
 Фурье 160, 164
- Характеристический многочлен 213, 217, 269  
 Характеристическое уравнение 213, 217  
 Хаусхолдер 153, 336, 339  
 Хессенберг 118, 336, 341  
 Хокней 350  
 Холецкий 291, 300, 331  
 Хорошо обусловленная матрица 48, 324
- Целевая функция 355  
 Цена игры 397, 402  
 Цепочка векторов 418
- Частичный выбор ведущего элемента 50, 324  
 Частное решение 74  
 Частота 215, 248  
 Четная перестановка 206  
 Численное интегрирование 99  
 Число обусловленности 323, 325, 328
- Шахматы 404, 406  
 Шварц 130, 133, 159  
 Ширина ленты 58  
 Шур 270
- Экономика 236, 238, 263, 376  
 Экспонента от матрицы 239, 240, 244, 422  
 Экспоненциальное решение 215, 239, 240, 244  
 Элементарное преобразование 186  
 Элементарные матрицы 31  
 Эллипс 293  
 Эллипсоид 294, 304, 311  
 Эпидемия 234  
 Эрмитова матрица 256, 272  
 Эрмитово сопряженная матрица 255
- Ядро 89  
 Якоби 336, 344, 351  
 Якобиан 183, 207  
 Янг 348  
 «Ящичные» ограничения 389
- LDU*-разложение 37, 38, 55  
*LU*-разложение 12, 34, 70  
 $\bar{L}\bar{U}$ -разложение 120, 172  
*QR*-алгоритм (метод) 323, 332, 340, 341  
*QR*-разложение 157, 338  
 $Q_1\Sigma Q_2^T$ -разложение 170, 339

# Оглавление

От редактора перевода . . . . .	5
Предисловие . . . . .	7
<b>Глава 1. МЕТОД ИСКЛЮЧЕНИЯ ГАУССА . . . . .</b>	<b>11</b>
§ 1.1. Введение . . . . .	11
§ 1.2. Пример применения метода исключения Гаусса . . . . .	13
§ 1.3. Матричные обозначения и умножение матриц . . . . .	17
§ 1.4. Эквивалентность метода исключения Гаусса и разложения на треугольные матрицы . . . . .	30
§ 1.5. Перестановки строк, обращения и ошибки округления . . . . .	39
§ 1.6. Ленточные матрицы, симметрические матрицы и их применения Обзорные упражнения . . . . .	52 60
<b>Глава 2. ТЕОРИЯ СИСТЕМ ЛИНЕЙНЫХ УРАВНЕНИЙ . . . . .</b>	<b>62</b>
§ 2.1. Векторные пространства и подпространства . . . . .	62
§ 2.2. Решение $m$ уравнений с $n$ неизвестными . . . . .	68
§ 2.3. Линейная независимость, базис и размерность . . . . .	77
§ 2.4. Четыре основных подпространства . . . . .	86
§ 2.5. Ортогональность векторов и подпространств . . . . .	100
§ 2.6. Пары подпространств и произведения матриц . . . . .	113
Обзорные упражнения . . . . .	123
<b>Глава 3. ОРТОГОНАЛЬНЫЕ ПРОЕКЦИИ И МЕТОД НАИМЕНЬШИХ КВАДРАТОВ . . . . .</b>	<b>125</b>
§ 3.1. Скалярные произведения и транспонирование . . . . .	125
§ 3.2. Проекции на подпространства и аппроксимации по методу на- именьших квадратов . . . . .	134

§ 3.3. Ортогональные базисы, ортогональные матрицы и ортогонализация Грама—Шмидта . . . . .	146
§ 3.4. Псевдообращение и сингулярное разложение . . . . .	164
§ 3.5. Взвешенные наименьшие квадраты . . . . .	174
Обзорные упражнения . . . . .	180
<b>Глава 4. ОПРЕДЕЛИТЕЛИ . . . . .</b>	<b>182</b>
§ 4.1. Введение . . . . .	182
§ 4.2. Свойства определителя . . . . .	185
§ 4.3. Формулы для определителя . . . . .	191
§ 4.4. Применения определителей . . . . .	200
Обзорные упражнения . . . . .	208
<b>Глава 5. СОБСТВЕННЫЕ ЗНАЧЕНИЯ И СОБСТВЕННЫЕ ВЕКТОРЫ . . . . .</b>	<b>210</b>
§ 5.1. Введение . . . . .	210
§ 5.2. Диагональная форма матрицы . . . . .	221
§ 5.3. Разностные уравнения и степени $A^k$ . . . . .	227
§ 5.4. Дифференциальные уравнения и экспонента $e^{At}$ . . . . .	239
§ 5.5. Комплексный случай: эрмитовы и унитарные матрицы . . . . .	251
§ 5.6. Преобразования подобия и треугольные формы . . . . .	267
Обзорные упражнения . . . . .	277
<b>Глава 6. ПОЛОЖИТЕЛЬНО ОПРЕДЕЛЕННЫЕ МАТРИЦЫ . . . . .</b>	<b>279</b>
§ 6.1. Максимумы, минимумы и седловые точки . . . . .	279
§ 6.2. Критерии положительной определенности . . . . .	286
§ 6.3. Полуопределенные и неопределенные матрицы. Обобщенная задача на собственные значения $Ax = \lambda Bx$ . . . . .	295
§ 6.4. Принципы минимума и отношение Релея . . . . .	304
§ 6.5. Принцип Релея—Ритца и метод конечных элементов . . . . .	315
<b>Глава 7. ВЫЧИСЛЕНИЯ С МАТРИЦАМИ . . . . .</b>	<b>322</b>
§ 7.1. Введение . . . . .	322
§ 7.2. Норма и число обусловленности матрицы . . . . .	324
§ 7.3. Вычисление собственных значений . . . . .	332
§ 7.4. Итерационные методы решения системы $Ax = b$ . . . . .	343
<b>Глава 8. ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ И ТЕОРИЯ ИГР . . . . .</b>	<b>353</b>
§ 8.1. Линейные неравенства . . . . .	353
§ 8.2. Симплекс-метод . . . . .	360

---

§ 8.3. Теория двойственности . . . . .	374
§ 8.4. Сетевые модели . . . . .	388
§ 8.5. Теория игр и теорема о минимаксе . . . . .	395
<i>Приложение А. ЛИНЕЙНЫЕ ПРЕОБРАЗОВАНИЯ, МАТРИЦЫ И ЗАМЕНЫ БАЗИСОВ . . . . .</i>	<b>407</b>
<i>Приложение В. ЖОРДАНОВА ФОРМА МАТРИЦЫ . . . . .</i>	<b>416</b>
Список литературы . . . . .	423
Решения . . . . .	424
Указатель . . . . .	446

#### УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., д. 2, издательство «Мир».

Г. Стренг

**ЛИНЕЙНАЯ АЛГЕБРА  
И ЕЕ ПРИМЕНЕНИЯ**

Научный редактор Г. М. Цукерман  
Мл. научный редактор Л. С. Суркова  
Художник М. К. Шевцов  
Художественный редактор В. И. Шаповалов  
Технический редактор Н. Б. Панфилова  
Корректор Н. А. Гиря

ИБ № 1374

Сдано в набор 12.10.79. Подписано к печати 04.02.80.  
Формат 60×90<sup>1/16</sup>. Бумага типографская № 2.  
Гарнитура литературная. Печать высокая.  
Объем: 14,25 бум. л. Усл. печ. л. 28,50.  
Уч.-изд. л. 25,57. Изд. № 1/9886.  
Тираж 14 000 экз. Заказ № 740. Цена 2 р. 10 к.

**ИЗДАТЕЛЬСТВО «МИР»**  
Москва, 1-й Рижский пер., 2

Ордена Октябрьской Революции и ордена Трудового Красного Знамени  
Первая Образцовая типография имени А. А. Жданова  
Союзполиграфпрома при Государственном комитете СССР  
по делам издательств, полиграфии и книжной торговли.  
Москва, М-54, Валовая, 28

