П.Ланкастер

ТЕОРИЯ МАТРИЦ

Главная редакция физико-математической литературы издательства «Наука», М., 1973, 280 с.

Книга предназначена быть основой для спецкурсов и справочным пособием для всех, интересующихся прикладными аспектами теории матриц. Ее можно рассматривать как хорошее дополнение к обычному курсу линейной алгебры (первые две главы — изложение линейной алгебры на матричном языке).

Строгое изложение основ теории матриц сочетается в ней с обсуждением прикладных вопросов, отчасти классических, отчасти новых.

ОГЛАВЛЕНИЕ

Глава 3
Вариационный метод

Глава 4
Минимальный многочлен и нормальные формы

Глава 5
Функции от матриц

АЛФАВИТНЫЙ УКАЗАТЕЛЬ

ПРЕДИСЛОВИЕ

Эта книга написана в основном для студентов по специальностям: прикладная математика, инженерное дело и различные естественные науки (биология, химия и др.), которые желают получить полное знакомство с теорией матриц, включая некоторые наиболее важные достижения теории за приблизительно последние двадцать лет. Однако ввиду не совсем обычного содержания последних глав я думаю, что многое здесь заинтересует также и математика.

Имея в виду предполагаемую аудиторию, я не счел себя обязанным придерживаться всех математических абстракций, в то же время стремился всюду к строгому изложению, которое было бы приемлемо как для математиков, так и для их собратьев-прикладников. Мы сочли также совершенно необходимым привести нетривиальные приложения. Некоторые из них разбираются в деталях, другие же появляются в упражнениях и примерах. Некоторые из этих приложений принадлежат области численных методов, и было бы возможно включить значительно большее количество таких приложений. Однако, чтобы книга не стала слишком громоздкой и чтобы не отклониться от основной линии развития теории, я должен был тщательно себя ограничить во избежание искушения перегрузить книгу численными методами.

Хотя предварительные знания для читателя не предполагаются обширными, они включают в себя полное владение алгеброй в объеме колледжа (в частности, сюда входят комплексные числа и полиномиальные уравнения), годовой курс дифференциального и интегрального исчисления и некоторое знакомство с алгеброй связанных векторов. Тем не менее опыт показывает, что, исходя из требуемого уровня восприятия студентов, материал следует излагать на третьем курсе. Его достаточно для полного годового курса. Главы 1, 2 и 4 охватывают материал, излагаемый во многих местах и предназначенный для полугодовых курсов, и мы рекомендуем главы 1, 2, 4 и 5 как ядро любого курса, основанного на данной книге. Затем могут быть усвоены главы 3, 6 и 7—9, в зависимости от интересов преподавателя и аудитории. Вторая часть главы 7 и § 5.8 используют теорию функций комплексной переменной, но эти разделы независимы от остальной части книги и могут быть при необходимости пропущены.

Сокращение упр. используется одновременно для обозначения как примеров, так и упражнений. Так как предполагается, что студенты уделят внимание как тем, так и другим, то лишь малое количество из них может быть пропущено. Следует подчеркнуть, что упражнения и примеры составляют существенную часть книги, а так как я придерживаюсь взгляда, что математика может быть усвоена лишь при помощи активного действия, то всякий серьезный читатель этой книги должен быть готов выполнить по крайней мере 50% упражнений. Большинство упражнений довольно простые, там же, где могут возникнуть некоторые трудности, даются краткие советы или указания на их полное решение. Отметим также, что некоторые из упражнений используются в изложении теории. Используемые таким образом упражнения отмечены звездочкой.

Каждая глава делится на параграфы. Так, § 5.3 — это третий параграф главы 5. Теоремы нумеруются при помощи номера параграфа, так что теорема 5.3.2 — это вторая теорема из § 5.3. Упражнения нумеруются последовательно 1, 2, 3, ... в каждом параграфе.

Книга такого рода во многом, естественно, должна быть обязана более ранним публикациям на эту тему. В данном изложении теории я воспользовался несколькими предыдущими работами, наиболее важные из которых упомянуты в дополнении 3. Однако я могу выделить «Теорию матриц» Гантмахера как особенно важный для справок труд и как ценный источник идей. Я благодарен нескольким машинисткам за помощь на различных стадиях подготовки рукописи, а особенно Лезии Хорелак за всегда квалифицированное содействие. Мои коллеги Ричард К. Гай и Девид Л. Рой помогли детальными и конструктивными замечаниями по главе 1 и Джон Г. Рокне принял участие в окончательном корректировании. Я искренне благодарен всем троим. Но так как я один несу ответственность за ошибки в тексте, прошу обратить на них мое внимание.

Питер Ланкастер

Глава 1

ЛИНЕЙНЫЕ ПРОСТРАНСТВА, АЛГЕБРА МАТРИЦ И ЛИНЕЙНЫЕ АЛГЕБРАИЧЕСКИЕ УРАВНЕНИЯ

1.1. Связанные векторы в трехмерном пространстве

Наша цель в этой главе — ввести понятие линейного пространства и, в частности, тех пространств, которые понадобятся нам при изучении матриц. Начнем с основного понятия: *упорядоченная n-ка* чисел есть n чисел, расположенных в определенном порядке. Повторения одного и того же числа в упорядоченной n-ке допустимы, но если положения двух неравных чисел в упорядоченной n-ке меняются, то в результате получается новая упорядоченная n-ка, не равная первоначальной. Нашей ближайшей целью является введение алгебры упорядоченных n-ок, т. е. мы хотим исследовать способы их комбинирования и преобразования друг в друга.

Многие читатели, по-видимому, знакомы до некоторой степени с алгеброй упорядоченных *троек* вещественных чисел, но, возможно, в другом виде. Мы имеем в виду изучение алгебры связанных векторов в трехмерном евклидовом пространстве. Чтобы начать наш анализ с того, что будет знакомо для большинства читателей, мы кратко (и с некоторой опорой на интуицию) обсудим алгебру связанных векторов и затем покажем, как изложенные идеи могут быть обобщены на изучение упорядоченных n-ок и на понятие n-мерного пространства.

С точкой P, имеющей декартовы координаты (x_1, x_2, x_3), ассоциируется связанный вектор $\boldsymbol{x}$. Начало O системы координат, т. е. $(0, 0, 0)$, определяет *нулевой вектор*. Точка P однозначно определяет либо $\boldsymbol{x}$, либо упорядоченную тройку вещественных чисел (x_1, x_2, x_3). Рассматривая $\boldsymbol{x}$ и (x_1, x_2, x_3) как различные обозначения одного и того же вектора, будем писать

$$\boldsymbol{x} = (x_1,\ x_2,\ x_3).$$

Длина вектора $\boldsymbol{x}$ задается формулой $|\boldsymbol{x}|=\sqrt{x_1^2+x_2^2+x_3^2}$, а его направление определено углами θ_1, θ_2, θ_3, где $0\leqslant\theta_i\leqslant\pi$ и

$$\cos\theta_i=\frac{x_i}{|x|},\qquad i=1,\ 2,\ 3.$$

Это *направляющие косинусы* вектора $\boldsymbol{x}$ (рис. 1).

Пусть точкам P, Q соответствуют связанные векторы $\boldsymbol{x}$, $\boldsymbol{y}$. Мы определяем *сумму* векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ как вектор

$$\boldsymbol{x}+\boldsymbol{y}=(x_1+y_1,\ x_2+y_2,\ x_3+y_3).$$

Легко проверить, что это определение эквивалентно правилу параллелограмма для сложения векторов, которое утверждает,

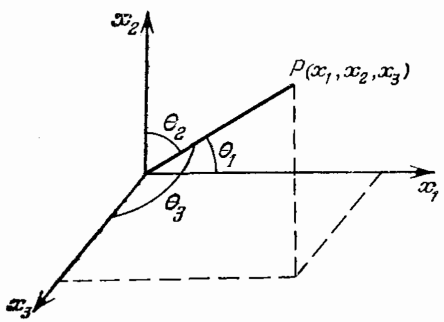

Рис. 1.

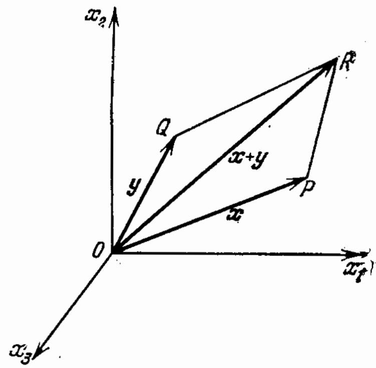

Рис. 2.

что если $\boldsymbol{x}+\boldsymbol{y}$ — связанный вектор точки R, то $OPRQ$ — параллелограмм (рис. 2).

Из определения и соответствующих свойств вещественных чисел следует, что сложение связанных векторов коммутативно и ассоциативно, т. е.

$$\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}$$

и

$$(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}=\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z}).$$

Последнее равенство означает, что можно писать $\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}$, не опасаясь двусмысленности.

Рассмотрим связанный вектор $\boldsymbol{z}$, который имеет длину $|\boldsymbol{x}|$, но направление, противоположное $\boldsymbol{x}$. Если направления векторов $\boldsymbol{x}$ и $\boldsymbol{z}$ задаются углами θ_i, φ_i соответственно $(i=0,2,3)$, то $\varphi_i=\pi-\theta_i$ и

$$\cos\varphi_i=-\cos\theta_i=-\frac{x_i}{|\boldsymbol{x}|}=\frac{-x_i}{|\boldsymbol{z}|}.$$

Таким образом, $z_i = -x_i$, $z = (-x_1, -x_2, -x_3)$ и $x + z = 0$. Поэтому мы записываем z как $-x$. Следовательно, для каждого связанного вектора x существует вектор $-x$ такой, что $x + (-x) = 0$.

Пусть k — произвольное вещественное число. Определим вектор $k(x_1, x_2, x_3)$, или kx, как (kx_1, kx_2, kx_3). Если x фиксирован и k пробегает все вещественные числа, то множество всех векторов вида kx представляет *все* точки на прямой, проходящей через начало координат и содержащей x. Ясно, что можно написать $(-x) = (-1)x$. Легко убедиться, что для любого вещественного числа α

$$\alpha(x + y) = \alpha(x_1 + y_1, x_2 + y_2, x_3 + y_3) = (\alpha(x_1 + y_1), \alpha(x_2 + y_2), \alpha(x_3 + y_3)) = \alpha x + \alpha y,$$

а также

$$(\alpha + \beta)x = \alpha x + \beta x, \quad \alpha(\beta x) = (\alpha\beta)x.$$

Единичные векторы $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$ особенно важны. Используя введенные выше операции, мы можем написать теперь для любого x:

$$x = (x_1, x_2, x_3) = (x_1, 0, 0) + (0, x_2, 0) + (0, 0, x_3) = x_1e_1 + x_2e_2 + x_3e_3.$$

1.2. Пространства $\mathcal{R}_n$ и $\mathcal{C}_n$

Перечислим теперь те свойства связанных векторов, которые обсуждались выше. Для *любых* связанных векторов x, y, z и любых вещественных чисел α, β имеем:

А1 $x + y$ — однозначно определенный вектор из того множества, которому принадлежат x, y.

А2 $x + y = y + x$.

А3 $(x + y) + z = x + (y + z)$.

А4 Существует такой вектор 0, что $x + 0 = x$.

А5 Для любого x существует такой вектор $-x$, что

$$x + (-x) = 0.$$

Б1 αx — вектор из того множества, которому принадлежит x.

Б2 $\alpha(\beta x) = (\alpha\beta)x$.

Б3 $\alpha(x + y) = \alpha x + \alpha y$.

Б4 $(\alpha + \beta)x = \alpha x + \beta x$.

Б5 $1 \cdot x = x$.

Мы используем теперь эти свойства для общего определения. Любое множество элементов (интуитивно соответствующих связанным векторам, вследствие чего эти элементы называются векторами), которые удовлетворяют условиям А1—А5 и Б1—Б5, называется *линейным пространством над полем вещественных чисел*.

Элементы множеств из некоторых последующих упражнений являются *функциями*. Если это так, то сумма двух функций есть функция, получаемая обычным образом при помощи сложения соответствующих функциональных значений для каждого допустимого значения независимой переменной. Подобно этому произведение вещественного числа и функции есть функция, получаемая умножением функциональных значений на заданное число.

Например, пусть функции f и g определены на отрезке [2, 3] формулами

$$f(x) = (1-x)^{-1} \quad \text{и} \quad g(x) = e^x.$$

Функции $f+g$ и αf (для любого вещественного α) определены на том же самом отрезке формулами

$$(f+g)(x) = f(x) + g(x) = (1-x)^{-1} + e^x,$$
$$(\alpha f)(x) = \alpha f(x) = \alpha (1-x)^{-1}.$$

Упр. 1. Проверить, что следующие множества являются линейными пространствами над полем вещественных чисел:

(а) множество всех многочленов степени не выше n с вещественными коэффициентами;

(б) множество всех вещественнозначных функций, определенных на [0, 1] и имеющих там непрерывную первую производную;

(в) множество всех ограниченных функций, определенных на [0, 1];

(г) множество всех сходящихся последовательностей (с соответствующими операциями сложения и умножения).

Упр. 2. Являются ли следующие множества линейными пространствами над полем вещественных чисел:

(а) множество всех функций, интегрируемых на [0, 1];

(б) множество всех вращений твердого тела вокруг осей, проходящих через фиксированную точку;

(в) множество всех связанных векторов единичной длины в трехмерном пространстве?

Пусть A, B и S — произвольные множества и $A \times B$ обозначает множество всех упорядоченных пар элементов (a, b), где a, b — элементы из A и B соответственно. *Бинарная операция* из $A \times B$ в S есть правило, по которому каждой упорядоченной паре из $A \times B$ ставится в соответствие единственный элемент из S. Например, если каждое из A, B и S есть множество всех вещественных чисел, то сложение, вычитание и умножение вещественных чисел являются бинарными операциями из $A \times A$ в A. Если за A_1 принять множество всех вещественных чисел, за исключением нуля, то деление будет бинарной операцией из $A \times A_1$ в A. Если V — множество всех связанных векторов, то бинарная операция *умножения на скаляр*, удовлетворяющего условиям Б1—Б5, отображает $A \times V$ в V. Бинарная операция *сложения векторов*, удовлетворяющего условиям А1—А5, отображает $V \times V$ в V.

Бинарная операция, определенная на множестве $R \times R$, называется *замкнутой*, если она отображает $R \times R$ в R. Так, сложение векторов — замкнутая бинарная операция.

В определении линейного пространства было использовано выражение «поле вещественных чисел». Мы не даем определения понятия «поле». Заинтересованный читатель, не знакомый пока с этим понятием, может найти его определение в любом изложении современной алгебры. В применении к этой книге поле, о котором идет речь, всегда будет либо множеством $\mathcal{R}$ всех вещественных чисел, либо множеством $\mathcal{C}$ всех комплексных чисел. Следовательно, тому, кто не знаком с общим определением, всегда можно при упоминании общего поля $\mathcal{F}$ представлять себе вместо $\mathcal{F}$ либо $\mathcal{R}$, либо $\mathcal{C}$.

Рассмотрим теперь множество $\mathcal{L}$ такое, что условия А1—А5 и Б1—Б5 выполняются для любых $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ из $\mathcal{L}$ и любых α, β из поля $\mathcal{F}$. В таком случае мы говорим, что $\mathcal{L}$ — линейное пространство над полем $\mathcal{F}$.

В частности, рассмотрим множество всех упорядоченных n-ок вещественных чисел. Если $\boldsymbol{x}$, $\boldsymbol{y}$ из этого множества, то мы пишем

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n), \quad \boldsymbol{y} = (y_1, y_2, \ldots, y_n)$$

и определяем сложение в этом множестве по правилу

$$\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n).$$

Умножение на вещественное число α определяется равенством

$$\alpha\boldsymbol{x} = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n).$$

Легко проверить, что введенные операции превращают рассматриваемое множество в линейное пространство над вещественными числами. Это пространство обозначается через $\mathcal{R}_n$. При $n = 3$ получаем уже рассмотренные связанные векторы и $\mathcal{R}_3$ соответствует обычному евклидову трехмерному пространству. При $n = 2$ получаем множество связанных векторов на плоскости.

Второй важный пример линейного пространства: множество всех упорядоченных n-ок комплексных чисел над полем комплексных чисел с операциями, определенными по тем же правилам, что и для $\mathscr{R}_n$. В общем случае через $\mathscr{F}_n$ обозначаем линейное пространство упорядоченных n-ок над полем $\mathscr{F}$.

В дальнейшем элемент $(x_1, x_2, \ldots, x_n)$ линейного пространства $\mathscr{F}_n$ будет упоминаться также как *вектор порядка n*.

1.3. Внутренние произведения

В пространстве $\mathscr{R}_3$ «скалярное произведение» $\boldsymbol{x} \cdot \boldsymbol{y} \stackrel{\prime}{=} x_1y_1 + x_2y_2 + x_3y_3$ — важное и полезное понятие. В частности, имеем

$$\boldsymbol{x} \cdot \boldsymbol{x} = |\boldsymbol{x}|^2 \quad \text{и} \quad \boldsymbol{x} \cdot \boldsymbol{y} = |\boldsymbol{x}| \cdot |\boldsymbol{y}| \cos \theta,$$

где θ — угол между $\boldsymbol{x}$ и $\boldsymbol{y}$, причем ненулевые векторы $\boldsymbol{x}$ и $\boldsymbol{y}$ по определению перпендикулярны, если $\theta = 90°$, т. е. если $\boldsymbol{x} \cdot \boldsymbol{y} = 0$.

В пространстве $\mathscr{R}_n$ мы вводим другое обозначение для того же самого произведения. Мы пишем

$$\langle x,\ y\rangle = x_1y_1 + x_2y_2 + \ldots + x_ny_n \tag{1.3.1}$$

и называем это число *внутренним произведением* векторов $\boldsymbol{x}$ и $\boldsymbol{y}$. Мы заимствуем также геометрический язык из $\mathscr{R}_3$ для определения

$$|\boldsymbol{x}| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} = \langle \boldsymbol{x},\ \boldsymbol{x}\rangle^{1/2}$$

как *длины* вектора $\boldsymbol{x}$ и говорим, что два вектора $\boldsymbol{x}$, $\boldsymbol{y}$ из $\mathscr{R}_n$ *ортогональны*, если $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$.

Обобщим это понятие на более широкий класс линейных пространств. В общем линейном пространстве $\mathscr{L}$ над полем $\mathscr{F}$ (в котором определено комплексное сопряжение) бинарная операция $\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ из $\mathscr{L} \times \mathscr{L}$ в $\mathscr{F}$ называется внутренним произведением, если для всех $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ из $\mathscr{L}$ и α, β из $\mathscr{F}$ выполняются условия:

B1 $\langle \boldsymbol{x}, \boldsymbol{x}\rangle \geqslant 0$ и $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = 0$ тогда и только тогда, когда $\boldsymbol{x} = 0$.

B2 $\langle \boldsymbol{x}, \alpha\boldsymbol{y} + \beta\boldsymbol{z}\rangle = \alpha\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \beta\langle \boldsymbol{x}, \boldsymbol{z}\rangle$.

B3 $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x}\rangle}$ (где черта сверху обозначает комплексное сопряжение).

Легко проверяется, что эти условия выполняются для бинарной операции (1.3.1) в $\mathscr{R}_n$.

Попытаемся теперь найти внутреннее произведение в $\mathscr{C}_n$. Будет ли таковым произведение, определенное формулой (1.3.1)? Согласно этой формуле мы имели бы

$$\langle \boldsymbol{x},\ \boldsymbol{x}\rangle = x_1^2 + x_2^2 + \ldots + x_n^2,$$

и ясно, что в общем случае условие B1 не выполняется. Если же положить

$$\langle \boldsymbol{x},\ \boldsymbol{y}\rangle = \bar{x}_1y_1 + \bar{x}_2y_2 + \ldots + \bar{x}_ny_n, \tag{1.3.2}$$

то $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = \bar{x}_1 x_1 + \ldots + \bar{x}_n x_n = |x_1|^2 + \ldots + |x_n|^2 \geqslant 0$, и условие B1 теперь выполняется. Проверка условия B2:

$$\langle \boldsymbol{x}, \alpha \boldsymbol{y} + \beta \boldsymbol{z}\rangle = \bar{x}_1(\alpha y_1 + \beta z_1) + \bar{x}_2(\alpha y_2 + \beta z_2) + \ldots + \bar{x}_n(\alpha y_n + \beta z_n) = \\ = \alpha(\bar{x}_1 y_1 + \ldots + \bar{x}_n y_n) + \beta(\bar{x}_1 z_1 + \ldots + \bar{x}_n z_n) = \alpha\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \beta\langle \boldsymbol{x}, \boldsymbol{z}\rangle.$$

Для проверки B3 пишем

$$\overline{\langle \boldsymbol{y}, \boldsymbol{x}\rangle} = \overline{(\bar{y}_1 x_1 + \ldots + \bar{y}_n x_n)} = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n = \langle \boldsymbol{x}, \boldsymbol{y}\rangle.$$

Следовательно, формула (1.3.2) действительно определяет внутреннее произведение в $\mathscr{C}_n$ и, более того, (1.3.1) — частный случай (1.3.2) для $\mathscr{R}_n$. Естественно, если $\boldsymbol{x}$ — элемент из $\mathscr{C}_n$, то мы называем $\langle \boldsymbol{x}, \boldsymbol{x}\rangle^{1/2}$ длиной $\boldsymbol{x}$.

В конце этого параграфа мы приведем одно важное неравенство, обычно известное как *неравенство Шварца*:

Теорема 1.3.1. *Если $\boldsymbol{x}$, $\boldsymbol{y}$ — элементы линейного пространства, в котором определено внутреннее произведение, то*

$$|\langle \boldsymbol{x}, \boldsymbol{y}\rangle|^2 \leqslant \langle \boldsymbol{x}, \boldsymbol{x}\rangle\langle \boldsymbol{y}, \boldsymbol{y}\rangle.$$

Доказательство. Положим $a = -\langle \boldsymbol{x}, \boldsymbol{y}\rangle$, $b = \langle \boldsymbol{x}, \boldsymbol{x}\rangle$ и $\boldsymbol{z} = a\boldsymbol{x} + b\boldsymbol{y}$. Мы должны доказать, что $|a|^2 \leqslant b\langle \boldsymbol{y}, \boldsymbol{y}\rangle$. По условиям B1 и B2

$$0 \leqslant \langle \boldsymbol{z}, \boldsymbol{z}\rangle = \langle a\boldsymbol{x} + b\boldsymbol{y}, a\boldsymbol{x} + b\boldsymbol{y}\rangle = a\langle a\boldsymbol{x} + b\boldsymbol{y}, \boldsymbol{x}\rangle + b\langle a\boldsymbol{x} + b\boldsymbol{y}, \boldsymbol{y}\rangle$$

и, используя упр. 1 ниже,

$$0 \leqslant a\bar{a}\langle \boldsymbol{x}, \boldsymbol{x}\rangle + a\bar{b}\langle \boldsymbol{y}, \boldsymbol{x}\rangle + b\bar{a}\langle \boldsymbol{x}, \boldsymbol{y}\rangle + b\bar{b}\langle \boldsymbol{y}, \boldsymbol{y}\rangle.$$

Используя определения a и b и то, что $\bar{b} = b$, приводим последнее неравенство к виду

$$0 \leqslant b(b\langle \boldsymbol{y}, \boldsymbol{y}\rangle - |a|^2).$$

Если $b = 0$, то $\boldsymbol{x} = \boldsymbol{0}$ и результат тривиально верен. Если $b > 0$, то получаем $|a|^2 \leqslant b\langle \boldsymbol{y}, \boldsymbol{y}\rangle$. ◀

***Упр. 1.** При выполнении условий B1—B3 показать, что

$$\langle \alpha\boldsymbol{x} + \beta\boldsymbol{y}, \boldsymbol{z}\rangle = \bar{\alpha}\langle \boldsymbol{x}, \boldsymbol{z}\rangle + \bar{\beta}\langle \boldsymbol{y}, \boldsymbol{z}\rangle.$$

***Упр. 2.** В $\mathscr{R}_3$ *проекция* вектора $\boldsymbol{x}$ на ненулевой вектор $\boldsymbol{y}$ определяется как $\alpha\hat{\boldsymbol{y}}$, где $\hat{\boldsymbol{y}}$ — вектор единичной длины с направлением вектора $\boldsymbol{y}$ и α — длина вектора $\boldsymbol{x}$, умноженная на косинус угла между $\boldsymbol{x}$ и $\boldsymbol{y}$. Показать, что

$$\alpha = \langle \boldsymbol{y}, \boldsymbol{x}\rangle / \langle \boldsymbol{y}, \boldsymbol{y}\rangle^{1/2}.$$

***Упр. 3** (теорема Пифагора). Доказать, что если $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$, то

$$\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{x}\rangle + \langle \boldsymbol{y}, \boldsymbol{y}\rangle.$$

Упр. 4. Доказать, что

$$\langle x+y,\ x+y\rangle+\langle x-y,\ x-y\rangle=2\,(\langle x,\ x\rangle+\langle y,\ y\rangle).$$

Этот результат иногда упоминается как теорема Аполлония или как теорема о параллелограмме.

1.4. Линейные комбинации

Если $x_1, x_2, \ldots, x_r$ — элементы из $\mathscr{L}$ и $a_1, a_2, \ldots, a_r$ — элементы из $\mathscr{F}$, то $a_1x_1+\ldots+a_rx_r$ принадлежит $\mathscr{L}$ и называется *линейной комбинацией* векторов $x_1, x_2, \ldots, x_r$. В следующей теореме рассматривается множество всех линейных комбинаций от $x_1, x_2, \ldots, x_r$, т. е. множество всех векторов вида $a_1x_1+\ldots+a_rx_r$, где $(a_1, \ldots, a_r)$ пробегает всевозможные упорядоченные наборы r чисел из поля $\mathscr{F}$.

Теорема 1.4.1. *Множество всех линейных комбинаций векторов $x_1, \ldots, x_r$ из $\mathscr{L}$ является линейным пространством.*

Доказательство. Очевидно, что сумма линейных комбинаций есть линейная комбинация (A1). A2 и A3 следуют из свойств $\mathscr{L}$. Отмечая, что $\mathbf{0}$ может быть выражен как линейная комбинация векторов $x_1, \ldots, x_r$, получаем A4. Свойства A5 и Б1—Б5 проверяются так же просто. ◀

Описанное в теореме пространство называется *порожденным* (или *натянутым на*) r данными векторами. Оно является *подпространством* в $\mathscr{L}$, т. е. подмножеством $\mathscr{L}$, которое само есть линейное пространство над $\mathscr{F}$. Если обозначить подпространство через $\mathscr{L}'$, то каждый элемент из $\mathscr{L}'$ будет элементом из $\mathscr{L}$; обратное утверждение не обязательно выполняется. Мы говорим, что $\mathscr{L}'$ «содержится в» $\mathscr{L}$ и пишем $\mathscr{L}'\subseteq\mathscr{L}$, или $\mathscr{L}\supseteq\mathscr{L}'$, что выражается словами: «$\mathscr{L}$ содержит $\mathscr{L}'$». Если известно, что $\mathscr{L}'\neq\mathscr{L}$, то мы пишем $\mathscr{L}'\subset\mathscr{L}$ или $\mathscr{L}\supset\mathscr{L}'$ и говорим, что *$\mathscr{L}'$ — собственное подпространство в $\mathscr{L}$.*

Евклидова геометрия пространства $\mathscr{R}_3$ дает нам несколько полезных примеров. Отметим, что любые два ненулевых связанных вектора x и y, которые не кратны друг другу, порождают плоскость в $\mathscr{R}_3$, проходящую через начало координат. Один вектор x порождает прямую, проходящую через начало координат, и любые три вектора, не лежащие в одной плоскости, порождают $\mathscr{R}_3$.

Упр. 1. Найти вектор x в $\mathscr{R}_3$, который ортогонален к каждому из векторов $(2, 1, -1)$ и $(1, 5, 2)$, и показать, что любое скалярное кратное вектора x также ортогонально этим векторам.

Упр. 2. Найти длину проекции вектора $(2, 7, -1)$ на вектор $(5, 7, 4)$.

Упр. 3. Найти ненулевой вектор, который принадлежит как пространству, натянутому на $(1,2,-1)$ и $(3,2,0)$, так и пространству, натянутому на $(2,-1,-1)$ и $(1,0,4)$. Дать геометрическую интерпретацию.

Упр. 4. Доказать, что $(2,-1,6,4)$ принадлежит пространству, порожденному $(1,5,-2,1)$ и $(11,11,18,19)$.

***Упр. 5.** Показать, что векторы $(2,-1,6)$ и $(-3,4,1)$ порождают то же самое пространство, что и $(-1, 3, 7)$ и $(8, -9, 4)$.

1.5. Матричная алгебра

Мы определяем *матрицу* как прямоугольную таблицу чисел

$$\begin{Vmatrix} a_{11} & a_{12} & a_{1n} \\ a_{21} & a_{22} & a_{2n} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & a_{mn} \end{Vmatrix},$$

где *элементы* матрицы a_{ij} $(1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n)$ — числа из поля $\mathcal{F}$. Для наших целей поле $\mathcal{F}$ всегда будет либо множеством всех вещественных чисел, либо множеством всех комплексных чисел. Размер матрицы в терминах числа строк m и числа столбцов n обозначается через $m \times n$. Если $m = n$, то говорят, что матрица *квадратная порядка n*. В общем случае матрица называется *прямоугольной*.

Каждой $m \times n$-матрице A с элементами a_{ij} соответствует $n \times m$-матрица с элементами a_{ji}. Она называется *транспонированной* к A и обозначается через A'. Ясно, что $(A')' = A$. Отметим, что строки матрицы A становятся столбцами в A' и столбцы матрицы A становятся строками в A'.

Определим теперь следующие операции:

(i) Сумма двух $m \times n$-матриц A и B с элементами a_{ij} и b_{ij} есть $m \times n$-матрица C с элементами $c_{ij} = a_{ij} + b_{ij}$, и мы пишем $C = A + B$.

(ii) Произведение матрицы A на число α поля $\mathcal{F}$ есть матрица C с элементами $c_{ij} = \alpha a_{ij}$, и мы пишем $C = \alpha A$.

(iii) Произведение $m \times l$-матрицы A на $l \times n$-матрицу B есть $m \times n$-матрица C с элементами $c_{ij} = \sum_{k=1}^{l} a_{ik} b_{kj}$, и мы пишем $C = AB$.

***Упр. 1.** Рассмотрим множество всех $m \times n$-матриц (с фиксированными m, n) с элементами из $\mathcal{F}$. Показать, что операции (i) и (ii) превращают это множество в линейное пространство над $\mathcal{F}$.

Результат этого упражнения наводит на мысль обозначить множество всех вещественных (комплексных) $m \times n$-матриц через $\mathscr{R}_{m\times n}$ ($\mathscr{C}_{m\times n}$) и вообще множество всех $m \times n$-матриц с элементами из поля $\mathscr{F}$ через $\mathscr{F}_{m\times n}$. Таким образом, последнее упражнение просто означает, что $\mathscr{F}_{m\times n}$ — линейное пространство над $\mathscr{F}$. Очень важно, чтобы читатель завершил доказательство этого утверждения. В результате он убедится в том, что сложение матриц коммутативно и ассоциативно, что умножение на скаляры ассоциативно, как и требует условие Б2, и дистрибутивно по отношению к сложению матриц, как и требует условие Б3.

Отметим, что операция (iii), известная как *умножение матриц*, определена только для *соответствующих* друг другу матриц, т. е. число столбцов первого множителя должно быть равно числу строк второго. Следовательно, как определено в (iii), умножение матриц является бинарной операцией из $\mathscr{F}_{m\times l} \times \mathscr{F}_{l\times n}$ в $\mathscr{F}_{m\times n}$.

Читателю следует также доказать, что *умножение матриц дистрибутивно по отношению к сложению матриц*, т. е. если A, B, C, D — матрицы, то

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

всегда, когда все входящие сюда операции определены. Наконец, следует доказать, что умножение матриц *ассоциативно*, т. е. если P, Q — соответствующие друг другу матрицы и Q, R тоже, то

$$P(QR) = (PQ)R.$$

Эти свойства легко проверяются прямым вычислением, исходя из определений сложения и умножения матриц.

Отметим, что, вообще говоря, $BA \neq AB$; если имеет место равенство, то мы говорим, что A и B *коммутируют*. Если A и B коммутируют, то они, очевидно, квадратные матрицы одного и того же порядка.

Упр. 2. Вычислить AB, BC, $A(BC)$, $(AB)C$, $(BC)'$, $C'B'$, где

$$A = \| 1 \quad 2 \quad 1 \|, \quad B = \begin{Vmatrix} 4 & 1 & 0 \\ 0 & -2 & 3 \\ 0 & -1 & 2 \end{Vmatrix}, \quad C = \begin{Vmatrix} 1 & -1 \\ 2 & 0 \\ 4 & 2 \end{Vmatrix}.$$

***Упр. 3.** Доказать, что

(i) если A и B из $\mathscr{F}_{m\times n}$, то $(A+B)' = A' + B'$;

(ii) $(\alpha A)' = \alpha A'$, где A из $\mathscr{F}_{m\times n}$ и α из $\mathscr{F}$;

(iii) если произведение AB существует, то $(AB)' = B'A'$, а следовательно, если $AB \ldots K$ существует, то $(A \ldots K)' = K' \ldots A'$.

Упр. 4. Найти все матрицы, коммутирующие с

$$\text{(i)}\quad \begin{Vmatrix} & 1 & 0 \\ -1 & 0 \end{Vmatrix}, \qquad \text{(ii)}\quad \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{Vmatrix}.$$

***Упр. 5.** Если A из $\mathcal{F}_{m\times n}$ и $AI = A$, $JA = A$, то I из $\mathcal{F}_{n\times n}$ и J из $\mathcal{F}_{m\times m}$.

***Упр. 6.** Доказать, что если A из $\mathcal{F}_{m\times l}$, B из $\mathcal{F}_{l\times n}$ и α из $\mathcal{F}$, то $(\alpha A)B = A(\alpha B) = \alpha(AB)$,

В дополнение к упр. 1 читателю предлагается убедиться в том, что $m\times n$-матрица из $\mathcal{F}_{m\times n}$, элементы которой все являются нулями, играет роль нулевого элемента. Очевидно, что такая матрица действует также как нулевой элемент в операции умножения матриц. Существует ли такая матрица, которая выполняет роль единицы, или единичного элемента, для умножения матриц? Иначе говоря, существует ли такая матрица I, что $AI = A$ и $IB = B$ всегда, когда произведения определены? Нетрудно убедиться (упр. 5), что если $AI = A$ и A из $\mathcal{F}_{m\times n}$, то I должна быть квадратной матрицей порядка n. Отметим теперь, что $n\times n$-матрица I, заданная как

$$I = \begin{Vmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \ldots & 1 \end{Vmatrix},$$

удовлетворяет требуемому свойству для *любой* A из $\mathcal{F}_{m\times n}$. Читателю предлагается проверить этот простой факт, исходя из определения умножения матриц. Нетрудно проверить также, что для любой B из $\mathcal{F}_{n\times p}$ имеем $IB = B$. Эти два свойства позволяют назвать матрицу I *единичной матрицей.*

Следует отметить, что для некоторой *фиксированной* матрицы A может существовать много матриц B, для которых $AB = A$, но I — единственная матрица с этим свойством для любой матрицы A из $\mathcal{F}_{m\times n}$. Доказательство этого станет очевидным, когда будут привлечены понятия ранга и вырождения.

Скалярная матрица определяется как скалярное кратное единичной матрицы. Следовательно, A — скалярная матрица в $\mathcal{F}_{n\times n}$ тогда и только тогда, когда существует такое α из $\mathcal{F}$, что $A = \alpha I$, где I — единичная матрица порядка n. Следует заметить, что умножение на скалярную матрицу дает тот же самый эффект, что и операция умножения на скаляр и наоборот. Это следует из того, что для любой B из $\mathcal{F}_{m\times n}$ имеем

$$B(\alpha I) = \alpha(BI) = \alpha B,$$

и подобно этому для C из $\mathcal{F}_{n\times p}$ и α из $\mathcal{F}$ будет $(\alpha I)C = \alpha(IC) = \alpha C$.

Скалярные матрицы являются подмножеством другого класса матриц, известных как *диагональные матрицы*.

Будем называть элементы a_{jk} матрицы A с $j = k$ *главной диагональю* A. Квадратная матрица, ненулевые элементы которой лежат на главной диагонали, называется *диагональной*. Ясно тогда, что скалярная матрица является диагональной матрицей, все элементы которой на главной диагонали равны. Если D — диагональная матрица с элементами d_{jk}, $1 \leqslant j, k \leqslant n$, то

$$d_{jk} = d_{jj}\delta_{jk} = \begin{cases} d_{jj} & \text{при} \quad j = k, \\ 0 & \text{при} \quad j \neq k, \end{cases}$$

где δ_{jk} — *кронекеровская дельта*, имеющая значение 1 при $j = k$ и 0 при $j \neq k$.

Диагональная матрица сокращенно записывается в виде

$$D = \operatorname{diag}\{d_{11}, d_{22}, \ldots, d_{nn}\}.$$

Если A — квадратная матрица, то определено произведение AA, и мы пишем $AA = A^2$. Используя ассоциативность произведения матриц, можно показать, что $AA \ldots A$ (p раз) — однозначно определенная матрица, которую мы записываем в виде A^p. Предположим также, что $A^0 = I$. В настоящий момент эта формула должна рассматриваться как удобное соглашение, но в главе 5 мы увидим, что она является логическим следствием соответствующего соглашения для комплексных чисел.

Упр. 7. Найти все матрицы J, для которых $AJ = A$, если

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

***Упр. 8.** Для $D = \operatorname{diag}\{d_1, \ldots, d_n\}$ показать, что результат матричного произведения AD сводится к умножению j-го столбца матрицы A на d_j, $j = 1, 2, \ldots, n$; дать подобную интерпретацию для произведения DB.

***Упр. 9.** Доказать, что диагональные матрицы одного и того же порядка коммутируют.

1.6. Разбиения матриц

Если $m \times n$-матрицу P, $m \times l$-матрицу Q, $k \times n$-матрицу R и $k \times l$-матрицу S записать в виде

$$\begin{array}{c} \begin{Vmatrix} P & Q \\ R & S \end{Vmatrix} \begin{array}{l} m \\ k \end{array}, \\ \;\; n \quad\; l \end{array} \qquad (1.6.1)$$

то они, очевидно, образуют некоторую $(m+k)\times(n+l)$-матрицу A. В таком случае P, Q, R, S могут быть названы *блоками* матрицы A и обозначены через A_{11}, A_{12}, A_{21} и A_{22} соответственно. Представление (1.6.1) называется *разбиением* матрицы A. Эта идея может быть обобщена на несколько подразделений между строками и столбцами; блоки матрицы можно представлять себе образованными разделяющими линиями между некоторыми строками и некоторыми столбцами (каждая разделяющая линия полностью рассекает матрицу по ширине или высоте).

Если матричное произведение AB существует и A, B разбиты на блоки A_{qp}, B_{pr}, а разбиение по столбцам матрицы A соответствует разбиению по строкам матрицы B, то можно ожидать, что AB имеет блоки $(AB)_{qr}$, задаваемые формулой

$$(AB)_{qr}=\sum_p A_{qp}B_{pr}.$$

Таким образом, мы предполагаем, что произведение матриц в терминах блоков, полученных при соответствующих разбиениях сомножителей, формально совпадает с произведением этих матриц в терминах скалярных элементов. Прежде чем дать доказательство этого результата, иллюстрируем высказанное предположение простым примером.

Упр. 1. Пусть

$$A=\left\|\begin{array}{c:cc} 2 & 1 & 3\\ 0 & -1 & 2\end{array}\right\|,\quad A_{11}=\left\|\begin{matrix}2\\0\end{matrix}\right\|,\quad A_{12}=\left\|\begin{matrix}1 & 3\\ -1 & 2\end{matrix}\right\|,$$

$$B=\left\|\begin{array}{cccc} 3 & -7 & -7 & 2\\ \hdashline -2 & 1 & 4 & 0\\ 0 & 2 & 4 & 0\end{array}\right\|,\qquad B_{11}=\|3\ \ -7\ \ -7\ \ 2\|,$$

$$B_{21}=\left\|\begin{matrix}-2 & 1 & 4 & 0\\ 0 & 2 & 4 & 0\end{matrix}\right\|.$$

Тогда мы требуем, чтобы

$$AB=\|A_{11}\ \ A_{12}\|\left\|\begin{matrix}B_{11}\\ B_{21}\end{matrix}\right\|=A_{11}B_{11}+A_{12}B_{21}=$$

$$=\left\|\begin{matrix}2\\0\end{matrix}\right\|\|3\ \ -7\ \ -7\ \ 2\|+\left\|\begin{matrix}1 & 3\\ -1 & 2\end{matrix}\right\|\left\|\begin{matrix}-2 & 1 & 4 & 0\\ 0 & 2 & 4 & 0\end{matrix}\right\|=$$

$$=\left\|\begin{matrix}6 & -14 & -14 & 4\\ 0 & 0 & 0 & 0\end{matrix}\right\|+\left\|\begin{matrix}-2 & 7 & 16 & 0\\ 2 & 3 & 4 & 0\end{matrix}\right\|=\left\|\begin{matrix}4 & -7 & 2 & 4\\ 2 & 3 & 4 & 0\end{matrix}\right\|,$$

что может быть проверено прямым вычислением AB.

Теорема 1.6.1. *Пусть матрица A из $\mathcal{F}_{m\times l}$ имеет блоки A_{qp}, где A_{qp} — $m_q\times l_p$-матрица, $1\leqslant p\leqslant\alpha$, и B — матрица из $\mathcal{F}_{l\times n}$*

с блоками B_{pr} размера $l_p \times n_r$. Тогда AB имеет блоки

$$(AB)_{qr} = \sum_{p=1}^{\alpha} A_{qp}B_{pr}.$$

Доказательство. Отметим сначала, что каждое произведение $A_{qp}B_{pr}$ существует и является $m_q \times n_r$-матрицей. Следовательно, $C_{qr} = \sum_{p=1}^{\alpha} A_{qp}B_{pr}$ существует и будет $m_q \times n_r$-матрицей. Далее, для фиксированного q каждое C_{qr} имеет n_r столбцов и для фиксированного r каждое C_{qr} имеет m_q строк, откуда следует, что C_{qr} — блоки некоторой $m \times n$-матрицы C.

Пусть c_{ij} — некоторый элемент матрицы C, расположенной в клетке (a, b) блока C_{qr}. Так как $C_{qr} = \sum_{p=1}^{\alpha} A_{qp}B_{pr}$, c_{ij} есть сумма элементов в клетках (a, b) матриц $A_{qp}B_{pr}$, $p = 1, 2, \ldots$ $\ldots, \alpha$. Но элемент матрицы $A_{qp}B_{pr}$ в клетке (a, b) является суммой произведений l_p элементов в строке a матрицы A_{qr} на элементы столбца b матрицы B_{pr}. Далее, элементы строки a матрицы A_{qp} совпадают с некоторыми элементами i-й строки в A, а именно, с a_{ik}, где индекс k определяется неравенствами

$$k \leqslant l_1, \qquad \text{если} \quad p = 1,$$

$$\sum_{\nu=1}^{p-1} l_\nu < k \leqslant \sum_{\nu=1}^{p} l_\nu, \quad \text{если} \quad p > 1.$$

Подобно предыдущему, элементы столбца b матрицы B_{pr} будут элементами b_{kj} в B. Следовательно,

$$c_{ij} = \sum_{k=1}^{l_1} a_{ik}b_{kj} + \sum_{k=l_1+1}^{l_1+l_2} a_{ik}b_{kj} + \ldots + \sum_{k=n-l_\alpha+1}^{n} a_{ik}b_{kj} = \sum_{k=1}^{n} a_{ik}b_{kj}. \blacktriangleleft$$

***Упр. 2.** (i) *Квадратная* матрица A с блоками A_{qr} имеет *симметричное разбиение*, если блоки A_{qq} — квадратные матрицы для каждого q. Исследовать вид степеней A^n для целого положительного числа n, если $A_{qr} = 0$, как только $q \neq r$.

(ii) Пусть $p(\lambda) = \lambda^3 + \lambda^2 + 2\lambda + 2$ и $p(A) = A^3 + A^2 + 2A + 2$. Показать, что если

$$A = \begin{Vmatrix} \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & \mu & 1 & 0 & 0 & 0 \\ 0 & 0 & \mu & 0 & 0 & 0 \\ 0 & 0 & 0 & \nu & 1 & 0 \\ 0 & 0 & 0 & 0 & \nu & 1 \\ 0 & 0 & 0 & 0 & 0 & \nu \end{Vmatrix},$$

то

$$p(A)=\left\|\begin{array}{cccccc} p(\lambda) & 0 & 0 & 0 & 0 & 0 \\ 0 & p(\mu) & p'(\mu) & 0 & 0 & 0 \\ 0 & 0 & p(\mu) & 0 & 0 & 0 \\ 0 & 0 & 0 & p(\nu) & p'(\nu) & \frac{1}{2}p''(\nu) \\ 0 & 0 & 0 & 0 & p(\nu) & p'(\nu) \\ 0 & 0 & 0 & 0 & 0 & p(\nu) \end{array}\right\|.$$

Упр. 3. Пусть A — имеющая симметричное разбиение матрица вида

$$A=\left\|\begin{array}{cc} A_{11} & A_{12} \\ 0 & I \end{array}\right\|$$

и значение многочлена от A определено, как в упр. 2. Доказать, что для любого целого положительного числа n

$$A^n=\left\|\begin{array}{cc} A_{11}^n & p(A_{11})A_{12} \\ 0 & I \end{array}\right\|,$$

где $p(\lambda)=(\lambda^n-1)/(\lambda-1)$.

1.7. Вектор-столбцы и вектор-строки

$n\times 1$-матрица (имеющая лишь один столбец) определена упорядоченной совокупностью своих n элементов из поля $\mathcal{F}$. Заметим, что операции сложения и умножения на скаляр, примененные к этим матрицам из одного столбца, или *вектор-столбцам*, в точности совпадают с операциями над компонентами, которые определены в $\mathcal{F}_n$. Действительно, пространство всех вещественных вектор-столбцов, например, в точности совпадает с $\mathcal{R}_n$, если только условиться записывать каждый вектор вертикально. Начиная с этого места, в дальнейшем $\boldsymbol{x}$ всегда будет обозначать *вектор-столбец*

$$\left\|\begin{array}{c} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{array}\right\|,$$

а говоря о векторах из $\mathcal{F}_n$, мы будем представлять себе их записанными в виде столбцов.

Мы могли бы равным образом рассматривать элементы из $\mathcal{F}_n$ как матрицы из одной строки, или *вектор-строки*, но мы отдаем предпочтение записи такой матрицы как транспонированной от вектор-столбца. Таким образом, для введенного выше вектор-столбца $\boldsymbol{x}$ имеем $\boldsymbol{x}' = \|x_1\, x_2 \ldots x_n\|$.

Внутреннее произведение векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ из $\mathscr{C}_n$ может быть теперь записано в виде

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \bar{x}_1 y_1 + \bar{x}_2 y_2 + \ldots + \bar{x}_n y_n = \bar{\boldsymbol{x}}' \boldsymbol{y},$$

где $\bar{\boldsymbol{x}}$ — элемент из $\mathscr{C}_n$, получаемый комплексным сопряжением компонент $\boldsymbol{x}$.

Если $\boldsymbol{x}$, $\boldsymbol{y}$ имеют порядки n, m соответственно, то произведение $\boldsymbol{x}\boldsymbol{y}'$ будет $n \times m$-матрицей, которая часто называется *внешним произведением*. Отметим, что $\boldsymbol{y}\boldsymbol{x}'$ имеет размеры $m \times n$ и $(\boldsymbol{y}\boldsymbol{x}')' = \boldsymbol{x}\boldsymbol{y}'$.

Введя все эти обозначения, совокупность *линейных алгебраических уравнений*

$$\begin{array}{llcll}
a_{11}x_1 & + a_{12}x_2 & + \ldots & + a_{1n}x_n & = y_1, \\
a_{21}x_1 & + a_{22}x_2 & + \ldots & + a_{2n}x_n & = y_2, \\
\vdots & \vdots & & \vdots & \vdots \\
a_{m1}x_1 & + a_{m2}x_2 & + \ldots & + a_{mn}x_n & = y_m
\end{array}$$

относительно неизвестных $x_1, x_2, \ldots, x_n$ можем записать в виде

$$A\boldsymbol{x} = \boldsymbol{y}, \tag{1.7.1}$$

где A — $m \times n$-матрица коэффициентов a_{ij} и $\boldsymbol{x}$, $\boldsymbol{y}$ — векторы порядков n, m соответственно Последнее уравнение может также рассматриваться как преобразование вектора $\boldsymbol{x}$ из $\mathscr{F}_n$ в вектор $\boldsymbol{y}$ из $\mathscr{F}_m$. Это наводит на мысль рассматривать A как преобразование $\mathscr{F}_n$ в $\mathscr{F}_m$.

Уравнение (1.7.1) называется *однородным* при $\boldsymbol{y} = \boldsymbol{0}$ и *неоднородным* при $\boldsymbol{y} \neq \boldsymbol{0}$. Решение $\boldsymbol{x}$ однородного уравнения называется *тривиальным* или *нетривиальным* в зависимости от того, будет ли $\boldsymbol{x} = \boldsymbol{0}$ или $\boldsymbol{x} \neq \boldsymbol{0}$.

Мы имеем теперь возможность сформулировать один частный случай теоремы 1.6.1 в векторном виде. Этот случай очень важен, и, так как в дальнейшем предстоит его часто использовать, мы выделяем этот результат отдельным следствием.

Следствие теоремы 1.6.1. *Если матрица AB существует и столбцами матрицы A являются $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$, а строками матрицы B являются $\boldsymbol{b}_1', \boldsymbol{b}_2', \ldots, \boldsymbol{b}_n'$, то*

$$AB = \sum_{k=1}^{n} \boldsymbol{a}_k \boldsymbol{b}_k'.$$

Прямое доказательство следствия намного проще доказательства основной теоремы и читателю предоставляется возможность провести его.

1.8. Аннулируемое подпространство и область значений

Введем еще одно полезное обозначение. Принадлежность к некоторому множеству будем обозначать при помощи символа $\in$ (который не следует смешивать с греческой буквой ε). Таким образом, если x представляет элемент множества A, то мы пишем $x \in A$. Символ $\in$ можно обычно заменять словами «из», «принадлежит» или «является элементом из»

Теорема 1.8.1. *Если $A \in \mathcal{F}_{m \times n}$ и $\mathbf{0}$ — нулевой вектор из $\mathcal{F}_m$, то множество всех решений уравнения $A\boldsymbol{x} = \mathbf{0}$ является подпространством в $\mathcal{F}_n$.*

Доказательство. Отметим сначала, что для существования произведения $A\boldsymbol{x}$ должно быть $\boldsymbol{x} \in \mathcal{F}_n$, т. е. множество векторов, удовлетворяющих уравнению, должно быть подмножеством в $\mathcal{F}_n$. Мы должны показать, что множество всех решений уравнения $A\boldsymbol{x} = \mathbf{0}$ удовлетворяет условиям А1—А5 и Б1—Б5 из § 1.2. Если $A\boldsymbol{x}_1 = \mathbf{0}$ и $A\boldsymbol{x}_2 = \mathbf{0}$, то по свойству дистрибутивности умножения матриц

$$A(\boldsymbol{x}_1 + \boldsymbol{x}_2) = A\boldsymbol{x}_1 + A\boldsymbol{x}_2 = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

так что А1 выполняется. Все другие условия непосредственно следуют из определений операций над матрицами. ◀

Определения. (а) Множество всех векторов $\boldsymbol{x}$, для которых $A\boldsymbol{x} = \mathbf{0}$, есть *аннулируемое подпространство* матрицы A и записывается как $\mathcal{N}(A)$.

(б) Если $A \in \mathcal{F}_{m \times n}$, то $\boldsymbol{x}$ принадлежит *области значений* матрицы A (записывается как $\mathcal{R}(A)$) тогда и только тогда, когда $\boldsymbol{x} = A\boldsymbol{y}$ для некоторого $\boldsymbol{y} \in \mathcal{F}_n$.

Отметим, что по теореме 1.8.1 $\mathcal{N}(A)$ — подпространство. Область значений матрицы A более наглядно описывается словом *образ* пространства $\mathcal{F}_n$ при действии A.

Предположим, что матрица A из $\mathcal{F}_{m \times n}$ имеет столбцы $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, ..., $\boldsymbol{a}_n$ (которые, очевидно, являются элементами из $\mathcal{F}_m$). Тогда из определения умножения матриц следует, что для любого $\boldsymbol{v} \in \mathcal{F}_n$ вектор $A\boldsymbol{v}$ будет линейной комбинацией векторов $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, ..., $\boldsymbol{a}_n$. На самом деле, следовательно, область значений матрицы A просто совпадает с множеством всех линейных комбинаций столбцов из A. По этой причине область значений матрицы A часто в литературе обозначается словами «пространство столбцов» для A.

***Упр. 1.** Будет ли $\mathcal{R}(A)$ подпространством?

Упр. 2. Уравнение $x+2y-3z=0$ можно записать в виде $A\boldsymbol{x}=\mathbf{0}$, где

$$A=\begin{Vmatrix} 1 & 2 & -3 \end{Vmatrix} \quad \text{и} \quad \boldsymbol{x}=\begin{Vmatrix} x \\ y \\ z \end{Vmatrix}.$$

Что такое $\mathcal{N}(A)$? В данном случае, если рассмотреть включение $\mathcal{N}(A)\subseteq\mathcal{R}_3$, имеется очень простая геометрическая интерпретация. Каждое решение уравнения ортогонально вектору $\boldsymbol{a}$ с $\boldsymbol{a}'=\begin{Vmatrix} 1 & 2 & -3 \end{Vmatrix}$. Множество векторов, удовлетворяющих уравнению, совпадает поэтому с множеством всех связанных векторов в $\mathcal{R}_3$, ортогональных к $\boldsymbol{a}$; они образуют плоскость, проходящую через начало координат. Область значений матрицы A совпадает с множеством всех вещественных чисел.

Упр. 3. Найти $\mathcal{N}(A)$, если

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

С геометрической точки зрения $\mathcal{N}(A)$ есть множество всех векторов из $\mathcal{R}_3$, ортогональных *одновременно* к

$$\begin{Vmatrix} 1 \\ -1 \\ 0 \end{Vmatrix} \quad \text{и} \quad \begin{Vmatrix} 1 \\ 1 \\ 1 \end{Vmatrix}.$$

Если какой-нибудь вектор ортогонален к обоим этим векторам, то он ортогонален к любой их линейной комбинации, т. е. к плоскости (подпространству), которую они порождают. Таким образом, можно ожидать, что $\mathcal{N}(A)$ — прямая линия в $\mathcal{R}_3$, проходящая через начало координат.

С алгебраической точки зрения, если

$$\begin{Vmatrix} x \\ y \\ z \end{Vmatrix}$$

из $\mathcal{N}(A)$, то

$$x-y=0, \quad x+y+z=0.$$

Таким образом, $x=y$, $z=-2y$ и каждый вектор из $\mathcal{N}(A)$ представляется в виде

$$\begin{Vmatrix} y \\ y \\ -2y \end{Vmatrix}, \quad \text{или} \quad y\begin{Vmatrix} 1 \\ 1 \\ -2 \end{Vmatrix}$$

при некотором y.

1.9. Линейная зависимость и размерность

Пусть $\mathscr{L}$ — линейное пространство над полем $\mathscr{F}$ и $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$ — векторы из $\mathscr{L}$. Если существуют числа $\alpha_1, \alpha_2, \ldots, \alpha_r$ из поля $\mathscr{F}$, не все равные нулю и такие, что

$$\alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2 + \ldots + \alpha_r\boldsymbol{x}_r = 0, \tag{1.9.1}$$

то векторы $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ *линейно зависимы*. Если из уравнения (1.9.1) следует, что $\alpha_1 = \alpha_2 = \ldots = \alpha_r = 0$, то векторы *линейно независимы*.

Основание для этой терминологии состоит в том, что если (1.9.1) выполняется и, например, α_i отлично от нуля, то уравнение можно преобразовать к виду, где $\boldsymbol{x}_i$ выражен через остальные $r-1$ векторов, и *зависимость* $\boldsymbol{x}_i$ от других векторов получена. Заметим, что в этом случае можно сказать, что $\boldsymbol{x}_i$ лежит в пространстве, порожденном другими $r-1$ векторами. Если из уравнения (1.9.1) следует, что все α равны нулю, то такой зависимости не существует — отсюда термин *независимы*.

Упр. 1. Векторы

$$\boldsymbol{x}_1 = \begin{Vmatrix} 1 \\ 2 \\ 0 \\ 4 \end{Vmatrix}, \quad \boldsymbol{x}_2 = \begin{Vmatrix} -1 \\ 0 \\ 5 \\ 1 \end{Vmatrix}, \quad \boldsymbol{x}_3 = \begin{Vmatrix} 1 \\ 6 \\ 10 \\ 14 \end{Vmatrix}$$

из $\mathscr{R}_4$ удовлетворяют уравнению $3\boldsymbol{x}_1 + 2\boldsymbol{x}_2 - \boldsymbol{x}_3 = 0$ и поэтому линейно зависимы. Любые два из этих векторов линейно независимы.

***Упр. 2.** Если $\mathscr{L}$ — линейное пространство, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r \in \mathscr{L}$ и линейно зависимы, то любое множество векторов из $\mathscr{L}$, содержащее $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$, линейно зависимо.

***Упр. 3.** Если в упр. 2 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$ линейно независимы, то векторы из любого подмножества в $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ линейно независимы.

Упр. 4. Найти нетривиальные линейные соотношения вида (1.9.1), которые выполняются для

$$\text{(а)} \begin{Vmatrix} 2 \\ 1 \\ 1 \end{Vmatrix}, \begin{Vmatrix} 3 \\ -4 \\ 6 \end{Vmatrix}, \begin{Vmatrix} 4 \\ -9 \\ 11 \end{Vmatrix}; \qquad \text{(б)} \begin{Vmatrix} 2 \\ 1 \end{Vmatrix}, \begin{Vmatrix} -1 \\ 3 \end{Vmatrix}, \begin{Vmatrix} 4 \\ 2 \end{Vmatrix};$$

$$\text{(в)} \begin{Vmatrix} 1 \\ 0 \\ 2 \\ 4 \end{Vmatrix}, \begin{Vmatrix} 0 \\ 1 \\ 9 \\ 2 \end{Vmatrix}, \begin{Vmatrix} -5 \\ 2 \\ 8 \\ -16 \end{Vmatrix}.$$

Определения. Конечное множество векторов $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$ является *базисом* для линейного пространства $\mathscr{L}$ в том и только в том случае, когда эти векторы линейно независимы и каждый $\boldsymbol{x} \in \mathscr{L}$ представляется в виде линейной комбинации от $\boldsymbol{x}_i$. (Базис для $\mathscr{L}$ *порождает* $\mathscr{L}$.)

Пусть $\mathscr{L}$ — линейное пространство. Если существует n линейно независимых векторов в $\mathscr{L}$, в то время как любое множество $n+1$ векторов из $\mathscr{L}$ является линейно зависимым, то $\mathscr{L}$ имеет *размерность* n. Если такого n не существует, по $\mathscr{L}$ — *бесконечномерное* пространство.

***Упр. 5.** Пусть $\boldsymbol{e}_i$ — вектор из $\mathscr{F}_n$, у которого на месте i стоит 1 и на всех других местах 0. Векторы $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ образуют базис в $\mathscr{F}_n$. В самом деле, пусть $\boldsymbol{x}$ есть вектор

$$\left\| \begin{matrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{matrix} \right\|$$

из $\mathscr{F}_n$; тогда

$$\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \ldots + x_n\boldsymbol{e}_n$$

и каждый $\boldsymbol{x} \in \mathscr{F}_n$ является линейной комбинацией от $\boldsymbol{e}_i$.

Для доказательства того, что векторы $\boldsymbol{e}_i$ линейно независимы, предположим, что это не так. Тогда существуют такие $\alpha_1, \alpha_2, \ldots, \alpha_n$, не все равные нулю, что $\alpha_1\boldsymbol{e}_1 + \ldots + \alpha_n\boldsymbol{e}_n = \boldsymbol{0}$. Но из этого уравнения, очевидно, следует, что $\alpha_i = 0$, $i = 1, 2, \ldots, n$ (противоречие).

Лемма 1. *Если $A \in \mathscr{F}_{m \times n}$ и $m < n$, то существует нетривиальное решение (с элементами из $\mathscr{F}$) однородного уравнения* $A\boldsymbol{x} = \boldsymbol{0}$.

Доказательство. Оно проводится по индукции относительно m. (Отметим, что $A\boldsymbol{x} = \boldsymbol{0}$ можно интерпретировать как m линейных однородных алгебраических уравнений для неизвестных $x_1, x_2, \ldots, x_n$.) Рассмотрим сначала случай $m = 1$. Тогда $A\boldsymbol{x} = \boldsymbol{0}$ сводится к единственному уравнению вида

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0.$$

Если $A = 0$, то любое $\boldsymbol{x}$ будет решением и результат тривиален. Если $A \neq 0$, то можно предположить (перенумеровывая переменные и коэффициенты, если необходимо), что $a_{11} \neq 0$. Переменным $x_2, x_3, \ldots, x_n$ можно придать тогда произвольные ненулевые значения и положить

$$x_1 = -\frac{1}{a_{11}}(a_{12}x_2 + a_{13}x_3 + \ldots + a_{1n}x_n),$$

что и даст нетривиальное решение исходного уравнения.

Мы предположим теперь, что утверждение верно для $m-1$ уравнений от более чем $m-1$ неизвестных, и покажем, что отсюда следует, что утверждение верно для m уравнений от более чем m неизвестных. Это завершит доказательство по индукции. Рассмотрим уравнение $A\boldsymbol{x}=\boldsymbol{0}$, где A — $m\times n$-матрица и $m<n$. Пусть строки матрицы A будут $\boldsymbol{a}'_1, \boldsymbol{a}'_2, \ldots, \boldsymbol{a}'_m$. Опять можно предположить, что $A\neq 0$. Переставляя уравнения

$$\begin{aligned}
\boldsymbol{a}'_1\boldsymbol{x} &\equiv a_{11}x_1+a_{12}x_2+\ldots+a_{1n}x_n=0,\\
\boldsymbol{a}'_2\boldsymbol{x} &\equiv a_{21}x_1+a_{22}x_2+\ldots+a_{2n}x_n=0,\\
&\vdots\\
\boldsymbol{a}'_m\boldsymbol{x} &\equiv a_{m1}x_1+a_{m2}x_2+\ldots+a_{mn}x_n=0
\end{aligned}\tag{1.9.2}$$

и перенумеровывая неизвестные и коэффициенты, если необходимо, можем считать $a_{11}\neq 0$. Воспользуемся первым уравнением для исключения x_1 из всех других уравнений. В результате получим $m-1$ уравнений от $x_2, x_3, \ldots, x_n$ вида

$$\begin{aligned}
\boldsymbol{a}'_2\boldsymbol{x}-\frac{a_{21}}{a_{11}}(\boldsymbol{a}'_1\boldsymbol{x})&=0,\\
\boldsymbol{a}'_3\boldsymbol{x}-\frac{a_{31}}{a_{11}}(\boldsymbol{a}'_1\boldsymbol{x})&=0,\\
&\vdots\\
\boldsymbol{a}'_m\boldsymbol{x}-\frac{a_{m1}}{a_{11}}(\boldsymbol{a}'_1\boldsymbol{x})&=0.
\end{aligned}$$

Так как $m-1<n-1$ (числа неизвестных), предположение индукции означает, что существует нетривиальное решение $x_2, x_3, \ldots, x_n$ для этих уравнений. Первое из уравнений (1.9.2) может быть использовано теперь для нахождения такого x_1, что $x_1, x_2, \ldots, x_n$ дают нетривиальное решение первоначальных m уравнений. Это завершает индукцию. ◀

Отметим, что в лемме утверждается всего лишь существование решения, в то время как теорема 1.8.1 означает, что решение не может быть единственным.

Теорема 1.9.1. *$\mathscr{F}_n$ — линейное пространство размерности n.*

Доказательство. Обратимся к определению размерности. Как мы видели, существует n линейно независимых векторов в $\mathscr{F}_n$, а именно, $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$. Мы должны доказать, что любые $n+1$ векторов $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{n+1}$ необходимо линейно зависимы. Пусть $X = \| \boldsymbol{x}_1 \boldsymbol{x}_2 \ldots \boldsymbol{x}_{n+1} \|$. Обращаясь к определению линейно зависимых векторов, видим, что $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{n+1}$ линейно независимы в том и только в том случае, когда существует ненулевой вектор

$$\boldsymbol{\alpha} = \begin{Vmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_{n+1} \end{Vmatrix}$$

такой, что

$$\sum_{i=1}^{n+1} \boldsymbol{x}_i \alpha_i = X\boldsymbol{\alpha} = \mathbf{0}.$$

Но X — $n \times (n+1)$-матрица и, согласно лемме, существует ненулевой вектор $\boldsymbol{\alpha}$, удовлетворяющий последнему уравнению. ◀

Из определения размерности теперь немедленно следует, что *любое множество r векторов из $\mathscr{F}_n$ при $r > n$ линейно зависимо.*

Приводимые ниже результаты выражают важные свойства конечномерных линейных пространств. В частности, будет получен результат о том, что все базисы для данного линейного пространства должны содержать одно и то же число векторов и что это число совпадает с размерностью пространства. Следует отчетливо понять, что линейное пространство может иметь разные множества базисных векторов. Например,

$$\begin{Vmatrix} 1 \\ 0 \\ 0 \end{Vmatrix}, \quad \begin{Vmatrix} 0 \\ 1 \\ 0 \end{Vmatrix} \quad \text{и} \quad \begin{Vmatrix} 1 \\ 1 \\ 0 \end{Vmatrix}, \quad \begin{Vmatrix} 0 \\ 1 \\ 0 \end{Vmatrix}$$

будут базисами для одного и того же подпространства в $\mathscr{R}_3$. Другой пример дает упр. 5 из § 1.4.

Лемма 2. *Если линейное пространство $\mathscr{L}$ порождается множеством V элементов, r из которых $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_r$ линейно независимы и каждое множество $r+1$ векторов из V линейно зависимо, то $\mathscr{L}$ порождается векторами $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_r$.*

Доказательство. Мы докажем, что из $\boldsymbol{a}_s \in V$ следует, что $\boldsymbol{a}_s$ выражается в виде линейной комбинации от $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_r$. По предположению $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r, \boldsymbol{a}_s$ линейно зависимы, так что существуют $\alpha_1, \alpha_2, \ldots, \alpha_r, \alpha_s$, не все равные нулю и такие, что

$$\alpha_1 \boldsymbol{a}_1 + \ldots + \alpha_r \boldsymbol{a}_r + \alpha_s \boldsymbol{a}_s = \mathbf{0}.$$

При $\alpha_s = 0$ получаем противоречие с независимостью элементов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$. Следовательно, $\alpha_s \neq 0$, а поэтому любой элемент $\boldsymbol{a}_s$ из V будет линейной комбинацией от $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$. Но каждый элемент из $\mathscr{L}$ является линейной комбинацией элементов из V и, так как каждый элемент из V выражается в виде линейной комбинации от $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$, то каждый элемент из $\mathscr{L}$ будет линейной комбинацией от $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$. ◀

Теорема 1.9.2. *Для линейного пространства $\mathscr{L}$ размерности d каждое множество порождающих $\mathscr{L}$ элементов содержит d и только d линейно независимых векторов.*

Доказательство. Пусть V — множество порождающих $\mathscr{L}$ элементов и $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_r$ — максимальная система линейно независимых векторов из V, т. е. каждое множество $r+1$ векторов из V линейно зависимо. Из определения размерности получаем, что $r \leqslant d$. Покажем теперь, что предположение $r < d$ ведет к противоречию.

Из определения размерности следует существование множества $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_d$ независимых векторов в $\mathscr{L}$. Но по лемме 2 $\mathscr{L}$ порождается $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$, а поэтому существуют константы b_{jk}, $1 \leqslant j \leqslant r$, $1 \leqslant k \leqslant d$, такие, что $\boldsymbol{x}_k = \sum_j b_{jk}\boldsymbol{a}_j$. Таким образом, для любого множества констант $c_1, c_2, \ldots, c_d$ имеем

$$\sum_{k=1}^{d} c_k\boldsymbol{x}_k = \sum_{k=1}^{d}\sum_{j=1}^{r} c_k b_{jk}\boldsymbol{a}_j = \sum_{j=1}^{r}\boldsymbol{a}_j\sum_{k=1}^{d} b_{jk}c_k. \qquad (1.9.3)$$

Для $r \times d$-матрицы B с элементами b_{jk} по лемме 1 получаем существование такого вектора $\boldsymbol{c} \neq \boldsymbol{0}$, что $B\boldsymbol{c} = \boldsymbol{0}$. Поэтому при таком выборе $\boldsymbol{c}$ имеем $\sum_k b_{jk}c_k = 0$ для всех j в правой части равенства (1.9.3) и, следовательно, $\sum_{k=1}^{d} c_k\boldsymbol{x}_k = \boldsymbol{0}$. Это противоречит тому, что x_i линейно независимы, откуда следует, что r *не* меньше d. Следовательно, $r = d$. ◀

Следствия. (а) *Каждый базис в $\mathscr{L}$ содержит d и только d элементов.*

(б) *Любые d независимых векторов в $\mathscr{L}$ образуют базис для $\mathscr{L}$.*

Упр. 6. Получить теорему 1.9.1 из теоремы 1.9.2.

Упр. 7. Показать, что векторы

$$\boldsymbol{x}_1 = \begin{Vmatrix} 1 \\ 0 \\ 2 \\ -1 \end{Vmatrix}, \quad \boldsymbol{x}_2 = \begin{Vmatrix} 0 \\ -1 \\ 2 \\ 0 \end{Vmatrix}$$

линейно независимы и порождают в $\mathscr{R}_4$ подпространство $\mathscr{L}$ размерности 2. Это пространство имеет $\boldsymbol{x}_1, \boldsymbol{x}_2$ в качестве базиса.

Показать, что векторы

$$y_1 = \left\| \begin{matrix} 1 \\ -1 \\ 4 \\ -1 \end{matrix} \right\|, \quad y_2 = \left\| \begin{matrix} -1 \\ -2 \\ 2 \\ 1 \end{matrix} \right\|$$

тоже принадлежат $\mathscr{L}$ и образуют базис в $\mathscr{L}$.

***Упр. 8.** Пусть заданы два подпространства $\mathscr{L}_1$ и $\mathscr{L}_2$ в $\mathscr{L}$ со свойством: любые ненулевые $x \in \mathscr{L}_1$ и $y \in \mathscr{L}_2$ линейно независимы. Показать, что если $x_1, \ldots, x_r \in \mathscr{L}_1$ и линейно независимы и $y_1, \ldots, y_s \in \mathscr{L}_2$ и линейно независимы, то $x_1, \ldots, x_r, y_1, \ldots, y_s$ линейно независимы.

1.10. Свойства базисных векторов

Теорема 1.10.1. *Если векторы $x_1, x_2, \ldots, x_m$ образуют базис линейного пространства $\mathscr{L}$, то каждый вектор из $\mathscr{L}$ может быть однозначно выражен в виде линейной комбинации от $x_1, x_2, \ldots, x_m$.*

Доказательство. Предположим, что теорема неверна и для $x \in \mathscr{L}$ существуют числа a_i, b_i, $i = 1, 2, \ldots, m$, с $a_i \neq b_i$ для некоторого i такие, что

$$x = \sum_{i=1}^{m} a_i x_i = \sum_{i=1}^{m} b_i x_i.$$

Тогда

$$\sum_{i=1}^{m} (a_i - b_i) x_i = \mathbf{0}$$

и по крайней мере один коэффициент $a_i - b_i$ ненулевой. Это означает, что базисные векторы зависимы (противоречие). Следовательно, $a_i = b_i$, $i = 1, 2, \ldots, m$, и теорема доказана. ◀

Теорема 1.10.2. *Пусть $\mathscr{L}$ — линейное пространство размерности m и $x_1, x_2, \ldots, x_r$, $r < m$, — линейно независимые векторы в $\mathscr{L}$. Тогда существуют векторы $x_{r+1}, \ldots, x_m$ в $\mathscr{L}$ такие, что $x_1, x_2, \ldots, x_m$ образуют базис для $\mathscr{L}$.*

Доказательство. Векторы $x_1, x_2, \ldots, x_r$ порождают собственное подпространство $\mathscr{T}_r$ в $\mathscr{L}$, потому что существует такой $x_{r+1} \in \mathscr{L}$, который *не* принадлежит $\mathscr{T}_r$, а поэтому *не* выражается как линейная комбинация от $x_1, x_2, \ldots, x_r$. Рассмотрим теперь уравнение

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_{r+1} x_{r+1} = \mathbf{0}.$$

Здесь $\alpha_{r+1} = 0$, ибо если это не так, то $\boldsymbol{x}_{r+1}$ линейно выразился бы через $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$. Независимость $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ означает тогда, что $\alpha_1 = \alpha_2 = \ldots = \alpha_r = 0$. Следовательно, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{r+1}$ линейно независимы и порождают, например, $\mathscr{T}_{r+1}$.

Если $r + 1 = m$, то доказательство окончено. Если $r+1 < m$, то повторяем рассуждение для получения $\boldsymbol{x}_{r+2}$, для которого $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{r+2}$ линейно независимы, и так далее для $m - r$ шагов, пока не получим $\mathscr{T}_m = \mathscr{L}$. ◀

Упр. 1. Показать, что каждое из следующих подмножеств является линейным пространством над $\mathscr{F}$:

(а) множество векторов в $\mathscr{F}_n$ $(n \geqslant 2)$, первые два элемента которых нулевые;

(б) множество векторов в $\mathscr{F}_n$ $(n \geqslant 2)$, первые два элемента которых равны;

(в) множество векторов в $\mathscr{F}_n$, сумма элементов которых равна нулю.

Упр. 2. Найти размерность и базис для каждого из пространств (а), (б) и (в) в упр. 1.

Мы видели, что *любое* множество r линейно независимых векторов в подпространстве $\mathscr{L} \subseteq \mathscr{F}_n$ размерности r образует базис. Исследуем два таких базиса $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$ и $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_r$. Так как $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$ образуют базис, каждый из $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_r$ является их линейной комбинацией. Таким образом, существуют числа α_{ji} $(1 \leqslant i, \ j \leqslant r)$ такие, что для $1 \leqslant i \leqslant r$

$$\boldsymbol{y}_i = \alpha_{1i}\boldsymbol{x}_1 + \alpha_{2i}\boldsymbol{x}_2 + \ldots + \alpha_{ri}\boldsymbol{x}_r.$$

Если определить $n \times r$-матрицы $X = \|\boldsymbol{x}_1\ \boldsymbol{x}_2 \ldots \boldsymbol{x}_r\|$ и $Y = \|\boldsymbol{y}_1\ \boldsymbol{y}_2 \ldots \boldsymbol{y}_r\|$ и $r \times r$-матрицу A с элементами α_{ji}, то будем иметь

$$\underset{n \times r}{Y} = \underset{(n \times r)}{X}\ \underset{(r \times r)}{A}. \tag{1.10.1}$$

Это можно выразить словами, сказав, что базисные векторы $\boldsymbol{x}_i$ преобразуются в $\boldsymbol{y}_i$ посредством умножения X на матрицу A справа и для любых двух множеств базисных векторов существует такая матрица A. В частности, существует $r \times r$-матрица B такая, что $X = YB$. Отметим, что если $\mathscr{L} = \mathscr{F}_n$, то X, Y, A и B все размера $n \times n$.

Мы имеем также

$$Y = Y(BA), \tag{1.10.2}$$

что приводит нас к вопросу о том, не будут ли преобразующие матрицы A, B таковы, что $BA = I$ (§ 1.5). Мы должны отложить ответ на этот вопрос на конец главы.

1.11. Определение функции определителя

Предполагается, что читатель знаком с основными идеями теории определителей. Поэтому без попыток дать какие-либо обоснования мы начинаем с формального определения и затем приводим наброски доказательств некоторых основных свойств. Для введения общего определения следует сначала исследовать некоторые свойства перестановок.

Пусть $j_1 j_2 \ldots j_n$ — числа $1, 2, \ldots, n$, записанные в любом порядке. Мы называем $j_1 j_2 \ldots j_n$ *перестановкой* чисел $1, 2, \ldots, n$. Обозначим такую перестановку через p и отметим, что существует в точности $n!$ различных перестановок.

Далее, для всех вещественных значений x определим функцию $\operatorname{sign} x$ (читается «сигнум x») так:

$$\operatorname{sign} x = \begin{cases} 1 & \text{для } x > 0, \\ 0 & \text{для } x = 0, \\ -1 & \text{для } x < 0. \end{cases}$$

Операция перестановки двух чисел в p приводит к новой перестановке, называемой *транспозицией*. Предположим теперь, что p может быть приведена к естественному (возрастающему) порядку чисел с помощью $t(p)$ транспозиций*). Мы утверждаем, что если

$$Q(p) = \prod_{1 \leqslant l < k \leqslant n} (j_k - j_l),$$

то

$$\operatorname{sign} Q(p) = (-1)^{t(p)}. \qquad (1.11.1)$$

Для доказательства этого заметим, что произведение $Q(p)$ может быть записано в виде

$$(j_2 - j_1)(j_3 - j_1) \ldots (j_n - j_1)(j_3 - j_2) \ldots (j_n - j_2) \ldots (j_n - j_{n-1}).$$

Ясно, что когда p соответствует естественному порядку чисел, скажем $p = p'$, тогда $\operatorname{sign} Q(p') = 1$.

Предположим, что p преобразуется в p_1 с помощью одной транспозиции. Мы утверждаем, что

$$Q(p_1) = (-1) Q(p). \qquad (1.11.2)$$

Если последнее равенство получено, то ясно, что

$$Q(p') = (-1)^{t(p)} Q(p),$$

а так как $\operatorname{sign} Q(p') = 1$, то отсюда будет следовать (1.11.1). Остается поэтому доказать лишь (1.11.2).

*) Мы предполагаем *существование* $t(p)$.

Предположим, что для получения p_1 из p мы переставляем j_h и j_k. Обозначим через F_1 произведение тех множителей в $Q(p)$, которые не содержат ни j_h, ни j_k. Имеется еще множитель $\pm(j_h - j_k)$, а также множители, содержащие j_h и j_k не вместе. Пусть $\pm F_2$, $\pm F_3$ обозначают произведения множителей, содержащих лишь j_h и лишь j_k соответственно, где

$$F_2 = \prod_{\substack{l=1 \\ l \neq h,\, k}}^{n} (j_l - j_h), \quad F_3 = \prod_{\substack{l=1 \\ l \neq h,\, k}}^{n} (j_l - j_k).$$

Тогда

$$Q(p) = \pm F_1 F_2 F_3 (j_h - j_k).$$

Результат перестановки чисел j_h и j_k теперь ясен. Множитель F_1 в $Q(p)$ остается неизменным, F_2F_3 переходит в F_3F_2 и $j_h - j_k$ заменяется на $j_k - j_h$. Результат (1.11.2) отсюда следует сразу же.

Ясно, что может быть много различных комбинаций транспозиций, которые преобразуют p к естественному порядку чисел. Однако верен следующий важный результат относительно $t(p)$.

Лемма. *Четность числа транспозиций, требуемых для преобразования фиксированной перестановки к естественному порядку, инвариантна относительно выбора транспозиций, т. е. для фиксированной p число $t(p)$ либо всегда нечетное, либо всегда четное.*

Доказательство. Предположим, что одновременно s и t транспозиций могут быть использованы для преобразования p к естественному порядку. Тогда из (1.11.1) следует, что $(-1)^t = (-1)^s$, откуда получаем утверждение. ◀

Получив эту лемму, введем теперь понятие определителя матрицы $A \in \mathcal{F}_{n \times n}$. Мы рассматриваем произведение элементов из A вида

$$a_{1j_1} a_{2j_2} \ldots a_{nj_n},$$

где $j_1\ j_2 \ldots j_n$ — перестановка p чисел $1, 2, \ldots, n$. Таким образом, каждое такое произведение содержит один и только один элемент из каждой строки A и один из каждого столбца A. Далее, существуют $n!$ таких произведений. Мы вводим теперь функцию определителя матрицы A, в записи $\det A$, посредством

$$\det A = \sum_{p} (-1)^{t(p)} a_{1j_1} a_{2j_2} \ldots a_{nj_n}, \tag{1.11.3}$$

где p пробегает все $n!$ перестановок чисел $1, 2, \ldots, n$.

Отметим прежде всего, что лемма была нужна для того, чтобы показать, что $(-1)^{t(p)}$ однозначно определено для каждого члена этой суммы и затем что $\det A$ — функция, определенная на $\mathscr{F}_{n\times n}$ со значениями в $\mathscr{F}$.

***Упр. 1.** Доказать, что для $n=2$

$$\det A = a_{11}a_{22} - a_{12}a_{21},$$

а для $n=3$

$$\det A = a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13}.$$

***Упр. 2.** Если $A = \operatorname{diag}\{a_1, a_2, \ldots, a_n\}$, то $\det A = a_1 a_2 \ldots a_n$.

1.12. Свойства определителей

Использованное выше для введения функции определителя суммирование означает, что в каждом члене суммы имеется один и только один элемент из каждого столбца. Таким образом, если $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ — столбцы матрицы A и, например, $\boldsymbol{a}_1 = \alpha \boldsymbol{b}_1 + \beta \boldsymbol{b}_2$, то

$$\det \| \alpha \boldsymbol{b}_1 + \beta \boldsymbol{b}_2 \ \boldsymbol{a}_2 \ldots \boldsymbol{a}_n \| = \alpha \det \| \boldsymbol{b}_1 \ \boldsymbol{a}_2 \ldots \boldsymbol{a}_n \| + \beta \det \| \boldsymbol{b}_2 \ \boldsymbol{a}_2 \ldots \boldsymbol{a}_n \|$$

и подобно этому для любого другого столбца матрицы A, т. е. можно сказать, что $\det A$ — аддитивная и однородная функция *) от j-го столбца матрицы A $(j = 1, 2, \ldots, n)$. Подобное же утверждение справедливо для строк матрицы A.

Так как каждый член в (1.11.3) содержит элемент из i-й строки, то, собирая вместе члены, содержащие a_{ij} для $j = 1, 2, \ldots, n$, получим выражение вида

$$\det A = a_{i1}A_{i1} + a_{i2}A_{i2} + \ldots + a_{in}A_{in}, \tag{1.12.1}$$

где коэффициенты A_{ij} при a_{ij} называются *алгебраическими дополнениями* для a_{ij}. Этот вид для $\det A$ известен как *разложение* определителя по его i-й строке. Очевидно, что существует подобное выражение для разложения $\det A$ по его i-му столбцу.

Мы сформулируем теперь и докажем пять наиболее важных свойств определителей. Напомним, что A' обозначает транспонированную матрицу к матрице A (§ 1.5).

Свойство 1. $\det A' = \det A$.

Чтобы найти выражение для $\det A'$, мы просто обращаем порядок индексов в выражении (1.11.3) для $\det A$. Таким образом,

$$\det A' = \sum_p (-1)^{t(p)} a_{j_1 1} a_{j_2 2} a_{j_n n}.$$

*) Функция f аддитивна в том и только в том случае, когда $f(a+b) = f(a) + f(b)$ для всех a, b и $a+b$ из области определения f, и f однородна над $\mathscr{F}$, если $f(\alpha x) = \alpha f(x)$ для всех $\alpha \in \mathscr{F}$.

Множители в $a_{j_1 1} a_{j_2 2} \ldots a_{j_n n}$ можно, очевидно, так переставить, что получится член $a_{1k_1} a_{2k_2} \ldots a_{nk_n}$ из $\det A$, за исключением, может быть, знака. Знак, который должен быть приписан этому члену в $\det A$, есть $(-1)^{t(p')}$, где p' — перестановка $k_1 k_2 \ldots k_n$ чисел $1, 2, \ldots, n$. Но p' получается из $1, 2, \ldots, n$ применением $t(p)$ транспозиций, используемых для приведения p к естественному порядку. Следовательно, простым обращением этой последовательности транспозиций p' приводится к естественному порядку. Поэтому $t(p') = t(p)$ и $(-1)^{t(p)} a_{j_1 1} a_{j_2 2} \ldots a_{j_n n}$ будет членом $\det A$. Таким образом, каждый член $\det A'$ является членом $\det A$ и наоборот. Следовательно, $\det A' = \det A$.

Свойство 2. *Если матрица B получается из $n \times n$-матрицы A перестановкой двух строк (или столбцов), то* $\det B = -\det A$.

Для упрощения обозначений мы докажем этот результат для случая, когда B получается из A перестановкой первых двух строк. По существу нет ничего нового в рассмотрении случая перестановки *любых* двух строк. Мы имеем

$$b_{1j_1} = a_{2j_1}, \qquad b_{2j_2} = a_{1j_2}, \qquad b_{rj_r} = a_{rj_r}, \qquad r = 3,\ 4, \ldots, n,$$

а поэтому разложение для $\det B$ таково:

$$\det B = \sum_p (-1)^{t(p)} a_{1j_2} a_{2j_1} a_{3j_3} \ldots a_{nj_n}.$$

Пусть q — перестановка $j_2 j_1 j_3 \ldots j_n$. Тогда можем написать $t(q) + 1 = t(p)$ и

$$\det B = (-1) \sum_q (-1)^{t(q)} a_{1j_2} a_{2j_1} \ldots a_{nj_n}.$$

Но суммирование по всем перестановкам q дает те же самые $n!$ членов, что и суммирование по всем перестановкам p. Таким образом,

$$\det B = (-1) \sum_q (-1)^{t(q)} a_{1j_2} a_{2j_1} \ldots a_{nj_n} = (-1) \det A.$$

Доказав свойство 2 для строк, результат о столбцах получаем сразу же применением свойства 1.

Свойство 3. *Если $n \times n$-матрица A имеет одинаковые две строки (или столбца), то* $\det A = 0$.

Если переставить две одинаковые строки (или столбца) матрицы A, то по свойству 2 получим, что $\det A = -\det A$, откуда следует результат*).

---

*) Приведенное доказательство проходит лишь в предположении, что характеристика поля $\mathcal{F}$ не равна 2. — *Прим. перев.*

Свойство 4. *Если* $s \neq r$, *то*

$$a_{s1}A_{r1} + a_{s2}A_{r2} + \dots + a_{sn}A_{rn} = 0$$

и (1.12.2)

$$a_{1s}A_{1r} + a_{2s}A_{2r} + \dots + a_{ns}A_{nr} = 0.$$

Рассмотрим матрицу B, полученную из A заменой строки r на строку s, $r \neq s$. Тогда B будет иметь одинаковые r-ю и s-ю строки; разлагая $\det B$ по r-й строке, получим первое из нужных равенств. Подобное рассуждение для столбцов приводит ко второму равенству.

Если $A \in \mathscr{F}_{n \times n}$, то *минор порядка* $n-1$ матрицы A — это определитель матрицы, получаемой из A выбрасыванием одной строки и одного столбца. Минор, получаемый выбрасыванием строки i и столбца j, обозначается через M_{ij}, $1 \leqslant i, j \leqslant n$. Следующий результат показывает, что существует тесная и, с первого взгляда, замечательная связь между алгебраическими дополнениями и минорами.

Свойство 5. $A_{ij} = (-1)^{i+j} M_{ij}$.

Мы разобьем доказательство на две части:

1. Сначала рассмотрим случай $i = j = 1$. Пользуясь определением $\det A$, можем написать

$$\det A = \sum_q (-1)^{t(q)} a_{11} a_{2j_2} \dots a_{nj_n} + \sum_{\substack{p \\ j_1 \neq 1}} (-1)^{t(p)} a_{1j_1} a_{2j_2} \dots a_{nj_n},$$

где q — обозначение перестановки $j_2\, j_3 \dots j_n$ чисел $2, 3, \dots, n$. Опять используя (1.11.3), убеждаемся в том, что

$$A_{11} = \sum_q (-1)^{t(q)} a_{2j_2} \dots a_{nj_n} = M_{11},$$

а поэтому утверждение в этом случае верно.

2. Для общих i, j перемещаем сначала элемент a_{ij} в положение $(1, 1)$ посредством $i-1$ последовательных перестановок *соседних* строк, за которыми следуют $j-1$ перестановок соседних столбцов. Обозначим полученную после этого матрицу через B. Соответствующий элементу a_{ij} минор тот же самый как в A, так и в B, потому что *относительные* расположения строк и столбцов в M_{ij} не меняются. Следовательно, как в части 1,

$$\det B = a_{ij} M_{ij} + (\text{члены, не содержащие } a_{ij}).$$

Но по свойству 2 $\det B = (-1)^{i-1}(-1)^{j-1} \det A = (-1)^{i+j} \det A$. Следовательно, $\det A = (-1)^{i+j} a_{ij} M_{ij} + (\text{члены, не содержащие } a_{ij})$. Результат следует теперь из сравнения этого равенства с (1.12.1).

Упр. 1. Вычислить определители следующих матриц:

$$\begin{Vmatrix} 3 & 2 \\ 2 & 5 \end{Vmatrix}, \quad \begin{Vmatrix} 2 & 1 & 2 \\ 0 & 3 & -1 \\ 1 & 2 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 3 & 1 & -1 & 2 \\ 0 & 2 & 0 & 0 \\ 4 & 1 & 0 & 2 \\ -2 & 1 & 3 & 4 \end{Vmatrix}.$$

***Упр. 2.** Квадратная матрица, в которой все элементы выше (ниже) главной диагонали нулевые, называется нижней (верхней) *треугольной* матрицей. Доказать, что если A треугольная, то $\det A$ будет равен произведению элементов главной диагонали матрицы A.

***Упр. 3.** Для $A \in \mathcal{F}_{n\times n}$ и $\lambda \in \mathcal{F}$ доказать, что

$$\det(\lambda A) = \lambda^n \det A.$$

***Упр. 4.** Если $A \in \mathcal{C}_{n\times n}$ и $\bar{A}$ — матрица, элементы которой — сопряженные комплексные числа к элементам матрицы A, то доказать, что

$$\det \bar{A} = \overline{\det A}.$$

Упр. 5. *Матрица Вандермонда* порядка n задается как

$$V = \begin{Vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{Vmatrix}.$$

Доказать, что $\det V = \prod\limits_{i>j} (x_i - x_j)$.

1.13. Присоединенная и обратная матрицы

Предположим, что $A \in \mathcal{F}_{n\times n}$ и, как выше, A_{ij} $(1 \leqslant i, j \leqslant n)$ — алгебраические дополнения a_{ij}. Мы определяем *присоединенную* матрицу к A, в записи $A^{\vee}$, как *транспонированную* к матрице алгебраических дополнений для A. Таким образом,

$$A^{\vee} = \begin{Vmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{Vmatrix}.$$

Рассмотрим матричное произведение $AA^\vee$. Элемент на месте i, j этого произведения есть $\sum_{k=1}^{n} a_{ik}A_{jk}$, что по равенству (1.12.1) и свойству 4 сводится к $\delta_{ij}\det A$. Этим способом мы получаем важный результат:

$$AA^\vee = A^\vee A = \begin{Vmatrix} \det A & 0 & & & \\ 0 & \det A & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & \det A \end{Vmatrix} = (\det A) I.$$

Определения. (i) матрица $A \in \mathscr{F}_{n\times n}$ называется *особой* или *неособой* в зависимости от того, будет ли $\det A$ нулем или нет.

(ii) Если A неособая, то определяем *обратную* матрицу к A, в записи A^{-1}, как

$$A^{-1} = \frac{1}{\det A} A^\vee .$$

Из написанных выше уравнений сразу же следует

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

Значение обратной матрицы лучше всего, пожалуй, иллюстрируется, если обратиться к уравнению (1.7.1) в случае, когда A неособая. Напомним, что (когда A размера $n \times n$) (1.7.1) — сокращенная запись для множества n алгебраических уравнений от n компонент вектора $\boldsymbol{x}$. Если A известна, то, умножая обе части уравнения на A^{-1} слева, видим, что решение может быть сразу же записано в виде $\boldsymbol{x} = A^{-1}\boldsymbol{y}$.

Упр. 1. Если A неособая и $AB = I$, то доказать, что $B = A^{-1}$.

Упр. 2. Доказать, что $I^{-1} = I$.

Упр. 3. Показать, что

$$\begin{Vmatrix} 3 & 1 \\ 2 & 4 \end{Vmatrix}^{-1} = \frac{1}{10} \begin{Vmatrix} 4 & -1 \\ -2 & 3 \end{Vmatrix},$$

и найти обратные (если они существуют) к матрицам из упр. 1 § 1.12.

Упр. 4. Показать, что уравнение $P^{-1} = P$ имеет решением

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

***Упр. 5.** Для нижней треугольной $n \times n$-матрицы A предположим что $a_{ii} \neq 0$, $i = 1, 2, \ldots, n$. Показать, что A^{-1} существует, будет тоже нижней треугольной и имеет диагональные элементы $1/a_{ii}$, $i = 1, 2, \ldots, n$.

Упр. 6. Доказать, что если A^{-1} существует, то $(A^{-1})' = (A')^{-1}$.

Упр. 7. Если A неособая, $B \neq 0$ и AB существует, то доказать, что $AB \neq 0$.

1.14. Формула Бине — Коши

Мы уже определили миноры порядка $n-1$ для $n \times n$-определителя. В более общем случае, если из $m \times n$-матрицы A выбросить все строки, *кроме* строк $i_1, i_2, \ldots, i_p$, и все столбцы, *кроме* столбцов $k_1, k_2, \ldots, k_p$, то определитель полученной в результате матрицы называется минором матрицы A порядка p, и мы пишем

$$A\begin{pmatrix} i_1 & i_2 & \ldots & i_p \\ k_1 & k_2 & \ldots & k_p \end{pmatrix} = \det \left\| \begin{matrix} a_{i_1 k_1} & \ldots & a_{i_1 k_p} \\ \vdots & & \vdots \\ a_{i_p k_1} & \ldots & a_{i_p k_p} \end{matrix} \right\| .$$

Миноры, для которых $i_1 = k_1$, $i_2 = k_2$, $\ldots$, $i_p = k_p$, называются *главными* для матрицы A. Если A — $n \times n$-матрица, то в этих обозначениях имеем

$$\det A = A\begin{pmatrix} 1 & 2 & \ldots & n \\ 1 & 2 & \ldots & n \end{pmatrix},$$

и алгебраическое дополнение A_{1n}, например, есть

$$A_{1n} = (-1)^{n+1} A\begin{pmatrix} 2 & 3 & \ldots & n \\ 1 & 2 & \ldots & n-1 \end{pmatrix}.$$

Отметим, что индексы i и k не обязательно расположены в возрастающем порядке. В действительности мы допускаем, чтобы они принимали повторяющиеся значения — с очевидной интерпретацией, что получающиеся в результате миноры равны нулю (§ 1.12, свойство 3).

Если квадратная матрица является произведением некоторых матриц (которые могут быть прямоугольными), то часто бывает важно иметь возможность выразить определитель произведения в терминах свойств множителей. Следующая теорема — мощный результат этого рода.

Теорема 1.14.1 (формула Бине — Коши). *Пусть A, B — $m \times n$- и $n \times m$-матрицы соответственно, $m \leqslant n$ и $C = AB$.*

Тогда

$$\det C = \sum_{1 \leqslant k_1 < k_2 < \ldots < k_m \leqslant n} A\begin{pmatrix} 1 & 2 & \ldots & m \\ k_1 & k_2 & \ldots & k_m \end{pmatrix} B\begin{pmatrix} k_1 & k_2 & \ldots & k_m \\ 1 & 2 & \ldots & m \end{pmatrix}. \tag{1.14.1}$$

Другими словами, при $m \leqslant n$ определитель матрицы C является суммой произведений всех возможных миноров порядка m в A на соответствующие миноры матрицы B того же самого порядка.

Упр. 1. Проиллюстрируем сначала обозначения примером. Пусть

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

Тогда

$$C = \begin{Vmatrix} 1 & 1 \\ 0 & 2 \end{Vmatrix}$$

и $\det C = 2$. По написанной выше формуле имеем

$$\det C = \sum_{1 \leqslant k_1 < k_2 \leqslant 3} A\begin{pmatrix} 1 & 2 \\ k_1 & k_2 \end{pmatrix} B\begin{pmatrix} k_1 & k_2 \\ 1 & 2 \end{pmatrix} =$$

$$= A\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} B\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} + A\begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} B\begin{pmatrix} 1 & 3 \\ 1 & 2 \end{pmatrix} + A\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} B\begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix} =$$

$$= \det\begin{Vmatrix} 1 & 0 \\ -1 & 1 \end{Vmatrix} \det\begin{Vmatrix} -1 & -1 \\ -2 & 0 \end{Vmatrix} + \det\begin{Vmatrix} 1 & 2 \\ -1 & 1 \end{Vmatrix} \det\begin{Vmatrix} -1 & -1 \\ 1 & 1 \end{Vmatrix} +$$

$$+ \det\begin{Vmatrix} 0 & 2 \\ 1 & 1 \end{Vmatrix} \det\begin{Vmatrix} -2 & 0 \\ 1 & 1 \end{Vmatrix} = 1 \cdot (-2) + 3 \cdot 0 + (-2) \cdot (-2) = 2.$$

Доказательство теоремы. Так как $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$, можно написать

$$C = \begin{Vmatrix} \sum_{\alpha_1=1}^{n} a_{1\alpha_1} b_{\alpha_1 1} & \cdots & \sum_{\alpha_m=1}^{n} a_{1\alpha_m} b_{\alpha_m m} \\ \vdots & & \vdots \\ \sum_{\alpha_1=1}^{n} a_{m\alpha_1} b_{\alpha_1 1} & \cdots & \sum_{\alpha_m=1}^{n} a_{m\alpha_m} b_{\alpha_m m} \end{Vmatrix}.$$

Мы отмечали, что определитель — аддитивная и однородная функция каждого из своих столбцов. Используя этот факт для каждого из m столбцов в $\det C$, выражаем $\det C$ в виде суммы n^m определителей:

$$\det C = \sum_{\alpha_1=1}^{n} \ldots \sum_{\alpha_m=1}^{n} \det \begin{Vmatrix} a_{1\alpha_1} b_{\alpha_1 1} & \ldots & a_{1\alpha_m} b_{\alpha_m m} \\ \vdots & & \vdots \\ a_{m\alpha_1} b_{\alpha_1 1} & \ldots & a_{m\alpha_m} b_{\alpha_m m} \end{Vmatrix} =$$

$$= \sum_{\alpha_1=1}^{n} \ldots \sum_{\alpha_m=1}^{n} A\begin{pmatrix} 1 & 2 & \ldots & m \\ \alpha_1 & \alpha_2 & \ldots & \alpha_m \end{pmatrix} b_{\alpha_1 1} b_{\alpha_2 2} \ldots b_{\alpha_m m}. \quad (1.14.2)$$

Те члены в суммировании, которые имеют совпадающими два или более индексов $\alpha_1, \alpha_2, \ldots, \alpha_m$, равны нулю, так как в этих случаях миноры будут иметь по крайней мере два совпадающих столбца. Таким образом, нужно рассматривать лишь те $n!/(n-m)!$ членов суммирования, в которых индексы α различны. Мы распределяем эти остающиеся члены на $\binom{n}{m}$ групп по $m!$ членов в каждой таким образом, что в каждой группе члены отличаются лишь *порядком* индексов $\alpha_1, \alpha_2, \ldots, \alpha_m$. Отметим также, что можно написать

$$A\begin{pmatrix} 1 & 2 & \ldots & m \\ \alpha_1 & \alpha_2 & \ldots & \alpha_m \end{pmatrix} = (-1)^{t(p)} A\begin{pmatrix} 1 & 2 & \ldots & m \\ k_1 & k_2 & \ldots & k_m \end{pmatrix},$$

где $k_1 < k_2 < \ldots < k_m$ и, в обозначении § 1.11, p — перестановка $\alpha_1 \alpha_2 \ldots \alpha_m$ чисел $k_1, k_2, \ldots, k_m$. Следовательно, сумма по $m!$ членам, в которых $\alpha_1 \alpha_2 \ldots \alpha_m$ — перестановка чисел $k_1, k_2, \ldots, k_m$, задается выражением

$$A\begin{pmatrix} 1 & 2 & \ldots & m \\ k_1 & k_2 & \ldots & k_m \end{pmatrix} \sum_p (-1)^{t(p)} b_{\alpha_1 1} b_{\alpha_2 2} \ldots b_{\alpha_m m}.$$

Переставляя элементы b так, чтобы первые индексы шли в возрастающем порядке, приводим это выражение к виду

$$A\begin{pmatrix} 1 & 2 & \ldots & m \\ k_1 & k_2 & \ldots & k_m \end{pmatrix} \sum_q (-1)^{t(q)} b_{k_1 j_1} b_{k_2 j_2} \ldots b_{k_m j_m},$$

где q — перестановка $j_1 j_2 \ldots j_m$ чисел $1, 2, \ldots, m$ и, как очевидно, $t(q) = t(p)$. Из определения функции определителя теперь следует, что это выражение есть просто

$$A\begin{pmatrix} 1 & 2 & \ldots & m \\ k_1 & k_2 & \ldots & k_m \end{pmatrix} B\begin{pmatrix} k_1 & k_2 & \ldots & k_m \\ 1 & 2 & \ldots & m \end{pmatrix}.$$

Поэтому равенство (1.14.2) сводится к (1.14.1). ◄

Следствие. *Определитель произведения двух квадратных матриц равен произведению определителей множителей.*

Это следует из теоремы при $m=n$.

Упр. 2. Снова может быть получено *неравенство Шварца*. Пусть

$$A=\begin{Vmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \end{Vmatrix} \quad \text{и} \quad B=\begin{Vmatrix} c_1 & d_1 \\ c_2 & d_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ c_n & d_n \end{Vmatrix}, \qquad n\geqslant 2;$$

тогда из теоремы следует *тождество Коши*

$$\det\begin{Vmatrix} \boldsymbol{a'c} & \boldsymbol{a'd} \\ \boldsymbol{b'c} & \boldsymbol{b'd} \end{Vmatrix}=\det\begin{Vmatrix} a_1c_1+\dots+a_nc_n & a_1d_1+\dots+a_nd_n \\ b_1c_1+\dots+b_nc_n & b_1d_1+\dots+b_nd_n \end{Vmatrix}=$$

$$=\sum_{1\leqslant i<k\leqslant n}\det\begin{Vmatrix} a_i & a_k \\ b_i & b_k \end{Vmatrix}\det\begin{Vmatrix} c_i & d_i \\ c_k & d_k \end{Vmatrix}.$$

Полагая $\boldsymbol{a}=\boldsymbol{c}$, $\boldsymbol{b}=\boldsymbol{d}$, получаем

$$\det\begin{Vmatrix} a_1^2+\dots+a_n^2 & a_1b_1+\dots+a_nb_n \\ a_1b_1+\dots+a_nb_n & b_1^2+\dots+b_n^2 \end{Vmatrix}=\sum_{1\leqslant i<k\leqslant n}\left(\det\begin{Vmatrix} a_i & a_k \\ b_i & b_k \end{Vmatrix}\right)^2.$$

Если a_i и b_i — *вещественные числа*, то правая часть в этом тождестве неотрицательна, а поэтому

$$(a_1^2+\dots+a_n^2)(b_1^2+\dots+b_n^2)\geqslant(a_1b_1+\dots+a_nb_n)^2,$$

или

$$(\boldsymbol{a'a})(\boldsymbol{b'c})\geqslant(\boldsymbol{a'b})^2.$$

Упр. 3. Пусть $\boldsymbol{a}$, $\boldsymbol{b}$ — векторы в $\mathscr{C}_n$. Использовать рассуждение в упр. 2 для доказательства теоремы 1.3.1 для $\mathscr{C}_n$.

***Упр. 4.** Для неособой матрицы A показать, что

$$\det(A^{-1})=(\det A)^{-1}.$$

***Упр. 5.** Для неособых $n\times n$-матриц A, B доказать, что $(AB)^{-1}$ существует и $(AB)^{-1}=B^{-1}A^{-1}$.

1.15. Ранг матрицы

Пусть A — $m\times n$-матрица. Обозначим столбцы матрицы A через $A_{*1}, A_{*2}, \dots, A_{*n}$ и ее строки через $A_{1*}, A_{2*}, \dots, A_{m*}$. Тогда

$$A=\|A_{*1}A_{*2}\ \dots\ A_{*n}\|=\begin{Vmatrix} A_{1*} \\ \cdot \\ \cdot \\ \cdot \\ A_{m*} \end{Vmatrix}.$$

Мы определяем *ранг по строкам* для матрицы A как наибольшее число линейно независимых (порядка n) векторов среди A_{i*}, $i = 1, 2, \ldots, m$, и *ранг по столбцам* для A как наибольшее число линейно независимых (порядка m) векторов среди A_{*j}, $j = 1, 2, \ldots, n$.

Теорема 1.15.1. *Для любой матрицы ранг по строкам и ранг по столбцам совпадают.*

Доказательство. Пусть r и c — ранг по строкам и ранг по столбцам соответственно для матрицы $A \in \mathscr{C}_{m\times n}$. Мы утверждаем, что матрица A^*, полученная из A перестановкой строк, будет иметь те же самые значения r и c, что и A. Для r это очевидно. Что касается столбцов, заметим, что столбцы матрицы A удовлетворяют линейному соотношению $\sum_{i=1}^{n} \alpha_i A_{*i} = \mathbf{0}$ в том и только в том случае, когда все p-е элементы каждого из этих векторов удовлетворяют тому же самому линейному соотношению, $1 \leqslant p \leqslant m$. Изменение порядка строк в матрице A не меняет этих соотношений между элементами, а поэтому столбцы в A^* удовлетворяют в точности тем же соотношениям, что и столбцы в A. Следовательно, c тоже не меняется при перестановке строк.

Мы можем предположить теперь, что после перестановки строк, если необходимо, матрица A имеет первые r строк линейно независимые, а последние $m-r$ ее строк являются линейными комбинациями первых r строк. Таким образом, если определить разбиение A:

$$B = \begin{Vmatrix} A_{1*} \\ \cdot \\ \cdot \\ \cdot \\ A_{r*} \end{Vmatrix}, \qquad C = \begin{Vmatrix} A_{r+1*} \\ \cdot \\ \cdot \\ \cdot \\ A_{m*} \end{Vmatrix}, \qquad A = \begin{Vmatrix} B \\ C \end{Vmatrix},$$

то существует такая $(m-r) \times r$-матрица T, что $C = TB$. Мы утверждаем, что матрицы A и B имеют одинаковые ранги по столбцам. Ибо если $A\boldsymbol{x} = \mathbf{0}$, то, очевидно, $B\boldsymbol{x} = \mathbf{0}$, а так как

$$A = \begin{Vmatrix} B \\ TB \end{Vmatrix} \quad \text{и} \quad A\boldsymbol{x} = \begin{Vmatrix} B\boldsymbol{x} \\ TB\boldsymbol{x} \end{Vmatrix},$$

то из $B\boldsymbol{x} = \mathbf{0}$ следует, что $A\boldsymbol{x} = \mathbf{0}$. Таким образом, столбцы матриц A и B удовлетворяют в точности одним и тем же линейным соотношениям, а следовательно, ранги по столбцам для A и B оба равны c.

Но столбцы матрицы B — векторы из $\mathscr{F}_r$, а поэтому $c \leqslant r$. Применяя точно ту же цепочку рассуждений к матрице A', находим, что (ранг по столбцам для A') $\leqslant$ (ранг по строкам для A'). Но строки матрицы A являются столбцами матрицы A' и наоборот. Поэтому $r \leqslant c$, откуда следует, что $r = c$. ◀

Мы можем дать теперь формальное определение: *ранг* некоторой матрицы есть наибольшее число ее линейно независимых строк (или столбцов). Имеется, однако, третий подход к понятию ранга, а именно: *ранг по минорам* определяется как порядок наибольшего отличного от нуля минора матрицы A. Нашей ближайшей целью будет доказательство замечательного результата о том, что ранг по минорам совпадает с рангом.

Лемма 1. *Если A — матрица ранга ρ по минорам и*

$$A\begin{pmatrix} i_1 & \dots & i_\rho \\ k_1 & \dots & k_\rho \end{pmatrix}$$

— не равный нулю минор в A, то любой столбец в A будет линейной комбинацией от $A_{*k_1}, A_{*k_2}, \dots, A_{*k_\rho}$.

Доказательство. Для удобства записи (чтобы избежать двойных индексов) мы предположим, что $i_1, i_2, \dots, i_\rho$ и $k_1, k_2, \dots$ $\dots, k_\rho$ совпадают с $1, 2, \dots, \rho$. Читатель может проверить, что приводимое доказательство применимо к произвольным множествам различных i и различных k. Если A — $m \times n$-матрица, то $\rho \leqslant \min(m, n)$. Рассмотрим

$$A\begin{pmatrix} 1 & 2 & \dots & \rho & k \\ 1 & 2 & \dots & \rho & s \end{pmatrix} = \det \begin{Vmatrix} a_{11} & \dots & a_{1\rho} & a_{1s} \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ a_{\rho 1} & \dots & a_{\rho\rho} & a_{\rho s} \\ a_{k1} & \dots & a_{k\rho} & a_{ks} \end{Vmatrix}.$$

Если $k \leqslant \rho$ или $s \leqslant \rho$, то минор равен нулю, так как в нем имеется повторяющаяся строка или столбец. Если $k > \rho$ и $s > \rho$, то минор равен нулю, так как ранг по минорам есть ρ. Таким образом, минор обращается в нуль для всех возможных выборов индексов k и s.

Пусть $c_1, c_2, \dots, c_\rho, c_s$ — алгебраические дополнения элементов в последней строке написанного выше определителя; разлагая этот определитель по последней строке, получим

$$a_{k1}c_1 + a_{k2}c_2 + \dots + a_{k\rho}c_\rho + a_{ks}c_s = 0$$

для $1 \leqslant k \leqslant m$. Заметим теперь, что коэффициенты c не зависимы от выбора k. Это означает, что

$$A_{*1}c_1 + A_{*2}c_2 + \dots + A_{*\rho}c_\rho + A_{*s}c_s = \mathbf{0}.$$

Но

$$c_s = A\begin{pmatrix} 1 & 2 & \dots & \rho \\ 1 & 2 & \dots & \rho \end{pmatrix} \neq 0,$$

а поэтому A_{*s} — линейная комбинация столбцов $A_{*1}, \ldots, A_{*\rho}$. Это верно для $1 \leqslant s \leqslant n$, следовательно, лемма доказана.

Лемма 2. *В обозначениях леммы* 1 *векторы* $A_{*k_1}, A_{*k_2}, \ldots, A_{*k_\rho}$ *линейно независимы.*

Доказательство. Мы опять считаем, что $i_1 = 1, \ldots, i_\rho = \rho$; $k_1 = 1, \ldots, k_\rho = \rho$. Предположим, что утверждение леммы неверно. Тогда существуют такие не все равные нулю числа $\alpha_1, \ldots, \alpha_\rho$, что

$$A_{*1}\alpha_1 + A_{*2}\alpha_2 + \ldots + A_{*\rho}\alpha_\rho = \mathbf{0}.$$

Пусть B — верхняя левая матрица при $\rho \times \rho$-разбиении матрицы A (т. е. матрица $A\begin{pmatrix} 1 & 2 & \ldots & \rho \\ 1 & 2 & \ldots & \rho \end{pmatrix}$). Тогда будем иметь

$$B\boldsymbol{\alpha} = \mathbf{0} \quad \text{с} \quad \det B \neq 0, \tag{1.15.1}$$

где $\boldsymbol{\alpha}' = \|\alpha_1\alpha_2 \ldots \alpha_\rho\| \neq \mathbf{0}'$. Но матрица B^{-1} существует, а поэтому, умножая обе части уравнения $B\boldsymbol{\alpha} = \mathbf{0}$ на B^{-1}, получим $\boldsymbol{\alpha} = \mathbf{0}$, что противоречит предположению. Следовательно, лемма доказана.

Теорема 1.15.2. *Для любой матрицы ранг по минорам совпадает с рангом.*

Доказательство. Из леммы 1 получаем, что если ρ — ранг по минорам, то ρ столбцов матрицы A порождают пространство, в котором лежат остальные $n - \rho$ столбцов. Следовательно, число линейно независимых столбцов не может превышать ρ. Но по лемме 2 имеем, что число линейно независимых вектор-столбцов в точности равно ρ, которое равно поэтому рангу по столбцам. Последнее же число совпадает с рангом. ◀

Введем теперь сокращение $r(A)$ для обозначения ранга матрицы A.

Теорема 1.15.3. *Если A — $n \times n$-матрица, то* $\det A = 0$ *в том и только в том случае, когда $r(A) < n$.*

Доказательство. Утверждение немедленно следует из теоремы 1.15.2 и определения ранга по минорам.

Следствие 1. *$n \times n$-матрица A неособая в том и только в том случае, когда $r(A) = n$.*

Следствие 2 (обобщение теоремы 1.14.1). *Пусть A, B — $m \times n$- и $n \times m$-матрицы соответственно и $C = AB$. Тогда*

$$\det C = \begin{cases} 0 & \textit{при} \quad m > n, \\ \displaystyle\sum_{1 \leqslant k_1 < k_2 < \ldots k_m \leqslant n} A\begin{pmatrix} 1 & 2 & \ldots & m \\ k_1 & k_2 & \ldots & k_m \end{pmatrix} B\begin{pmatrix} k_1 & k_2 & \ldots & k_m \\ 1 & 2 & \ldots & m \end{pmatrix} & \textit{при} \quad m \leqslant n. \end{cases}$$

Доказательство. Каждый столбец из AB является линейной комбинацией столбцов матрицы A. Следовательно, $r(C) \leqslant r(A)$. Так как A имеет лишь n столбцов, очевидно, что $r(A) \leqslant n$, а поэтому при $m > n$ имеем $r(C) \leqslant n < m$. Но C размера $m \times m$, следовательно, из теоремы следует, что $\det C = 0$. При $m \leqslant n$ результат совпадает с теоремой 1.14.1. ◀

***Упр. 1.** Показать, что размерность пространства $\mathscr{R}(A)$ равна $r(A)$.

***Упр. 2.** Если произведение матриц AB определено, то доказать, что

$$r(AB) \leqslant \min\{r(A),\ r(B)\}.$$

***Упр. 3.** Пусть A — неособая матрица и B, C таковы, что AB, CA определены. Доказать, что $r(AB) = r(B)$ и $r(CA) = r(C)$.

***Упр. 4.** Если сумма матриц $A + B$ определена, то доказать, что

$$r(A + B) \leqslant r(A) + r(B).$$

1.16. Решение уравнений

Мы рассмотрим теперь некоторые важные результаты относительно решений множеств линейных алгебраических уравнений, описанных в § 1.7.

Теорема 1.16.1. *Если $A \in \mathscr{F}_{n\times n}$, то уравнение $A\boldsymbol{x} = \boldsymbol{0}$ имеет нетривиальное решение в том и только в том случае, когда A — особая.*

Доказательство предоставляется читателю.

Теорема 1.16.2. *Если $A \in \mathscr{F}_{m\times n}$, то A имеет ранг r в том и только в том случае, когда $\mathscr{N}(A)$ имеет размерность $n - r$.*

Доказательство. 1) Если A имеет ранг r, то можно предположить для простоты, что A имеет разбиение вида

$$\begin{Vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{Vmatrix},$$

где A_{11} размера $r \times r$ и $\det A_{11} \neq 0$ (теорема 1.15.2). Этого всегда можно достигнуть с помощью перестановки компонент вектора $\boldsymbol{x}$ и столбцов матрицы A и затем строк в A.

Для $j = 1, 2, \ldots, n - r$ пусть $\boldsymbol{e}_j \in \mathscr{F}_{n-r}$ — вектор, имеющий 1 на j-м месте и нули на всех других местах (упр. 5 из § 1.9). Определим векторы

$$\xi_j = \begin{Vmatrix} -A_{11}^{-1}A_{12}\boldsymbol{e}_j \\ \boldsymbol{e}_j \end{Vmatrix},$$

и пусть $\mathscr{M}$ — подпространство в $\mathscr{F}_n$, порожденное этими векторами. Так как ξ_j, очевидно, линейно независимы, $\mathscr{M}$ имеет размерность $n-r$. Мы докажем теорему, показав, что $\mathscr{N}(A)=\mathscr{M}$.

2) Имеем

$$A\xi_j=\left\|\begin{matrix} -A_{12}e_j+A_{12}e_j \\ -A_{21}A_{11}^{-1}A_{12}e_j+A_{22}e_j\end{matrix}\right\|=\left\|\begin{matrix} 0 \\ -A_{21}A_{11}^{-1}A_{12}e_j+A_{22}e_j\end{matrix}\right\|,$$

а так как последние $n-r$ строк матрицы A получаются в виде линейных комбинаций первых r строк, отсюда следует, что $A\xi_j=\mathbf{0}$. Таким образом, $\xi_j\in\mathscr{N}(A)$ для каждого j и, так как $\mathscr{M}$ порождается этими векторами, $\mathscr{M}\subseteq\mathscr{N}(A)$.

3) Обратно, если $x\in\mathscr{N}(A)$ и мы запишем $x'=\|x_1',x_2'\|$, где $x_1\in\mathscr{F}_r$, $x_2\in\mathscr{F}_{n-r}$, то существуют такие числа $c_1, c_2, \ldots, c_{n-r}$, что $x_2=\sum_{j=1}^{n-r}c_je_j$. Но равенство $Ax=\mathbf{0}$ означает, что $x_1=-A_{11}^{-1}A_{12}x_2$, откуда следует, что $x=\sum_{j=1}^{n-r}c_j\xi_j$. Следовательно, $\mathscr{N}(A)\subseteq\mathscr{M}$.

4) Результаты пунктов 2) и 3) означают, что $\mathscr{N}(A)=\mathscr{M}$, а так как $\mathscr{M}$ имеет размерность $n-r$, теорема доказана. ◄

Теорема 1.16.3. *Если $A\in\mathscr{F}_{m\times n}$, то уравнение $Ax=c$ имеет решение в том и только в том случае, когда $m\times(n+1)$-пополненная матрица $B=\|A,c\|$ имеет ранг, равный $r(A)$.*

Доказательство. Если x — решение уравнения $Ax=c$, то $\sum_{i=1}^{n}A_{*i}x_i=c$ и c — линейная комбинация столбцов матрицы A. Следовательно, $r(B)=r(A)$. Обратно, если $r(B)=r(A)$, то c будет линейной комбинацией столбцов матрицы A и коэффициенты этой комбинации дадут решение x нужного уравнения. Теорема доказана. ◄

***Упр. 1.** Для $A\in\mathscr{F}_{m\times n}$ доказать, что сумма размерностей пространств $\mathscr{N}(A)$ и $\mathscr{R}(A)$ равна n.

Упр. 2. Уравнение

$$\left\|\begin{matrix}1 & 2 & 0\\ 2 & -3 & 7\\ 1 & 4 & -2\end{matrix}\right\|\left\|\begin{matrix}x_1\\ x_2\\ x_3\end{matrix}\right\|=\left\|\begin{matrix}11\\ -2\\ -7\end{matrix}\right\|$$

не имеет решения. Заметим сначала, что $r(A)=2$, так как $A_{*3}=2A_{*1}-A_{*2}$, и затем, что $r(B)=3$. Отметим, что c не может быть линейной комбинацией векторов A_{*1} и A_{*2}, так как ортогонален к ним обоим.

Упр. 3. Существуют решения уравнения

$$\begin{Vmatrix} 1 & 2 & 0 \\ 2 & -3 & 7 \\ 1 & 4 & -2 \end{Vmatrix} \begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix} = \begin{Vmatrix} 1 \\ 9 \\ -1 \end{Vmatrix}.$$

В этом случае $\boldsymbol{c} = A_{*1} + A_{*3}$, следовательно, $r(B) = r(A)$.

Следующая теорема позволяет описать множество всех решений уравнения $A\boldsymbol{x} = \boldsymbol{c}$ в терминах некоторого частного решения в комбинации с элементами из $\mathcal{N}(A)$.

Теорема 1.16.4. *Пусть s — некоторое частное решение уравнения $A\boldsymbol{x} = \boldsymbol{c}$. Тогда* (i) *если $t \in \mathcal{N}(A)$,* то *$\boldsymbol{s} + \boldsymbol{t}$ — решение уравнения $A\boldsymbol{x} = \boldsymbol{c}$ и* (ii) *для каждого решения $\boldsymbol{x}$ уравнения $A\boldsymbol{x} = \boldsymbol{c}$ существует такой вектор $\boldsymbol{t} \in \mathcal{N}(A)$, что $\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{t}$.*

Доказательство. Что касается первой части, имеем $A(\boldsymbol{s} + \boldsymbol{t}) = \boldsymbol{c} + \boldsymbol{0} = \boldsymbol{c}$, а следовательно, $\boldsymbol{s} + \boldsymbol{t}$ — вектор-решение уравнения $A\boldsymbol{x} = \boldsymbol{c}$.

С другой стороны, если $A\boldsymbol{x} = \boldsymbol{c}$, то $A(\boldsymbol{x} - \boldsymbol{s}) = \boldsymbol{0}$, следовательно, $\boldsymbol{x} - \boldsymbol{s} \in \mathcal{N}(A)$. Записав $\boldsymbol{x} - \boldsymbol{s} = \boldsymbol{t}$, получим $\boldsymbol{t} \in \mathcal{N}(A)$ и $\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{t}$. ◀

Упр. 4. В упр. 3 можно взять

$$\boldsymbol{s} = \begin{Vmatrix} 1 \\ 0 \\ 1 \end{Vmatrix}.$$

Вектор-решение для уравнения $A\boldsymbol{x} = \boldsymbol{0}$ есть

$$\boldsymbol{t} = \begin{Vmatrix} 2 \\ -1 \\ -1 \end{Vmatrix} \quad \text{и} \quad \boldsymbol{s} + \alpha\boldsymbol{t} = \begin{Vmatrix} 1 + 2\alpha \\ -\alpha \\ 1 - \alpha \end{Vmatrix}$$

является решением уравнения $A\boldsymbol{x} = \boldsymbol{c}$ для любого α.

Теорема 1.16.5. *Для квадратной матрицы A уравнение $A\boldsymbol{x} = \boldsymbol{c}$ имеет единственное решение в том и только в том случае, когда A неособая.*

Доказательство. Если A неособая, то $\boldsymbol{x} = A^{-1}\boldsymbol{c}$ является решением и единственным решением уравнения $A\boldsymbol{x} = \boldsymbol{0}$ будет $\boldsymbol{x} = \boldsymbol{0}$ (теорема 1.16.1). Следовательно, по теореме 1.16.4 каждый вектор-решение уравнения $A\boldsymbol{x} = \boldsymbol{c}$ будет иметь вид $A^{-1}\boldsymbol{c} + \boldsymbol{0} = A^{-1}\boldsymbol{c}$, т. е. имеется лишь одно решение.

Обратно, если решение единственно, то в предыдущей теореме мы можем иметь лишь $\boldsymbol{t} = \boldsymbol{0}$. Поэтому по теореме 1.16.1 матрица A неособая. ◀

Предыдущие теоремы дают нам лишь теоретическую картину относительно решений уравнения $A\boldsymbol{x} = \boldsymbol{c}$. Однако мы не обсудили еще никаких практических приемов для отыскания этих решений. В следующем параграфе мы приводим конструктивный метод для нахождения решения, применимый к случаю, когда матрица A неособая. Более глубокий анализ случая, когда матрица A особая, откладывается до упр. 5 и 6 в конце главы 3. Дальнейшая информация по этому вопросу может быть получена из дополнения 2 об «обобщенно обратных» матрицах.

1.17. Правило Крамера

Мы сформулируем теперь ставшее уже классическим правило Крамера для решения уравнения $\boldsymbol{Ax} = \boldsymbol{c}$ в случае, когда $\det A$ существует и отличен от нуля. Как мы знаем, в этом случае

$$\boldsymbol{x} = A^{-1}\boldsymbol{c} = \frac{1}{\det A} A^{\vee}\boldsymbol{c},$$

т. е.

$$\left\| \begin{matrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{matrix} \right\| = \frac{1}{\det A} \left\| \begin{matrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{matrix} \right\| \left\| \begin{matrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_n \end{matrix} \right\|.$$

Таким образом,

$$x_i = \frac{1}{\det A} \sum_{j=1}^{n} c_j A_{ji}, \qquad i = 1, 2, \dots, n.$$

Но последняя сумма совпадает с разложением $\det A$ по i-му столбцу, если элементы этого столбца заменены на $c_1, c_2, \dots, c_n$. Следовательно,

$$x_1 = \frac{1}{\det A} \det \left\| \begin{matrix} c_1 & a_{12} & \dots & a_{1n} \\ c_2 & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ c_n & a_{n2} & \dots & a_{nn} \end{matrix} \right\|$$

и для $i = 1, 2, \dots, n$ число x_i будет частным двух определителей. Числитель является определителем матрицы, получаемой из A заменой столбца i на $\boldsymbol{c}$, а знаменатель совпадает с $\det A$.

Упр. 1. Решить уравнение

$$\left\| \begin{matrix} 2 & -1 & -2 \\ 4 & 1 & 2 \\ 8 & -1 & 1 \end{matrix} \right\| \left\| \begin{matrix} x_1 \\ x_2 \\ x_3 \end{matrix} \right\| = \left\| \begin{matrix} 5 \\ 1 \\ 5 \end{matrix} \right\|.$$

Имеем $\det A = 18$ и по теореме 1.16.5 решение единственно:

$$x_1 = \frac{1}{18} \det \begin{Vmatrix} 5 & -1 & -2 \\ 1 & 1 & 2 \\ 5 & -1 & 1 \end{Vmatrix} = \frac{18}{18} = 1,$$

$$x_2 = \frac{1}{18} \det \begin{Vmatrix} 2 & 5 & -2 \\ 4 & 1 & 2 \\ 8 & 5 & 1 \end{Vmatrix} = \frac{18}{18} = 1,$$

$$x_3 = \frac{1}{18} \det \begin{Vmatrix} 2 & -1 & 5 \\ 4 & 1 & 1 \\ 8 & -1 & 5 \end{Vmatrix} = \frac{-36}{18} = -2.$$

Читателю следует отметить, что хотя правило Крамера является важным теоретическим результатом и даже полезным в случае, когда порядок матрицы не слишком велик, оно не рекомендуется в качестве целенаправленного алгоритма для численного решения широкого класса уравнений.

Смешанные упражнения

1. Пусть V — линейное пространство всех многочленов от t с вещественными коэффициентами, степень которых не превосходит 2. Пусть M — подпространство в V, порожденное $t-1$, t^2+1, $3t^2+2t+1$. Найти базис для V, содержащий базис для M.

2. Для произвольной матрицы A доказать, что $\det(\bar{A}'A) \geqslant 0$. Показать, как этот результат можно использовать для обобщения неравенства Шварца на $\mathscr{C}_n$.

3. Если A — 6×6-матрица, то какой знак следует приписать следующим членам в разложении $\det A$:

$$a_{23}a_{31}a_{42}a_{56}a_{14}a_{65}, \quad a_{32}a_{43}a_{14}a_{51}a_{66}a_{25}?$$

4. Если элементы $n \times n$-матрицы A — дифференцируемые функции от x и $A = \|\boldsymbol{a}_1\boldsymbol{a}_2 \ldots \boldsymbol{a}_n\|$, то доказать, что

$$\frac{d}{dx}\det A = \det\left\| \frac{d}{dx}\boldsymbol{a}_1\boldsymbol{a}_2 \ldots \boldsymbol{a}_n \right\| + \det\left\| \boldsymbol{a}_1 \frac{d}{dx}\boldsymbol{a}_2 \ldots \boldsymbol{a}_n \right\| + \ldots$$
$$\ldots + \det\left\| \boldsymbol{a}_1\boldsymbol{a}_2 \ldots \frac{d}{dx}\boldsymbol{a}_n \right\|.$$

5. Доказать, что в равенствах (1.10.1) и (1.10.2) $B = A^{-1}$, если $r = n$.

6. Если квадратная матрица A имеет разбиение вида

$$\begin{Vmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{Vmatrix},$$

где A_{11} и A_{22} — квадратные матрицы, то доказать, что $\det A = (\det A_{11})(\det A_{22})$. [Указание. В случае, когда A_{11} — неособая, будем иметь

$$A = \begin{Vmatrix} A_{11} & 0 \\ 0 & I \end{Vmatrix} \begin{Vmatrix} I & 0 \\ 0 & A_{22} \end{Vmatrix} \begin{Vmatrix} I & A_{11}^{-1}A_{12} \\ 0 & I \end{Vmatrix}.]$$

7. Найти базисы для $\mathcal{N}(A)$ и $\mathcal{R}(A)$, если

$$A = \begin{Vmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 \\ 3 & 4 & 3 & 4 \end{Vmatrix}.$$

8. Если A — квадратная матрица и

$$A^2 + 2A + I = 0,$$

то показать, что A неособая. Как бы вы стали вычислять A^{-1}?

9. Дано, что D — диагональная неособая матрица и $D = (I + A)^{-1}A$. Доказать, что A диагональная.

10. Рассмотрим $n \times n$-матрицу

$$M = \begin{Vmatrix} x+\lambda & x & x & \dots & x \\ x & x+\lambda & x & \dots & x \\ \vdots & \vdots & \vdots & & \vdots \\ x & x & x & \dots & x+\lambda \end{Vmatrix}.$$

а) Доказать, что $\det M = \lambda^{n-1}(nx + \lambda)$.

б) Доказать, что M^{-1} (когда она существует) имеет тот же вид, что и M.

11. Пусть A — квадратная матрица. Исследовать ранг матрицы $A^{\vee}$.

12. а) Если A — матрица вида

$$\begin{Vmatrix} 1 & a & 0 & 0 \\ a & 1 & 0 & 0 \\ 1 & a & 1 & a \\ 1 & -1 & a & 1 \end{Vmatrix},$$

то показать, что необходимым условием существования *неединственного* решения уравнения $A\boldsymbol{x} = \boldsymbol{c}$ будет равенство $a = 1$ или $a = -1$.

б) В случае $a = 1$ рассмотреть уравнение, в котором $\boldsymbol{c}' = \| \alpha, \beta, \gamma, 0 \|$. Найти условия на α, β, γ, при выполнении которых неоднозначные решения существуют, и написать общее решение уравнения $A\boldsymbol{x} = \boldsymbol{c}$ в этом случае.

*13. Если $A \in \mathscr{C}_{n\times n}$, $\boldsymbol{x}$, $\boldsymbol{y} \in \mathscr{C}_n$ и α — комплексное число, то доказать, что

$$\det \begin{Vmatrix} A & \boldsymbol{x} \\ \boldsymbol{y}' & \alpha \end{Vmatrix} = \alpha \det A - \boldsymbol{y}' A^{\vee} \boldsymbol{x},$$

а следовательно,

$$\det \begin{Vmatrix} A & A\xi \\ \xi' A & \alpha \end{Vmatrix} = \det A (\alpha - \xi' A \xi).$$

Глава 2

СОБСТВЕННЫЕ ЗНАЧЕНИЯ И СОБСТВЕННЫЕ ВЕКТОРЫ

2.1. Характеристическое уравнение

Предположим, что $A \in \mathcal{F}_{n\times n}$ и $\boldsymbol{x} \in \mathcal{F}_n$. Вектор $A\boldsymbol{x}$ принадлежит $\mathcal{F}_n$ и будет элементом из области значений матрицы A. Нам будут особенно интересны те векторы $\boldsymbol{x}$, которые при умножении на A переходят в кратные им векторы, т. е. такие векторы $\boldsymbol{x} \neq 0$, для которых существует число μ_j из $\mathcal{F}$ с $A\boldsymbol{x} = \mu_j\boldsymbol{x}$. Такой ненулевой вектор $\boldsymbol{x}$ называется *правым собственным вектором* матрицы A и μ_j — соответствующим *собственным значением*. Написанное выше уравнение может быть переписано в виде $(\mu_j I - A)\boldsymbol{x} = 0$. Применяя теорему 1.16.1, получаем, что μ_j — собственное значение матрицы A в том и только в том случае, когда $\det(\mu_j I - A) = 0$. Определение функции определителя означает теперь, что μ_j удовлетворяет полиномиальному уравнению с коэффициентами в $\mathcal{F}$; это уравнение известно как *характеристическое уравнение* матрицы A. Многочлен $c(\mu) \equiv \det(\mu I - A)$ называется *характеристическим многочленом* матрицы A.

Следует заметить, что правый собственный вектор матрицы A, соответствующий данному собственному значению μ_j, не может быть единственным. Правые собственные векторы совпадают в действительности с множеством всех ненулевых элементов из $\mathcal{N}(\mu_j I - A)$.

Изучение свойств характеристического многочлена даст нам первое приближение к изучению свойств собственных значений. Отметим, что при произвольном поле $\mathcal{F}$ собственные значения не обязательно принадлежат $\mathcal{F}$. Это происходит потому, что корни многочлена с коэффициентами в некотором поле $\mathcal{F}$ не обязательно принадлежат $\mathcal{F}$. Классическим примером этого являются многочлены с вещественными коэффициентами, имеющие комплексные корни. Для избежания всех таких трудностей мы будем всегда поле $\mathcal{F}$ предполагать таким, что корни всех многочленов с коэффициентами в $\mathcal{F}$ также принадлежат $\mathcal{F}$.

Поле с таким свойством называется *алгебраически замкнутым*. В частности, отметим, поскольку это касается структуры данной книги, что поле комплексных чисел алгебраически замкнуто, а поле вещественных чисел нет.

Первый результат такого рода уже был отмечен:

Теорема 2.1.1. *Если $A \in \mathcal{F}_{n\times n}$, то $\mu_j \in \mathcal{F}$ — собственное значение матрицы A в том и только в том случае, когда μ_j — корень характеристического многочлена для A.*

Следующим шагом нашего анализа будет исследование коэффициентов характеристического многочлена. Сначала сформулируем вспомогательное утверждение.

Лемма. *Пусть $A \in \mathcal{F}_{n\times n}$ и столбцы $i_1, i_2, \ldots, i_k$ матрицы A совпадают с единичными векторами $\boldsymbol{e}_{i_1}, \boldsymbol{e}_{i_2}, \ldots, \boldsymbol{e}_{i_k}$. Тогда $\det A$ равен главному минору матрицы A, полученному выбрасыванием строк и столбцов $i_1, i_2, \ldots, i_k$.*

Доказательство. Разлагая $\det A$ по столбцу i_1, получаем

$$\det A = (-1)^{i_1+i_1} A\begin{pmatrix} 1 & 2 \ldots i_1-1 & i_1+1 \ldots n \\ 1 & 2 \ldots i_1-1 & i_1+1 \ldots n \end{pmatrix} = $$
$$= A\begin{pmatrix} 1 & 2 \ldots i_1-1 & i_1+1 \ldots n \\ 1 & 2 \ldots i_1-1 & i_1+1 \ldots n \end{pmatrix}.$$

Далее вычисляем этот минор с помощью разложения по столбцу с номером i_2. В результате находим, что $\det A$ равен главному минору, полученному выбрасыванием строк и столбцов i_1 и i_2. Продолжая таким образом разложение по столбцам $i_3, \ldots, i_k$, получаем нужный результат. ◀

Теорема 2.1.2. *Пусть c_r — сумма всех главных миноров порядка r матрицы A, $1 \leqslant r \leqslant n$. Тогда*

$$c(\mu) = \mu^n - c_1\mu^{n-1} + c_2\mu^{n-2} - \ldots + (-1)^n c_n.$$

Доказательство. Пусть $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ — столбцы матрицы A. Тогда, используя единичные векторы, можем записать

$$(-1)^n c(\mu) = \det(A - \mu I) = \det \| \boldsymbol{a}_1 - \mu\boldsymbol{e}_1 \quad \boldsymbol{a}_2 - \mu\boldsymbol{e}_2 \ldots \boldsymbol{a}_n - \mu\boldsymbol{e}_n \|.$$

Так как определитель — однородная линейная функция от столбцов, написанный определитель можно выразить в виде суммы 2^n определителей, имеющих столбцы вида $\boldsymbol{a}_i$ или $-\mu\boldsymbol{e}_i$. Соберем вместе определители, содержащие точно r столбцов вида $-\mu\boldsymbol{e}_i$ для некоторого i. Таких определителей имеется $\binom{n}{r}$ и они получаются из матрицы A заменой r ее столбцов на $(-\mu)\cdot$(единичный вектор) всеми возможными способами. Таким образом, из леммы следует, что каждый из этих определителей равен некоторому главному минору порядка $n-r$ в A, умноженному на $(-\mu)^r$. С другой стороны, каждый главный минор порядка $n-r$ встречается при суммировании. Результат следует теперь из рассмотрения значений $r=0, 1, 2, \ldots, n$. ◀

В частности, отметим, что $c_1 = \sum_{r=1}^{n} a_{rr}$. Эта величина известна как *след* матрицы A и обозначается через $\operatorname{tr}(A)$. Отметим также, что $c_n = \det A$. Последний результат может быть получен непосредственно, если положить $\mu = 0$ в определении характеристического многочлена. Из основного результата теории полиномиальных уравнений следует, что c_r будет также суммой всех произведений по r штук из n корней многочлена $c(\mu)$. Таким образом, если $\mu_1, \mu_2, \ldots, \mu_n$ — собственные значения матрицы A (не обязательно различные), то

$$\operatorname{tr}(A) = \sum_{r=1}^{n} \mu_r \quad \text{и} \quad \det A = \prod_{r=1}^{n} \mu_r.$$

***Упр. 1.** Доказать, что $\det A = 0$ в том и только в том случае, когда A имеет нулевое собственное значение.

Упр. 2. Показать, что для

$$A = \begin{Vmatrix} -3 & 1 \\ -7 & 5 \end{Vmatrix}$$

собственными значениями будут $\mu_1 = 4$, $\mu_2 = -2$.

***Упр. 3.** Исследовать собственные значения диагональных и треугольных матриц, обратив особое внимание на единичную и нулевую матрицы в $\mathcal{F}_{n\times n}$.

Упр. 4. Использовать теорему 2.1.2 для нахождения характеристического многочлена матрицы

$$\begin{Vmatrix} 3 & -1 & 0 \\ 2 & 1 & 1 \\ -1 & 1 & 0 \end{Vmatrix}.$$

Упр. 5. (а) Найти собственные значения матрицы

$$\begin{Vmatrix} 0 & 1 \\ 0 & 0 \end{Vmatrix}$$

и показать, что разные матрицы могут иметь одинаковые собственные значения.

(б) Найти собственные значения матриц

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

***Упр. 6.** Показать, что A и A' имеют одинаковые собственные значения.

Упр. 7. Пусть A — неособая матрица. Исследовать связь между собственными значениями матриц A и A^{-1}.

2.2. Кратность собственного значения

Из предположения, что поле $\mathcal{F}$ алгебраически замкнуто, следует, что характеристический многочлен $c(\mu)$ может быть разложен в произведение n линейных множителей:

$$c(\mu)=\prod_{j=1}^{n}(\mu-\mu_j), \tag{2.2.1}$$

где $\mu_1, \mu_2, \ldots, \mu_n$ — собственные значения матрицы A. Может случиться, что некоторые из этих собственных значений равны. Мы поэтому определяем *кратность* собственного значения μ_j матрицы A как число множителей $\mu-\mu_j$, входящих в разложение (2.2.1) характеристического многочлена матрицы A на линейные множители.

Мы убедились ранее, что μ_j — собственное значение матрицы A в том и только в том случае, когда $\mu_j I-A$ имеет ранг, который меньше n. Нашей целью теперь будет выявление некоторой связи между рангом матрицы $\mu_j I-A$ и кратностью собственного значения μ_j.

Лемма. *$n\times n$-матрица ранга $n-\alpha$ $(1\leqslant\alpha\leqslant n)$ имеет нуль собственным значением кратности $m\geqslant\alpha$.*

Доказательство. Пусть A — $n\times n$-матрица ранга $n-\alpha$. Из определения ранга по минорам следует, что все миноры матрицы A порядка выше $n-\alpha$ равны нулю. Поэтому по теореме 2.1.2 $c_{n-\alpha+1}=\ldots=c_{n-1}=c_n=0$ и характеристический многочлен матрицы A приводится к виду

$$\mu^n-c_1\mu^{n-1}+\ \ldots\ +(-1)^{n-\alpha}c_{n-\alpha}\mu^{\alpha}.$$

Следовательно, кратность нулевого собственного значения $\geqslant\alpha$. ◀

Если в написанном выше характеристическом многочлене $c_{n-\alpha}\neq 0$, то очевидно, что нуль будет собственным значением матрицы A кратности α. Может ли эта кратность превзойти α? Чтобы убедиться в возможности этого, нужно указать лишь подходящий пример. Пусть

$$A=\begin{Vmatrix}0&1\\0&0\end{Vmatrix};$$

тогда $r(A)=1$, так что имеем $n=2$, $\alpha=1$. Но $\det(\mu I-A)=\mu^2$, так что A имеет нуль собственным значением кратности 2, т. е. $m=2>\alpha$.

Читатель может без труда привести пример, в котором $m = \alpha$.

Теорема 2.2.1. *Если $n \times n$-матрица A имеет собственное значение μ_j и $\mu_j I - A$ — матрица ранга $n - \alpha$, то μ_j имеет кратность $m \geqslant \alpha$.*

Доказательство. Из леммы следует, что $\mu_j I - A$ имеет нуль собственным значением кратности $m \geqslant \alpha$. Для любой константы k соотношение

$$\det\{(\mu_j - k) I - (A - kI)\} = 0$$

показывает, что собственные значения матрицы $A - kI$ совпадают с собственными значениями матрицы A, уменьшенными на k. Другими словами, собственные значения матрицы A совпадают с собственными значениями матрицы $A - kI$, увеличенными на k. В частности, собственные значения матрицы A совпадают с собственными значениями матрицы $A - \mu_j I$, увеличенными на μ_j. Таким образом, то, что $A - \mu_j I$ имеет нуль собственным значением кратности m, означает, что A имеет собственное значение μ_j с кратностью m. ◀

Следует заметить, что в формулировке теоремы число α равно размерности пространства $\mathcal{N}(\mu_j I - A)$ (теорема 1.16.2), которое называется *геометрической кратностью* μ_j. Таким образом, теорему можно сформулировать также в таком виде:

Кратность собственного значения не меньше его геометрической кратности.

2.3. Собственные векторы

Как мы видели в упр. 6 из § 2.1, собственные значения матриц A и A' совпадают. Следовательно, если μ_j — собственное значение матрицы A, то существует такое $\mathbf{y} \neq \mathbf{0}$, что

$$(\mu_j I - A')\, \mathbf{y} = \mathbf{0},$$

что можно записать также в виде

$$\mathbf{y}' (\mu_j I - A) = \mathbf{0}.$$

Это соотношение наводит на мысль называть $\mathbf{y}$ *левым собственным вектором матрицы* A, соответствующим собственному значению μ_j.

Поэтому иногда удобно обозначать пространства $\mathcal{N}(\mu_j I - A)$ и $\mathcal{N}(\mu_j I - A')$ как *правое* и *левое собственные пространства* μ_j соответственно.

Следующий наш результат показывает, в частности, что один и тот же вектор не может принадлежать двум правым собственным пространствам, соответствующим разным собственным значениям матрицы A.

Теорема 2.3.1. *Правые собственные векторы, соответствующие различным собственным значениям, линейно независимы.*

Доказательство. Пусть $\mu_1, \mu_2, \ldots, \mu_s$ — различные собственные значения $n \times n$-матрицы A и $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_s$ — правые собственные векторы, соответствующие $\mu_1, \mu_2, \ldots, \mu_s$. Предположим, что существуют такие числа $c_1, c_2, \ldots, c_s$, что

$$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \ldots + c_s\boldsymbol{x}_s = \boldsymbol{0}.$$

Умножим обе части этого равенства слева на $\mu_1 I - A$, отметив, что $(\mu_1 I - A)\boldsymbol{x}_1 = \boldsymbol{0}$, $A\boldsymbol{x}_i = \mu_i\boldsymbol{x}_i$, $i = 2, 3, \ldots, s$. В результате получим

$$(\mu_1 - \mu_2)c_2\boldsymbol{x}_2 + (\mu_1 - \mu_3)c_3\boldsymbol{x}_3 + \ldots + (\mu_1 - \mu_s)c_s\boldsymbol{x}_s = \boldsymbol{0}.$$

Умножая теперь полученное равенство на $\mu_2 I - A$, получим

$$(\mu_1 - \mu_3)(\mu_2 - \mu_3)c_3\boldsymbol{x}_3 + \ldots + (\mu_1 - \mu_s)(\mu_2 - \mu_s)c_s\boldsymbol{x}_s = \boldsymbol{0}.$$

После $s - 1$ таких операций будем иметь

$$(\mu_1 - \mu_s)(\mu_2 - \mu_s)\ldots(\mu_{s-1} - \mu_s)c_s\boldsymbol{x}_s = \boldsymbol{0},$$

откуда следует, что $c_s = 0$. Но порядок собственных значений и собственных векторов произволен, а поэтому можно также доказать, что $c_1 = c_2 = \ldots = c_{s-1} = 0$. Следовательно, векторы $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_s$ линейно независимы. ◀

Очевидно, подобный результат верен и для левых собственных векторов.

2.4. Преобразования подобия и простые матрицы

Матрицы A и B из $\mathscr{F}_{n\times n}$ называются *подобными*, если существует такая неособая матрица $T \in \mathscr{F}_{n\times n}$, что $A = T^{-1}BT$. Можно также сказать, что A получается из B *преобразованием подобия*.

Если $\boldsymbol{x}$ — некоторый элемент из $\mathscr{F}_n$, то мы говорили про вектор $\boldsymbol{y} = A\boldsymbol{x}$, что он является образом $\boldsymbol{x}$ при преобразовании, определенном A. Предположим теперь, что $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_n$ — произвольный базис в $\mathscr{F}_n$, и пусть $Z = \| \boldsymbol{z}_1\, \boldsymbol{z}_2 \ldots \boldsymbol{z}_n \|$; это — неособая матрица, ибо ее ранг по столбцам равен n. Существуют такие числа $a_1, a_2, \ldots, a_n$ и $b_1, b_2, \ldots, b_n$, что

$$\boldsymbol{x} = \sum_{j=1}^{n} a_j\boldsymbol{z}_j = Z\boldsymbol{a} \quad \text{и} \quad \boldsymbol{y} = \sum_{j=1}^{n} b_j\boldsymbol{z}_j = Z\boldsymbol{b}.$$

Можно сказать, что векторы $\boldsymbol{a}$, $\boldsymbol{b}$ представляют $\boldsymbol{x}$, $\boldsymbol{y}$ соответственно в том смысле, что $\boldsymbol{a}$, $\boldsymbol{b}$ определяют $\boldsymbol{x}$, $\boldsymbol{y}$ как их координаты в базисе векторов $\mathbf{z}$. Уравнение $\mathbf{y} = A\mathbf{x}$ переходит теперь в $Z\mathbf{b} = AZ\mathbf{a}$, или $\mathbf{b} = Z^{-1}AZ\mathbf{a}$. Это уравнение дает возможность интерпретировать матрицу $Z^{-1}AZ$, подобную A, как то же самое *преобразование*, что и A, но представленное на языке базиса векторов $\mathbf{z}$. Это — важная алгебраическая идея и она приводит к заключению, что подобные матрицы являются просто представлением одного и того же преобразования на основании разных базисных векторов. Однако приведенное соображение не более чем эвристическое, так как мы не даем (и не будем использовать) формального определения «преобразования».

Заметим, что преобразования подобия симметричны в том смысле, что если A и B подобны, то таковыми же будут B и A. Ибо $A = T^{-1}BT$ означает, что $B = TAT^{-1} = S^{-1}AS$, где $S = T^{-1}$. Более того, если A и B подобны и B и C подобны, то A и C подобны. Читателю следует это проверить. Это свойство известно как транзитивность преобразований подобия*).

Отметим также, что

$$T^{-1}A_1T + T^{-1}A_2T + \ldots + T^{-1}A_kT = \\ = T^{-1}(A_1 + A_2 + \ldots + A_k)T \qquad (2.4.1)$$

и, если p — произвольное целое положительное число,

$$(T^{-1}AT)^p = (T^{-1}AT)(T^{-1}AT)\ldots(T^{-1}AT) = T^{-1}A^pT. \qquad (2.4.2)$$

Если $f(\lambda)$ — скалярный многочлен $a_0\lambda^k + a_1\lambda^{k-1} + \ldots + a_n$, то мы определяем $f(M)$, где M — произвольная квадратная матрица, как

$$f(M) = a_0M^k + a_1M^{k-1} + \ldots + a_{n-1}M + a_nI.$$

При таком определении значения многочлена от матрицы из свойств (2.4.1) и (2.4.2) следует, что

$$f(T^{-1}AT) = T^{-1}f(A)T.$$

Теорема 2.4.1. *Подобные матрицы имеют одинаковые собственные значения.*

Доказательство. Мы докажем, что подобные матрицы имеют одинаковые характеристические многочлены и, следовательно, по теореме 2.1.1 одинаковые собственные значения. Предположим, что $A = T^{-1}BT$, так что A и B подобны, и пусть $c(\mu) = \det(\mu I - A)$ — характеристический многочлен матрицы A.

*) Эти свойства вместе с рефлексивностью (матрица подобна самой себе) означают, что подобие матриц является *отношением эквивалентности.*

Тогда

$$c(\mu) = \det(\mu I - T^{-1}BT) = \det T^{-1}(\mu I - B)T =$$
$$= (\det T^{-1})(\det(\mu I - B))(\det T) = \det(\mu I - B),$$

так как $(\det T^{-1})(\det T) = 1$ (упр. 4 из § 1.14). Таким образом, характеристические многочлены матриц A и B совпадают. ◀

К сожалению, матрицы, имеющие одинаковые собственные значения, не обязательно подобны. Следующий пример это показывает.

Упр. 1. Доказать, что матрицы

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

имеют одинаковые собственные значения, но не подобны.

Для данной $n \times n$-матрицы A мы теперь ставим вопрос: каково максимальное число линейно независимых правых собственных векторов для A? Ясно, что это число не превосходит n. В этой главе нас особенно будет интересовать случай, когда оно равно в точности n. В § 2.3 мы определили максимальное число линейно независимых правых собственных векторов, соответствующих μ (т. е. размерность пространства $\mathcal{N}(\mu I - A)$), как геометрическую кратность μ. Теорема 2.3.1 и упр. 8 из § 1.9 означают тогда, что максимальное число линейно независимых правых собственных векторов A равно в точности сумме геометрических кратностей собственных значений матрицы A. Так как сумма кратностей собственных значений равна n, теорема 2.2.1 означает, что сумма геометрических кратностей либо меньше, либо равна n. Кроме того, у A имеется n линейно независимых правых собственных векторов в том и только в том случае, когда геометрическая кратность каждого собственного значения равна его кратности. Это приводит нас к следующему определению.

Квадратная матрица A называется *простой*, если для каждого собственного значения матрицы A его кратность равна геометрической кратности. Непростая квадратная матрица называется *дефектной*.

Если $A \in \mathscr{C}_{n \times n}$, то можно также сказать, что собственные векторы матрицы A порождают $\mathscr{C}_n$ в том и только в том случае, когда A — простая. Предположим, что A имеет множество независимых правых собственных векторов $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, соответствующих собственным значениям $\mu_1, \mu_2, \ldots, \mu_n$ (не обязательно различным). Тогда имеем

$$A\boldsymbol{x}_j = \mu_j \boldsymbol{x}_j, \qquad j = 1, 2, \ldots, n.$$

Определим $U = \operatorname{diag}\{\mu_1, \mu_2, \ldots, \mu_n\}$ и $X = \|\boldsymbol{x}_1 \boldsymbol{x}_2 \ldots \boldsymbol{x}_n\|$ и заметим, что, так как ранг по столбцам для X равен n, X — неособая. Написанные выше уравнения можно записать теперь в виде матричного уравнения

$$AX = XU, \tag{2.4.3}$$

откуда

$$A = XUX^{-1}. \tag{2.4.4}$$

Таким образом, простая матрица подобна диагональной матрице из своих собственных значений.

Предположим теперь, что заданы диагональная матрица U и неособая матрица X, для которой $A = XUX^{-1}$. Тогда $AX = XU$ и (в обозначениях § 1.15)

$$AX_{*j} = \mu_j X_{*j}, \qquad j = 1, 2, \ldots, n.$$

Таким образом, столбцы матрицы X могут быть интерпретированы как n линейно независимых правых собственных векторов для A с элементами μ в качестве соответствующих собственных значений. Следовательно, A простая. Нами доказана

Теорема 2.4.2. *Некоторая матрица простая в том и только в том случае, когда она подобна диагональной матрице.*

Упр. 2. Доказать, что $n \times n$-матрица с n различными собственными значениями простая.

***Упр. 3.** Если A простая, то доказать, что A' простая и, следовательно, A имеет n линейно независимых левых собственных векторов.

Мы рассмотрим теперь конструкцию для левых собственных векторов, подобную использованной выше для правых собственных векторов. Если A — простая матрица, то существуют такие n линейно независимых левых собственных векторов $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$, что

$$\boldsymbol{y}_j' A = \mu_j \boldsymbol{y}_j', \qquad j = 1, 2, \ldots, n.$$

Полагая $Y = \|\boldsymbol{y}_1 \boldsymbol{y}_2 \ldots \boldsymbol{y}_n\|$, получаем

$$Y'A = UY. \tag{2.4.5}$$

Но из равенства (2.4.3) следует, что $X^{-1}A = UX^{-1}$. Сравнивая эти равенства, замечаем, что строки матрицы X^{-1} являются левыми собственными векторами для A. В частности, элементы $\boldsymbol{y}$ можно выбрать так, что $Y' = X^{-1}$, откуда

$$Y'X = I \quad \text{и} \quad XY' = I.$$

Первое из этих соотношений можно записать в виде

$$\begin{Vmatrix} \boldsymbol{y}_1' \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{y}_n' \end{Vmatrix} \| \boldsymbol{x}_1 \boldsymbol{x}_2 \ldots \boldsymbol{x}_n \| = I$$

или, используя символ Кронекера,

$$\boldsymbol{y}_i' \boldsymbol{x}_j = \delta_{ij}, \qquad i, j = 1, 2, \ldots, n.$$

Мы говорим в таком случае, что два множества векторов $\boldsymbol{x}_1, \ldots$ $\ldots, \boldsymbol{x}_n$ и $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ *квазибиортогональны*. Обоснование этой терминологии будет дано в § 2.6.

Заметим, что каждая из рассматриваемых систем векторов является базисом в $\mathscr{C}_n$ и, если выражать произвольный вектор через один из этих базисов, коэффициенты разложения без труда находятся при помощи другого базиса. Например, если $\boldsymbol{a} \in \mathscr{C}_n$, то существуют такие комплексные числа $\alpha_1, \alpha_2, \ldots, \alpha_n$, что

$$\boldsymbol{a} = \alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \ldots + \alpha_n \boldsymbol{x}_n.$$

Умножая слева это равенство на $\boldsymbol{y}_j'$ и используя условие квазибиортогональности, убеждаемся в том, что

$$\alpha_j = \boldsymbol{y}_j' \boldsymbol{a}, \qquad j = 1, 2, \ldots, n.$$

Теорема 2.4.3. *Если A — простая матрица, то ее правые и левые собственные векторы можно определить таким образом, что $Y'X = I$ и $A = XUY'$.*

Доказательство. Мы уже убедились в том, что $Y'X = I$ и $A = XUX^{-1}$. Полагая $X^{-1} = Y'$ во втором равенстве, завершаем доказательство.

Упр. 4. Показать, что матрица

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

простая, и найти такие X, Y', что $Y'X = I$ и $A = XUY'$.

2.5. Спектральная теорема и многочлены от матриц

Множество всех собственных значений некоторой квадратной матрицы известно как ее *спектр*. Про теорему 2.4.3 можно было бы сказать, что она спектральная, потому что в ней дается выражение A через матрицы, связанные со спектром A. Следующий наш результат обобщает теорему 2.4.3. Утверждение теоремы мы записываем в таком виде, который подчеркивает ее спектральный характер.

Теорема 2.5.1. *Пусть A — простая $n \times n$-матрица и p — скалярный многочлен. Если $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ и $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$ — квазибиортогональные системы правых и левых собственных векторов соответственно, то*

$$p(A) = \sum_{j=1}^{n} p(\mu_j) G_j,$$

где сопутствующие матрицы G_j задаются равенствами

$$G_j = \boldsymbol{x}_j \boldsymbol{y}'_j, \quad j = 1, 2, \ldots, n.$$

Доказательство. Для целого положительного числа r по теореме 2.4.3 имеем

$$A^r = (XUY')(XUY') \ldots (XUY')$$

и $Y'X = I$. Следовательно,

$$A^r = XU^rY' = \| X_{*1} \ \ldots \ X_{*n} \| \begin{Vmatrix} \mu_1^r Y'_{*1} \\ \mu_2^r Y'_{*2} \\ \cdot \\ \cdot \\ \cdot \\ \mu_n^r Y'_{*n} \end{Vmatrix} = \| \boldsymbol{x}_1 \ \ldots \ \boldsymbol{x}_n \| \begin{Vmatrix} \mu_1^r \boldsymbol{y}'_1 \\ \cdot \\ \cdot \\ \cdot \\ \mu_n^r \boldsymbol{y}'_n \end{Vmatrix}.$$

Теорема 1.6.1 (или ее следствие в § 1.7) означает теперь, что

$$A^r = \sum_{j=1}^{n} \mu_j^r \boldsymbol{x}_j \boldsymbol{y}'_j = \sum_{j=1}^{n} \mu_j^r G_j.$$

Если $p(\mu) = p_0 \mu^l + p_1 \mu^{l-1} + \ldots + p_l$, то получаем

$$p(A) = p_0 \sum_j \mu_j^l G_j + p_1 \sum_j \mu_j^{l-1} G_j + \ldots + p_l I.$$

Далее, $XY' = I$ и, следовательно, опять используя теорему 1.6.1, получаем $I = \sum_j G_j$. Таким образом,

$$p(A) = \sum_{j=1}^{n} \left(p_0 \mu_j^l + p_1 \mu_j^{l-1} + \ldots + p_l \right) G_j. \quad \blacktriangleleft$$

Следствие. 1. *Сопутствующие матрицы обладают следующими свойствами:*

(i) $\sum_{j=1}^{n} G_j = I$;

(ii) $G_j G_k = 0, \ j \neq k; \ j, k = 1, 2, \ldots, n$;

(iii) $G_j^2 = G_j, \ j = 1, 2, \ldots, n$.

Свойство (i) было получено выше. Свойства (ii) и (iii) следуют сразу же из квазибиортогональности собственных векторов. Матрица A, для которой $A^2 = A$, называется *идемпотентной.* Таким образом, сопутствующие матрицы идемпотентны. Структура идемпотентных матриц будет подробно проанализирована в § 2.11.

Следствие 2. *Собственные значения матрицы $p(A)$ совпадают с $p(\mu_1), \ldots, p(\mu_n)$, и принадлежащие им собственные подпространства совпадают с соответствующими собственными подпространствами матрицы A.*

Проверка этого утверждения предоставляется читателю. Однако первая часть следствия может быть немедленно обобщена, а именно, можно убрать условие о том, что матрица A простая.

Теорема 2.5.2. *Если $\mu_1, \mu_2, \ldots, \mu_n$ — собственные значения некоторой $n \times n$-матрицы A и p — скалярный многочлен, то собственными значениями матрицы $p(A)$ будут $p(\mu_1), p(\mu_2), \ldots$ $\ldots, p(\mu_n)$.*

Доказательство. Пусть g — многочлен степени l с корнями $\mu^{(1)}, \mu^{(2)}, \ldots, \mu^{(l)}$. Тогда существует такая константа $g_0 \neq 0$, что

$$g(\mu) = g_0 \prod_{j=1}^{l} (\mu^{(j)} - \mu),$$

и мы имеем

$$g(A) = g_0 \prod_{j=1}^{n} (\mu^{(j)} I - A).$$

Если записать $c(\mu) = \det(\mu I - A)$ — характеристическому многочлену матрицы A и вспомнить, что $c(\mu) = \prod_{i=1}^{n} (\mu - \mu_i)$, то будем иметь

$$\det g(A) = g_0^n c(\mu^{(1)})\, c(\mu^{(2)}) \ldots c(\mu^{(l)}) =$$

$$= g_0^n \sum_{i=1}^{n} (\mu^{(1)} - \mu_i) \ldots \sum_{i=1}^{n} (\mu^{(l)} - \mu_i) =$$

$$= g_0^n \prod_{j=1}^{l} (\mu^{(j)} - \mu_1) \ldots \prod_{j=1}^{l} (\mu^{(j)} - \mu_n) = g(\mu_1)\, g(\mu_2) \ldots g(\mu_n).$$

Пусть $g(\mu) = \lambda - p(\mu)$, так что $p(\mu)$ — тоже многочлен от μ степени l. Имеем

$$g(A) = \lambda I - p(A),$$

откуда

$$\det(\lambda I - p(A)) = \{\lambda - p(\mu_1)\}\{\lambda - p(\mu_2)\} \ldots \{\lambda - p(\mu_n)\}.$$

Так как это равенство верно для любого $\lambda \in \mathcal{F}$, нами получено разложение характеристического многочлена матрицы $p(A)$ на линейные множители. Отсюда сразу же следует, что собственными значениями матрицы $p(A)$ будут $p(\mu_1)$, $p(\mu_2)$, ..., $p(\mu_n)$. ◀

Упр. 1. Если A — простая матрица и p — скалярный многочлен, то показать, что $p(A)$ — простая матрица.

Упр. 2. Показать, что матрица

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

простая. Найти сопутствующие матрицы для A и использовать их для нахождения A^{20}.

***Упр. 3.** Если A — простая матрица, то показать, что в обозначениях теоремы 2.5.1 имеем

$$(\mu I - A)^{-1} = \sum_{j=1}^{n} \frac{G_j}{\mu - \mu_j}, \qquad \mu \neq \mu_j.$$

***Упр. 4.** Доказать, что собственные значения идемпотентной матрицы все равны либо единице, либо нулю.

Упр. 5. Матрица A называется *нильпотентной* степени m, если существует такое целое число $m \geqslant 2$, что $A^m = 0$ и $A^{m-1} \neq 0$. Показать, что собственные значения нильпотентной матрицы все равны нулю и что такая матрица не может быть простой.

Упр. 6. Если μ — собственное значение простой матрицы A, то выбор базисных векторов в $\mathcal{N}(\mu I - A)$ не однозначен. Показать, что сумма сопутствующих матриц для μ тем не менее определена однозначно.

Упр. 7. (Мощный метод вычисления собственных значений.) Пусть A — простая матрица из $\mathcal{C}_{n \times n}$ с собственными значениями $\mu_1, \mu_2, \ldots, \mu_n$ и $|\mu_1| > |\mu_j|$, $j = 2, 3, \ldots, n$. Процессы предельного перехода к матрицам и векторам интерпретировать как примененные к соответствующим элементам.

(i) Показать, что $(\mu_1^{-1}A)^r \to G_1$ при $r \to \infty$, где G_1 — сопутствующая матрица собственного значения μ_1, т. е. $G_1 = \boldsymbol{x}_1 \boldsymbol{y}_1'$.

(ii) Для произвольного $\boldsymbol{a}_0 \in \mathcal{C}_n$ определить последовательность векторов $\boldsymbol{a}_0, \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots$ при помощи рекуррентного соотношения

$$\boldsymbol{a}_{r+1} = \frac{1}{k_{r+1}} A \boldsymbol{a}_r, \qquad r = 0, 1, 2, \ldots,$$

где k_{r+1} определяется так, чтобы наибольшая по модулю компонента вектора $\boldsymbol{a}_{r+1}$ была равна 1. Исследовать $\lim \boldsymbol{a}_r$ и $\lim k_r$ при $r \to \infty$, предполагая, что $\boldsymbol{y}_1' \boldsymbol{a}_0 \neq 0$.

(iii) Исследовать собственные значения и собственные векторы матрицы $A_1 = A - \mu_1 G_1$ и показать, что $A_1^r = A^r - \mu_1^r G_1$.

Упр. 8. Использовать упр. 7 для нахождения наибольшего по модулю собственного значения матрицы

$$A = \begin{Vmatrix} 2 & 4 & 6 \\ 8 & 9 & 15 \\ 4 & 16 & 36 \end{Vmatrix} \quad \text{с} \quad \boldsymbol{a}_0 = \begin{Vmatrix} 0 \\ 0 \\ 1 \end{Vmatrix}.$$

(При вычислении векторов $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, ... использовать четыре десятичных знака.)

Упр. 9. Доказать, что если A — простая матрица с характеристическим многочленом $c(\mu)$, то $c(A) = 0$. (Это частный случай теоремы Кели — Гамильтона, в полной общности доказываемой в главе 4.)

2.6. Ортогональные и квазиортогональные векторы

В этом параграфе мы вводим два понятия ортогональности. Оба они имеют важное значение в $\mathscr{C}_n$ и, как мы увидим, совпадают в $\mathscr{R}_n$. Хотя нас в основном интересуют эти два пространства, определения будут даны в виде, годном для более широкого приложения.

Рассмотрим сначала линейное пространство $\mathscr{L}$, в котором определено внутреннее произведение (§ 1.3). Пара ненулевых векторов $\boldsymbol{x}$, $\boldsymbol{y} \in L$ называется *ортогональной*, если $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$. Развивая эту идею, говорим, что множества $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ и $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_k$ из $\mathscr{L}$ *биортогональны* в том и только в том случае, когда

$$\langle \boldsymbol{x}_i, \boldsymbol{y}_j \rangle = \delta_{ij}, \qquad 1 \leqslant i, \quad j \leqslant k.$$

Если, в частности,

$$\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = \delta_{ij}, \qquad 1 \leqslant i, \quad j \leqslant k,$$

то множество $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ *ортонормированное*.

Мы начали главу 1 с формализации некоторых привычных понятий в $\mathscr{R}_3$ и затем использовали их для определения линейного пространства. Идя дальше в этом направлении, определяем *угол* θ между двумя ненулевыми векторами $\boldsymbol{x}$ и $\boldsymbol{y}$ из $\mathscr{L}$ по формуле

$$\theta = \arccos \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\langle \boldsymbol{x}, \boldsymbol{x} \rangle^{1/2} \langle \boldsymbol{y}, \boldsymbol{y} \rangle^{1/2}}.$$

Заметим, что при $\mathscr{L} = \mathscr{C}_n$ внутреннее произведение принимает в общем случае комплексные значения и тогда θ будет комплексным. Это не должно удивлять, ибо в пространстве векторов, определенных комплексными числами, можно ожидать, что углы будут комплексными. В частности, угол между $\boldsymbol{x}$ и $\boldsymbol{y}$ будет комплексным сопряженным к углу между $\boldsymbol{y}$ и $\boldsymbol{x}$. Тем не менее мы по-прежнему можем сказать, что $\boldsymbol{x}$ и $\boldsymbol{y}$ ортогональны, если $\theta = \pi/2$, и что угол между вектором и им самим необходимо равен нулю.

При $\mathscr{L} = \mathscr{R}_n$ неравенство Шварца (теорема 1.3.1) гарантирует, что $|\cos\theta| \leqslant 1$, а поэтому в этом случае можно всегда выбрать $0 \leqslant \theta \leqslant \pi$.

Второе определение ортогональности вводится для линейного пространства $\mathscr{F}_n$. Пара ненулевых векторов $\boldsymbol{x}$, $\boldsymbol{y} \in \mathscr{F}_n$ называется *квазиортогональной*, если $\boldsymbol{x}'\boldsymbol{y} = 0$. Мы говорим, что множества $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ и $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_k$ из $\mathscr{F}_n$ *квазибиортогональны*, в том и только в том случае, когда

$$\boldsymbol{x}_i'\boldsymbol{y}_j = \delta_{ij}, \qquad i, j = 1, 2, \ldots, k.$$

Если, в частности,

$$\boldsymbol{x}_i'\boldsymbol{x}_j = \delta_{ij}, \qquad i, j = 1, 2, \ldots, k,$$

то множество $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ *квазиортонормировано.*

Сравнивая определения ортогональных и квазиортогональных векторов в $\mathscr{R}_n$, убеждаемся в том, что, так как $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \bar{\boldsymbol{x}}'\boldsymbol{y} = \boldsymbol{x}'\boldsymbol{y}$, эти два определения совпадают. То же самое относится к определениям о множествах векторов. Оба приближения могут поэтому считаться обобщениями хорошо знакомой ортогональности в $\mathscr{R}_3$.

Упр. 1. Показать, что единичные векторы $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ в $\mathscr{C}_n$ ортонормированы и квазиортонормированы.

Упр. 2. (а) Построим пример вектора, квазиортогонального к себе самому.

(б) Пусть $\mathscr{L}$ — пространство с внутренним произведением и $\boldsymbol{x} \in \mathscr{L}$. Показать, что если $\boldsymbol{x} \neq \boldsymbol{0}$ и $\boldsymbol{y} = \boldsymbol{x}/\langle \boldsymbol{x}, \boldsymbol{x} \rangle^{1/2}$, то $\langle \boldsymbol{y}, \boldsymbol{y} \rangle = 1$. Мы говорим, что $\boldsymbol{y}$ получен *нормированием* из $\boldsymbol{x}$.

Упр. 3. Показать, что следующая система векторов из $\mathscr{C}_n$ ортонормирована, но не квазиортонормирована:

$$\boldsymbol{x}_1 = \frac{1}{\sqrt{2}} \left\| \begin{matrix} 0 \\ 0 \\ -i \\ i \end{matrix} \right\|, \qquad \boldsymbol{x}_2 = \frac{1}{\sqrt{2}} \left\| \begin{matrix} 0 \\ 0 \\ 1 \\ 1 \end{matrix} \right\|, \qquad \boldsymbol{x} = \frac{1}{\sqrt{3}} \left\| \begin{matrix} 1+i \\ i \\ 0 \\ 0 \end{matrix} \right\|.$$

Найти такие числа $\alpha_1, \alpha_2, \alpha_3$, что $\alpha_1\boldsymbol{x}_1$, $\alpha_2\boldsymbol{x}_2$ и $\alpha_3\boldsymbol{x}_3$ будут квазиортонормированы.

Упр. 4. Для $A \in \mathcal{F}_{n\times n}$ с $\det A \neq 0$ запишем $\boldsymbol{x}_j = A_{*j}$, $\boldsymbol{y}'_j = (A^{-1})_{j*}$, $j = 1, 2, \ldots, n$. Показать, что системы векторов $\{\boldsymbol{x}_i\}$, $\{\boldsymbol{y}_j\}$ квазибиортогональны.

***Упр. 5.** Для $A \in \mathcal{F}_{n\times n}$ и $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{C}_n$ доказать, что

$$\langle \boldsymbol{x}, A\boldsymbol{y}\rangle = \langle A^*\boldsymbol{x}, \boldsymbol{y}\rangle,$$

где $A^* = \bar{A}'$, транспонированная сопряженная к матрице A.

Упр. 6. Если $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{C}_n$, то доказать, что $|\boldsymbol{x}'\boldsymbol{y}|^2 \leqslant \langle \boldsymbol{x}, \boldsymbol{x}\rangle\langle \boldsymbol{y}, \boldsymbol{y}\rangle$.

Теорема 2.6.1. (а) *Если $\mathscr{L}$ — линейное пространство, в котором определено внутреннее произведение, и множества векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ и $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ в $\mathscr{L}$ биортогональны, то векторы $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ линейно независимы.*

(б) *Если множества векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ и $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ принадлежат $\mathcal{F}_n$ и квазибиортогональны, то $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ линейно независимы.*

Доказательство. (а) Пусть $\alpha_1, \alpha_2, \ldots, \alpha_k$ — числа, для которых

$$\alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2 + \ldots + \alpha_k\boldsymbol{x}_k = \mathbf{0}. \tag{2.6.1}$$

Тогда для $r = 1, 2, \ldots, k$ имеем

$$0 = \langle \alpha_1\boldsymbol{x}_1 + \ldots + \alpha_k\boldsymbol{x}_k, \boldsymbol{y}_r\rangle = \bar{\alpha}_1\langle \boldsymbol{x}_1, \boldsymbol{y}_r\rangle + \ldots + \bar{\alpha}_k\langle \boldsymbol{x}_k, \boldsymbol{y}_r\rangle = \bar{\alpha}_r,$$

так как $\langle \boldsymbol{x}_s, \boldsymbol{y}_r\rangle = \delta_{sr}$ по свойству биортогональности. Таким образом, $\alpha_r = 0$ для $r = 1, 2, \ldots, k$ и, значит, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ линейно независимы.

(б) В этом случае умножаем обе части равенства (2.6.1) слева на $\boldsymbol{y}'_r$. Из квазиортогональности тогда опять следует, что $\alpha_r = 0$. Так как это верно для $r = 1, 2, \ldots, k$, элементы опять линейно независимы. ◀

Очевидно, что множество ортонормированных (или квазиортонормированных) векторов линейно независимо. Это — частный случай теоремы, в котором $\boldsymbol{y}_r = \boldsymbol{x}_r$, $r = 1, 2, \ldots, k$.

Мы не будем глубоко заниматься сходством между ортогональностью и квазиортогональностью. Более плодотворно сосредоточиться теперь на понятиях ортогональности и ортонормированных систем.

2.7. Ортонормированные системы

Во всем этом параграфе $\mathscr{L}$ будет обозначать линейное пространство, в котором определено внутреннее произведение. Мы хотим показать, что для заданного множества из k линейно независимых векторов в $\mathscr{L}$ можно построить некоторое множество из k ортонормированных векторов, беря линейные комбинации в заданном множестве. Другими словами, если $\mathscr{L}_k$ — подпространство в $\mathscr{L}$, порожденное линейно независимыми векторами $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$, то существуют ортонормированные векторы $\boldsymbol{z}_1$, $\boldsymbol{z}_2, \ldots, \boldsymbol{z}_k$ в $\mathscr{L}_k$. Мы построим сначала множество ненулевых векторов $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_k$ в $\mathscr{L}_k$, для которых $\langle \boldsymbol{y}_r, \boldsymbol{y}_s\rangle = 0$ при $r \neq s$, $1 \leqslant r, s \leqslant k$. Ортонормированное множество тогда строится из этого при помощи

$$\boldsymbol{z}_r = \frac{\boldsymbol{y}_r}{\langle \boldsymbol{y}_r, \boldsymbol{y}_r\rangle^{1/2}}, \qquad r = 1, 2, \ldots, k.$$

Пусть $\mathscr{L}_p$ — подпространство, порожденное $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p$, где $1 \leqslant p \leqslant k$; тогда $\mathscr{L}_1 \subset \mathscr{L}_2 \subset \ldots \subset \mathscr{L}_k$ и $\boldsymbol{x}_r$ не принадлежит $\mathscr{L}_p$ при $r > p$. Полагая $\boldsymbol{y}_1 = \boldsymbol{x}_1$, находим затем вектор $\boldsymbol{y}_2 \in \mathscr{L}_2$, не принадлежащий $\mathscr{L}_1$ и для которого $\langle \boldsymbol{y}_1, \boldsymbol{y}_2\rangle = 0$. Записав

$$\boldsymbol{y}_2 = a_{12}\boldsymbol{y}_1 + \boldsymbol{x}_2,$$

определяем a_{12} из условия $\langle \boldsymbol{y}_1, \boldsymbol{y}_2\rangle = 0$. Должно быть

$$a_{12} = -\frac{\langle \boldsymbol{y}_1, \boldsymbol{y}_2\rangle}{\langle \boldsymbol{y}_1, \boldsymbol{y}_1\rangle},$$

так что

$$\boldsymbol{y}_2 = -\frac{\langle \boldsymbol{y}_1, \boldsymbol{x}_2\rangle}{\langle \boldsymbol{y}_1, \boldsymbol{y}_1\rangle}\boldsymbol{y}_1 + \boldsymbol{x}_2.$$

Далее, равенство $\boldsymbol{y}_2 = \boldsymbol{0}$ означало бы, что $\boldsymbol{x}_2 \in \mathscr{L}_1$, что не так. Следовательно, $\boldsymbol{y}_2 \neq \boldsymbol{0}$.

Мы могли бы теперь продолжить рассуждение, положив

$$\boldsymbol{y}_3 = a_{13}\boldsymbol{y}_1 + a_{23}\boldsymbol{y}_2 + \boldsymbol{x}_3,$$

но, применяя индукцию, мы запишем сразу для $r = 2, 3, \ldots, k$

$$\boldsymbol{y}_r = a_{1r}\boldsymbol{y}_1 + \ldots + a_{r-1,r}\boldsymbol{y}_{r-1} + \boldsymbol{x}_r, \tag{2.7.1}$$

где

$$\left.\begin{array}{l}\text{(i) } \langle \boldsymbol{y}_i, \boldsymbol{y}_j\rangle = 0, \ i \neq j;\\ \text{(ii) } \boldsymbol{y}_i \neq \boldsymbol{0};\\ \text{(iii) } \boldsymbol{y}_i \in \mathscr{L}_i \text{ и } \boldsymbol{y}_i \notin \mathscr{L}_{i-1},\end{array}\right\} \quad i, j = 1, 2, \ldots, r-1.$$

Мы докажем, что константы $a_{1r}, \ldots, a_{r-1,r}$ могут быть выбраны так, что (i), (ii) и (iii) будут выполняться для $i, j = 1, 2, \ldots, r$.

Определение (2.7.1) сразу же дает, что (iii) выполняется для $i = r$. Придав теперь константам значения

$$a_{ir} = -\frac{\langle \boldsymbol{y}_i, \boldsymbol{x}_r\rangle}{\langle \boldsymbol{y}_i, \boldsymbol{y}_i\rangle}, \qquad i = 1, 2, \ldots, r-1,$$

без труда проверяем, что (i) выполняется для $i, j = 1, 2, \ldots, r$. Используя (iii) при $i = r$, убеждаемся в том, что равенство

$\boldsymbol{y}_r = \mathbf{0}$ противоречило бы тому, что $\boldsymbol{y}_r \notin L_{r-1}$, а поэтому $\boldsymbol{y}_r \neq \mathbf{0}$. Таким образом, может быть найдена нужная последовательность $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_k$ и ортонормированная система $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k$ может быть получена при помощи нормирования из элементов $\boldsymbol{y}$. Описанная конструкция известна как процесс ортогонализации Грама — Шмидта.

Теорема 2.7.1. *Каждое подпространство в $\mathscr{L}$ размерности k содержит ортонормированный базис.*

Доказательство. Начиная с любого множества базисных векторов $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ для подпространства, ортонормированный базис можно получить при помощи процесса Грама — Шмидта. Полученное ортонормированное множество будет тогда базисом, потому что его члены линейно независимы (теорема 2.6.1). ◀

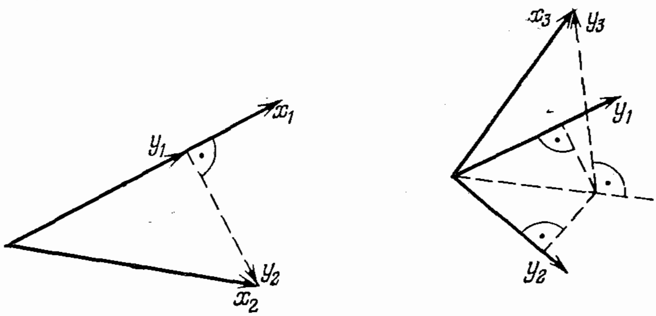

Рис. 3.

В помощь интуиции можно дать геометрическую интерпретацию получению элементов $\boldsymbol{y}$, использованных при процессе ортогонализации. Полагаем опять $\boldsymbol{y}_1 = \boldsymbol{x}_1$ и затем записываем (рис. 3)

$$\boldsymbol{y}_2 = \boldsymbol{x}_2 - (\text{проекция } \boldsymbol{x}_2 \text{ на } \boldsymbol{y}_1).$$

Если $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{L}$ и $\boldsymbol{y} \neq \mathbf{0}$, то, обобщая упр. 2 из § 1.3, определяем проекцию $\boldsymbol{x}$ на $\boldsymbol{y}$ как вектор $\{\langle \boldsymbol{y}, \boldsymbol{x}\rangle / \langle \boldsymbol{y}, \boldsymbol{y}\rangle\}\boldsymbol{y}$. Таким образом,

$$\boldsymbol{y}_2 = \boldsymbol{x}_2 - \frac{\langle \boldsymbol{y}_1, \boldsymbol{x}_2\rangle}{\langle \boldsymbol{y}_1, \boldsymbol{y}_1\rangle}\boldsymbol{y}_1.$$

В качестве $\boldsymbol{y}_3$ мы используем элемент из $\mathscr{L}_3$, получаемый в виде

$$\boldsymbol{y}_3 = \boldsymbol{x}_3 - \{(\text{проекция } \boldsymbol{x}_3 \text{ на } \boldsymbol{y}_1) + (\text{проекция } \boldsymbol{x}_3 \text{ на } \boldsymbol{y}_2)\} =$$
$$= \boldsymbol{x}_3 - \frac{\langle \boldsymbol{y}_1, \boldsymbol{x}_3\rangle}{\langle \boldsymbol{y}_1, \boldsymbol{y}_1\rangle}\boldsymbol{y}_1 - \frac{\langle \boldsymbol{y}_2, \boldsymbol{x}_3\rangle}{\langle \boldsymbol{y}_2, \boldsymbol{y}_2\rangle}\boldsymbol{y}_2.$$

Продолжая этот способ получения векторов далее, приходим к тому же самому множеству векторов, которое было получено выше при алгебраическом процессе ортогонализации.

Поучительно описать в общих чертах третий подход к задаче ортогонализации. Чтобы определить взаимно ортогональное множество $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_k$ в $\mathscr{L}_k$, мы опять берем $\boldsymbol{y}_1 = \boldsymbol{x}_1$ и затем полагаем

$$\boldsymbol{y}_r = a_{1r}\boldsymbol{x}_1 + a_{2r}\boldsymbol{x}_2 + \ldots + a_{r-1,r}\boldsymbol{x}_{r+1} + \boldsymbol{x}_r, \qquad r = 2, 3, \ldots, k.$$

Таким образом, как и прежде, $\boldsymbol{y}_r \in \mathscr{L}_r$ и $\boldsymbol{y}_r \notin \mathscr{L}_{r-1}$. Для определения коэффициентов $a_{1r}, \ldots, a_{r-1,r}$ теперь мы наложим условие, что $\boldsymbol{y}_r$ ортогонален к каждому вектору из $\mathscr{L}_{r-1}$. Так как $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{r-1}$ порождают $\mathscr{L}_{r-1}$, это означает, что $\langle \boldsymbol{x}_j, \boldsymbol{y}_r\rangle = 0$ для $j = 1, 2, \ldots, r-1$. Таким образом,

$$\begin{aligned}
&\langle \boldsymbol{x}_1, \boldsymbol{x}_1\rangle a_{1r} + \langle \boldsymbol{x}_1, \boldsymbol{x}_2\rangle a_{2r} + \ldots + \langle \boldsymbol{x}_1, \boldsymbol{x}_{r-1}\rangle a_{r-1,r} = -\langle \boldsymbol{x}_1, \boldsymbol{x}_r\rangle,\\
&\qquad\qquad\vdots\\
&\langle \boldsymbol{x}_{r-1}, \boldsymbol{x}_1\rangle a_{1r} + \langle \boldsymbol{x}_{r-1}, \boldsymbol{x}_2\rangle a_{2r} + \ldots + \langle \boldsymbol{x}_{r-1}, \boldsymbol{x}_{r-1}\rangle a_{r-1,r} =\\
&\qquad\qquad = -\langle \boldsymbol{x}_{r-1}, \boldsymbol{x}_r\rangle.
\end{aligned}$$

Это множество уравнений относительно $a_{1r}, a_{2r}, \ldots, a_{r-1,r}$ имеет единственное решение тогда и только тогда, когда определитель из коэффициентов не равен нулю. Этот определитель есть *грамиан* или *определитель Грама* от $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{r-1}$. В записи

$$G(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{r-1}) = \det \begin{Vmatrix} \langle \boldsymbol{x}_1, \boldsymbol{x}_2\rangle & \ldots & \langle \boldsymbol{x}_1, \boldsymbol{x}_{r-1}\rangle \\ \vdots & & \vdots \\ \langle \boldsymbol{x}_{r-1}, \boldsymbol{x}_1\rangle & \ldots & \langle \boldsymbol{x}_{r-1}, \boldsymbol{x}_{r-1}\rangle \end{Vmatrix}.$$

В случае $\mathscr{L} = \mathscr{C}_n$ можно определить $n \times (r-1)$-матрицу $X = \| \boldsymbol{x}_1 \boldsymbol{x}_2 \ldots \boldsymbol{x}_{r-1} \|$; тогда

$$G(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{r-1}) = \det(X^*X),$$

где X^* — сопряженная транспонированная матрица к X. Согласно следующей теореме этот определитель положителен или равен нулю в зависимости от того, будут ли векторы $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{r-1}$ линейно независимы или зависимы. По предположению эти векторы линейно независимы, так что их грамиан не равен нулю и для $a_{1r}, a_{2r}, \ldots, a_{r-1,r}$ имеется нетривиальное решение. Следовательно, по индукции мы можем завершить процесс ортогонализации.

Теорема 2.7.2 (критерий Грама). *Векторы $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k \in \mathscr{C}_n$ линейно независимы или зависимы согласно тому, будет ли $G(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k)$ положителен или равен нулю.*

Доказательство этой теоремы является простым применением формулы Бине — Коши (теорема 1.14.1). Детали этого предоставляются читателю.

Теорема 2.7.3. *Если $X \in \mathscr{C}_{n\times k}$ и $r(X) = k$, то существуют такие неособая верхняя треугольная матрица T в $\mathscr{C}_{k\times k}$ и матрица $Z \in \mathscr{C}_{n\times k}$ с ортонормированными столбцами, что $X = ZT$.*

Доказательство. Напомним, что при процессе ортогонализации взаимно ортогональные векторы $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_k$ ненулевые и имеют вид

$$\boldsymbol{y}_r = a_{1r}\boldsymbol{x}_1 + a_{2r}\boldsymbol{x}_2 + \ldots + a_{r-1,r}\boldsymbol{x}_{r-1} + \boldsymbol{x}_r, \qquad r = 1, 2, \ldots, k,$$

а следовательно, существуют такие числа $t_{1r}, t_{2r}, \ldots, t_{rr}$, что векторы

$$\boldsymbol{z}_r = t_{1r}\boldsymbol{x}_1 + t_{2r}\boldsymbol{x}_2 + \ldots + t_{rr}\boldsymbol{x}_r, \qquad r = 1, 2, \ldots, k, \qquad (2.7.2)$$

ортонормированы. Пусть теперь $X_{*j} = \boldsymbol{x}_j$ и $Z_{*j} = \boldsymbol{z}_j$, $j = 1, 2, \ldots, k$, так что столбцы в Z ортонормированы и столбцы в X линейно независимы. Кроме того, равенства (2.7.2) можно записать в виде $Z = XT_1$, где

$$T_1 = \begin{Vmatrix} t_{11} & t_{12} & \ldots & t_{1k} \\ & t_{22} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & t_{kk} \end{Vmatrix}.$$

Но $t_{11}, \ldots, t_{kk}$ все не равны нулю, так что T_1 — неособая и имеет верхнюю треугольную обратную $T_1^{-1} = T$ (упр. 5 из § 1.13). Следовательно, существует матрица T с требуемыми свойствами, для которой $X = ZT$. ◀

Упр. 1. Построить ортонормированный базис для подпространства в $\mathscr{R}_4$, порожденного векторами

$$\begin{Vmatrix} 1 \\ 1 \\ 0 \\ 0 \end{Vmatrix}, \quad \begin{Vmatrix} 0 \\ 1 \\ 1 \\ 0 \end{Vmatrix} \quad \text{и} \quad \begin{Vmatrix} 0 \\ 1 \\ 0 \\ 1 \end{Vmatrix}.$$

Упр. 2. (i) Если $\boldsymbol{x} \in \mathscr{C}_n$ и $\boldsymbol{x}'\boldsymbol{y} = 0$ для каждого $\boldsymbol{y} \in \mathscr{C}_n$, то доказать, что $\boldsymbol{x} = \boldsymbol{0}$.

(ii) Если $\boldsymbol{x} \in \mathscr{C}_n$ и $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$ для каждого $\boldsymbol{y} \in \mathscr{C}_n$, то доказать, что $\boldsymbol{x} = \boldsymbol{0}$.

***Упр. 3.** Если $A \in \mathscr{C}_{n\times n}$ и $\langle \boldsymbol{x}, A\boldsymbol{y} \rangle = 0$ для всех $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{C}_n$, то доказать, что $A = 0$.

2.8. Специальные типы матриц

Мы предполагаем всюду в следующих определениях, что $A \in \mathscr{C}_{n\times n}$ и имеет элементы a_{jk}, $1 \leqslant j, k \leqslant n$. Строки матрицы A обозначаются через $A_{1*}, \ldots, A_{n*}$ и столбцы — через $A_{*1}, \ldots, A_{*n}$.

Матрица A называется *симметрической*, если $A = A'$. Это условие и условия $a_{jk} = a_{kj}$, $1 \leqslant j, k \leqslant n$, как легко видеть, эквивалентны. Третья эквивалентная формулировка: $A_{*j} = (A_{j*})'$, $j = 1, 2, \ldots, n$. Следует заметить, что не существует различия между правыми и левыми собственными векторами симметрической матрицы.

Матрица A называется *кососимметрической*, если $A = -A'$. В этом случае имеем $a_{jk} = -a_{kj}$ для $j \neq k$ и $a_{jj} = 0$. Еще одна формулировка: $A_{*j} = -(A_{j*})'$.

Упр. 1. Доказать, что если A — симметрическая матрица, то таковой же будет и A^r для любого целого положительного числа r. Показать на примере, что если A и B симметрические, то AB не обязательно симметрическая.

Упр. 2. Доказать, что:

(i) если μ — собственное значение кососимметрической матрицы, то таковым же будет и $-\mu$;

(ii) кососимметрическая матрица нечетного порядка особая.

Начиная с этого места, A^* будет обозначать матрицу, получаемую из A транспонированием матрицы, элементы которой комплексно сопряжены к элементам A. Таким образом, $A^* = \bar{A}'$. Матрица A называется *эрмитовой*, если $A = A^*$. Опять имеем эквивалентные формулировки:

$$a_{jk} = \bar{a}_{kj}, \qquad 1 \leqslant j, k \leqslant n,$$

и

$$A_{*j} = (\bar{A}_{j*})', \qquad j = 1, 2, \ldots, n.$$

Заметим, что вещественная симметрическая матрица (т. е. симметрическая матрица с вещественными элементами) будет эрмитовой. Матрица A, для которой $A = -A^*$, называется *косоэрмитовой*.

Матрица A называется *ортогональной*, если $A'A = I$. Как очевидно, для ортогональной матрицы $(\det A)^2 = 1$, так что такая матрица неособая. Определяющее уравнение можно поэтому записать в виде $A^{-1} = A'$ или $AA' = I$. Соотношение $A'A = I$ также означает, что

$$A'_{*j} A_{*k} = \delta_{jk}, \qquad 1 \leqslant j, k \leqslant n.$$

Таким образом, столбцы ортогональной матрицы образуют квазиортонормированную систему векторов. То же утверждение справедливо для строк ортогональной матрицы.

Матрица A называется *унитарной*, если $A^*A = I$. Без труда получаем, что унитарная матрица неособая, что $A^{-1} = A^*$ и что $AA^* = I$. Заметим, что вещественная ортогональная матрица унитарна.

Рассуждение, только что примененное к ортогональным матрицам, может быть использовано для доказательства того, что столбцы (или строки) унитарной матрицы образуют ортонормированную систему базисных векторов для $\mathscr{C}_n$. Наоборот, если столбцы некоторой квадратной матрицы U ортонормированы, то $U^*U = I$ и эта матрица унитарна. Теорема 2.7.3 означает теперь: если X — произвольная неособая матрица, то существуют такие неособая верхняя треугольная матрица T и унитарная матрица U, что $X = UT$.

Упр. 3. Можно ли последнее утверждение расширить до включения случая, когда матрица X особая?

Упр. 4. Будет ли множество всех эрмитовых матриц порядка n подпространством в $\mathscr{C}_{n\times n}$? Дать ответ на тот же вопрос для вещественных симметрических матриц в $\mathscr{R}_{n\times n}$.

Упр. 5. При $A \in \mathscr{C}_{m\times n}$, $\boldsymbol{x} \in \mathscr{C}_m$ и $\boldsymbol{y} \in \mathscr{C}_n$ показать, что

$$\langle \boldsymbol{x}, A\boldsymbol{y} \rangle = \langle A^*\boldsymbol{x}, \boldsymbol{y} \rangle.$$

2.9. Эрмитовы матрицы

Свойства и структура эрмитовых матриц особенно просты и изящны, а так как такие матрицы часто возникают в приложениях, познакомиться с ними в полном объеме очень важно. Множество всех эрмитовых матриц порядка n — собственное подмножество в $\mathscr{C}_{n\times n}$, структура элементов которого более полно будет изучена в главе 4. Ни одно из свойств, получаемых в этом параграфе для эрмитовых матриц, в общем виде не выполняется в $\mathscr{C}_{n\times n}$.

Теорема 2.9.1. *Собственные значения эрмитовой матрицы вещественны.*

Доказательство. Пусть A — эрмитова матрица с собственным значением μ. Тогда существует такой вектор $\boldsymbol{x} \neq \boldsymbol{0}$, что

$$A\boldsymbol{x} = \mu\boldsymbol{x},$$

откуда, беря сопряженное равенство и транспонируя обе его части, получаем

$$\boldsymbol{x}^*A = \bar{\mu}\boldsymbol{x}^*,$$

так как $A^* = A$. Умножая первое из уравнений слева на $\boldsymbol{x}^*$ и второе справа на $\boldsymbol{x}$, находим, что $\mu\boldsymbol{x}^*\boldsymbol{x} = \bar{\mu}\boldsymbol{x}^*\boldsymbol{x}$. Но $\boldsymbol{x} \neq \boldsymbol{0}$, так что $\boldsymbol{x}^*\boldsymbol{x} > 0$, откуда получаем $\mu = \bar{\mu}$. ◀

Следствие. *Собственные значения вещественной симметрической матрицы вещественны.*

Как мы знаем, для произвольной квадратной матрицы правые собственные векторы, *соответствующие различным собственным значениям*, обязательно линейно независимы (теорема 2.3.1). Если матрица эрмитова, то можно утверждать нечто большее.

Теорема 2.9.2. *Правые собственные векторы эрмитовой матрицы, соответствующие различным собственным значениям, ортогональны.*

Доказательство. Пусть λ, μ — различные собственные значения эрмитовой матрицы A, и предположим, что $A\boldsymbol{x} = \lambda\boldsymbol{x}$ и $A\boldsymbol{y} = \mu\boldsymbol{y}$, где $\boldsymbol{x}$, $\boldsymbol{y} \neq \boldsymbol{0}$. Теорема 2.9.1 означает, что μ вещественно, так что, беря сопряженные и транспонируя, второе равенство можем записать в виде $\boldsymbol{y}^*A = \mu\boldsymbol{y}^*$. Умножая это равенство на $\boldsymbol{x}$ справа и равенство $A\boldsymbol{x} = \lambda\boldsymbol{x}$ на $\boldsymbol{y}^*$ слева, получаем

$$\boldsymbol{y}^*A\boldsymbol{x} = \lambda\boldsymbol{y}^*\boldsymbol{x} = \mu\boldsymbol{y}^*\boldsymbol{x},$$

откуда $(\lambda - \mu)\boldsymbol{y}^*\boldsymbol{x} = 0$. Так как $\lambda \neq \mu$ и $\boldsymbol{y}^*\boldsymbol{x} = \langle \boldsymbol{y}, \boldsymbol{x}\rangle$, получаем, что $\boldsymbol{x}$ и $\boldsymbol{y}$ ортогональны.

Теорема 2.9.3. *$n \times n$-матрица A эрмитова в том и только в том случае, когда* $\langle A\boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, A\boldsymbol{y}\rangle$ для *всех* $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{C}_n$.

Доказательство. Предположим сначала, что A эрмитова. Тогда для любых $\boldsymbol{x}$, $\boldsymbol{y} \in \mathscr{C}_n$

$$\langle A\boldsymbol{x}, \boldsymbol{y}\rangle = (A\boldsymbol{x})^* \boldsymbol{y} = \boldsymbol{x}^*A\boldsymbol{y} = \langle \boldsymbol{x}, A\boldsymbol{y}\rangle.$$

Наоборот, если дано, что $\langle A\boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, A\boldsymbol{y}\rangle$ для всех $\boldsymbol{x}$, $\boldsymbol{y} \in \mathscr{C}_n$, то, так как $\langle A\boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, A^*\boldsymbol{y}\rangle$, имеем

$$\langle \boldsymbol{x}, (A^* - A)\,\boldsymbol{y}\rangle = 0$$

для всех $\boldsymbol{x}$, $\boldsymbol{y} \in \mathscr{C}_n$. Нетрудно убедиться в том (упр. 3 из § 2 7), что это означает выполнение равенства $A^* - A = 0$, и, следовательно, A эрмитова. ◀

Отметим, что приведенная характеризация эрмитовых матриц наводит на мысль дать одно общее определение для преобразований на произвольном линейном пространстве с внутренним произведением. И действительно, рассмотренное свойство используется для определения *самосопряженных* операторов в теории гильбертовых пространств.

В начальных параграфах этой главы мы некоторое время занимались исследованием понятия простой матрицы. Мы докажем теперь, что эрмитовы матрицы обязательно простые. Таким образом, {эрмитовы матрицы порядка n} $\subset$ {простые матрицы порядка n} $\subset \mathscr{C}_{n \times n}$.

Теорема 2.9.4. *Каждая эрмитова матрица простая.*

Доказательство. Доказательство разбивается на три отдельных шага. (1) Пусть A — эрмитова матрица с собственным значением μ. Так как μ вещественно (теорема 2.9.1), то $B = \mu I - A$ тоже эрмитова. Шаг 1 состоит в доказательстве того, что $\mathscr{N}(B) = \mathscr{N}(B^2)$.

Сразу же ясно, что из $B\boldsymbol{x} = \boldsymbol{0}$ следует $B^2\boldsymbol{x} = \boldsymbol{0}$, так что $\mathscr{N}(B) \subseteq \mathscr{N}(B^2)$. Мы покажем, что на самом деле здесь имеется равенство, доказав, что каждый элемент из $\mathcal{N}(B^2)$ принадлежит также $\mathcal{N}(B)$. Из теоремы 2.9.3 получаем, что

$$\langle B^2\boldsymbol{x}, \boldsymbol{x}\rangle = \langle B\boldsymbol{x}, B\boldsymbol{x}\rangle,$$

так что из $B^2\boldsymbol{x} = \boldsymbol{0}$ следует $\langle B\boldsymbol{x}, B\boldsymbol{x}\rangle = 0$, следовательно, $B\boldsymbol{x} = \boldsymbol{0}$ (аксиома B1 для внутреннего произведения). Таким образом, $\mathcal{N}(B) = \mathcal{N}(B^2)$.

(2) Предположим, что μ имеет геометрическую кратность α, так что $\mathcal{N}(B)$ имеет размерность α. Согласно шагу 1 $\mathcal{N}(B^2)$ имеет размерность α. Мы покажем, что это означает, что алгебраическая кратность μ равна тоже α.

Рассмотрим некоторый главный минор матрицы B^2 порядка p. Его можно записать в виде

$$B^2\begin{pmatrix} i_1 & i_2 & \dots & i_p \\ i_1 & i_2 & \dots & i_p \end{pmatrix} = \det\left(\begin{Vmatrix} B_{i_1 *} \\ \cdot \\ \cdot \\ \cdot \\ B_{i_p *} \end{Vmatrix} \begin{Vmatrix} B_{*i_1} & \dots & B_{*i_p} \end{Vmatrix}\right),$$

а так как B эрмитова, то $B_{i_k *} = \overline{B}_{*i_k}$. Поэтому (в обозначениях теоремы 2.7.2)

$$B^2\begin{pmatrix} i_1 & i_2 & \dots & i_p \\ i_1 & i_2 & \dots & i_p \end{pmatrix} = G(B_{*i_1}, \dots, B_{*i_p}) \geqslant 0$$

и все главные миноры матрицы B^2 неотрицательны. Кроме того, так как B имеет ранг $n-\alpha$, B имеет $n-\alpha$ линейно независимых столбцов, так что (по теореме 2.7.2) существуют положительные главные миноры у B^2 порядков 1, 2, ..., $n-\alpha$. Из теоремы 2.1.2 теперь следует, что характеристический многочлен для B^2 имеет вид

$$\mu^n - c_1\mu^{n-1} + \dots \pm c_{n-\alpha}\mu^{\alpha},$$

где $c_{n-\alpha} \neq 0$. Таким образом, B^2 имеет нулевое собственное значение с алгебраической кратностью α, и теорема 2.5.2 тогда означает, что B должна тоже обладать этим свойством.

(3) Так как $B = \mu I - A$, то мы только что доказали, что собственное значение μ матрицы A имеет одинаковые кратность и геометрическую кратность. Но это верно для любого собственного значения матрицы A, поэтому, согласно определению простой матрицы, A простая. ◀

Из теоремы 2.4.2 теперь следует, что эрмитова матрица подобна вещественной диагональной матрице ее собственных значений. На самом деле мы в § 2.4 видели, как, исходя из собственных векторов данной матрицы, может быть построена трансформирующая матрица. То, что правые собственные векторы эрмитовой матрицы, соответствующие различным собственным значениям, ортогональны, позволяет нам в данном случае сказать о трансформирующей матрице нечто большее. Но сначала: мы говорим, что матрицы A и B *унитарно* (*ортогонально*) *подобны*, если существует такая унитарная (ортогональная) матрица U, что $A = U^{-1}BU = U^*BU$.

Теорема 2.9.5. *Эрмитова матрица унитарно подобна диагональной матрице ее собственных значений.*

Доказательство. Пусть μ_j — собственное значение кратности α_j эрмитовой матрицы A порядка n. Теоремы 2.7.1 и 2.9.4 означают, что существуют α_j ортонормированных правых собственных векторов матрицы A, соответствующих μ_j. Для каждого отдельного собственного значения матрицы A строим такое множество векторов. Из теоремы 2.9.2 тогда следует, что получающееся в результате множество из n векторов $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ может быть предположено ортонормированным. В таком случае матрица $X = \| \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \|$ будет унитарной. Но последняя матрица как раз совпадает с трансформирующей матрицей X равенства (2.4.4), так что утверждение доказано.

Следствие. *Вещественная симметрическая матрица ортогонально подобна диагональной матрице ее собственных значений.*

Чтобы убедиться в этом, заметим, что в доказательстве теоремы правые собственные векторы могут быть все выбраны вещественными. Из определения получаем, что вещественная унитарная матрица ортогональна.

Упр. 1. Если A и B эрмитовы, то обязательно ли эрмитова AB? Если A и B унитарны, то обязательно ли унитарна AB?

Упр. 2. Найти вещественную ортогональную матрицу X и диагональную матрицу U, для которых $A = XUX'$, если

$$A = \begin{Vmatrix} 1 & 0 & -4 \\ 0 & 5 & 4 \\ -4 & 4 & 3 \end{Vmatrix}.$$

Упр. 3. Пусть A — эрмитова матрица с правым собственным вектором $\boldsymbol{x}$, соответствующим собственному значению μ. Доказать, что $\bar{\boldsymbol{x}}$ — левый собственный вектор с собственным значением μ.

Упр. 4. Доказать, что если A — эрмитова матрица, то $\det A$ — вещественное число.

Упр. 5. Доказать, что если $\operatorname{tr}(A^*A) = 0$, то $A = 0$.

Упр. 6. Доказать, что если U унитарна и эрмитова, то $U^2 = I$.

Упр. 7. Доказать, что если μ — собственное значение унитарной матрицы, то $|\mu| = 1$.

***Упр. 8.** Доказать, что любую матрицу $A \in \mathscr{C}_{n \times n}$ можно однозначно представить в виде $A = H + S$, где H эрмитова и S косоэрмитова.

***Упр. 9.** Пусть $A = A^*$ и $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle = 0$ для каждого $\boldsymbol{x} \in \mathscr{C}_n$. Доказать, что $A = 0$.

2.10. Унитарно подобные преобразования

Рассмотрим некоторую матрицу с комплексными собственными значениями, которая вместе с тем унитарно подобна диагональной матрице из ее собственных значений. Очевидно, что такая матрица не эрмитова. Возникает вопрос: можно ли характеризовать множество всех матриц в $\mathscr{C}_{n \times n}$, которые унитарно подобны диагональным матрицам из их собственных значений? Как мы увидим в теореме 2.10.2, это может быть сделано в весьма изящном виде. Доказательство этой теоремы будет облегчено, однако, если мы сначала получим следующий результат относительно приведения произвольной матрицы из $\mathscr{C}_{n \times n}$ с помощью унитарно подобных преобразований. Эта теорема принадлежит Шуру и Тёплицу (1910) и имеет важное значение сама по себе.

Теорема 2.10.1. *Любая матрица $A \in \mathscr{C}_{n \times n}$ унитарно подобна верхней треугольной матрице.*

Заметим прежде всего, что если A подобна треугольной матрице T, то элементы главной диагонали в T будут собственными значениями для A. Ибо если c — характеристический многочлен матрицы A и $A = S^{-1}TS$, то

$$c(\mu) = \det(\mu I - A) = \det\left(\mu I - S^{-1}TS\right) = \det(\mu I - T) = \prod_{j=1}^{n} (\mu - t_{jj}).$$

Так как одновременно $c(\mu) = \prod\limits_{j=1}^{n} (\mu - \mu_j)$, то требуемое утверждение доказано.

Доказательство. (1) Теорема доказывается индукцией по n. Предположим сначала, что $n = 2$ и A имеет собственное значение μ_1 с соответствующим ему правым собственным вектором $\boldsymbol{x}_1$, для которого $\boldsymbol{x}_1^*\boldsymbol{x}_1 = 1$. Мы строим 2×2-матрицу U, первый столбец которой совпадает с $\boldsymbol{x}_1$ и второй столбец выбирается так, чтобы U была унитарна. Тогда

$$U^*AU = \begin{Vmatrix} \boldsymbol{x}_1^* \\ \boldsymbol{x}_2^* \end{Vmatrix} A \,\| \boldsymbol{x}_1 \boldsymbol{x}_2 \| = \begin{Vmatrix} \boldsymbol{x}_1^*A\boldsymbol{x}_1 & \boldsymbol{x}_1^*A\boldsymbol{x}_2 \\ \boldsymbol{x}_2^*A\boldsymbol{x}_1 & \boldsymbol{x}_2^*A\boldsymbol{x}_2 \end{Vmatrix}.$$

Так как $A\boldsymbol{x}_1 = \mu_1\boldsymbol{x}_1$ и $\boldsymbol{x}_2^*\boldsymbol{x}_1 = 0$, то $\boldsymbol{x}_2^*A\boldsymbol{x}_1 = 0$. Следовательно, U^*AU — верхняя треугольная матрица, так что $A = UTU^*$ и теорема верна в случае $n = 2$.

(2) Мы покажем теперь, что если теорема верна для некоторого $n \geqslant 2$, то она с необходимостью верна для $n+1$. Пусть опять $\boldsymbol{x}_1$ — правый собственный вектор матрицы A с собственным значением μ_1, A порядка $n+1$ и $\boldsymbol{x}_1^*\boldsymbol{x}_1 = 1$. Построим множество из n векторов $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ порядка $n+1$, для которого $\boldsymbol{x}_1, \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ ортонормированы, т. е. матрица $U = \| \boldsymbol{x}_1\, \boldsymbol{a}_1\, \boldsymbol{a}_2 \ldots \ldots \boldsymbol{a}_n \|$ унитарна. (Заметим, что для такого построения всегда может быть использован процесс Грама — Шмидта.) Будем иметь тогда

$$U^*AU = \begin{Vmatrix} \boldsymbol{x}_1^*A\boldsymbol{x}_1 & \boldsymbol{x}_1^*A\boldsymbol{a}_1 & \ldots & \boldsymbol{x}_1^*A\boldsymbol{a}_n \\ \boldsymbol{a}_1^*A\boldsymbol{x}_1 & & & \\ \vdots & & B & \\ \boldsymbol{a}_n^*A\boldsymbol{x}_1 & & & \end{Vmatrix},$$

где B — $n \times n$-матрица. Но $A\boldsymbol{x}_1 = \mu_1\boldsymbol{x}_1$ и $\boldsymbol{a}_j^*\boldsymbol{x}_1 = 0$ для $j = 1, 2, \ldots \ldots, n$ означают, что $\boldsymbol{a}_j^*A\boldsymbol{x}_1 = 0$, $j = 1, 2, \ldots, n$, так что при некотором векторе $\boldsymbol{b} \in \mathscr{C}_n$

$$U^*AU = \begin{Vmatrix} \mu_1 & \boldsymbol{b}' \\ \boldsymbol{0} & B \end{Vmatrix}.$$

По предположению индукции существует унитарная матрица V порядка n, для которой V^*BV — верхняя треугольная. Определим унитарную матрицу U_1 порядка $n+1$ как

$$U_1 = \begin{Vmatrix} 1 & 0' \\ 0 & V \end{Vmatrix}$$

и заметим, что произведение UU_1 тоже унитарно. Как легко видеть,

$$(UU_1)^*A(UU_1) = \begin{Vmatrix} \mu_1 & \boldsymbol{b}'V \\ \boldsymbol{0} & V^*BV \end{Vmatrix},$$

а так как V^*BV — верхняя треугольная, то унитарная матрица UU_1 приводит A к верхнему треугольному виду. ◀

Эта теорема приводит ко второму, изящному доказательству теоремы 2.9.5. Ибо если $U^*AU = T$ — треугольная матрица и $A = A^*$, то $T = T^*$. Но единственными матрицами, для которых это верно, являются диагональные матрицы, откуда и следует результат.

Мы введем теперь еще один важный тип матриц. Матрица $A \in \mathscr{C}_{n \times n}$ *нормальна*, если $AA^* = A^*A$, т. е. если A коммутирует со своей сопряженной транспонированной. Без труда убеждаемся в том, что все унитарные, эрмитовы, косоэрмитовы и диагональные матрицы будут также нормальными. Значение нормальных матриц выяснено в следующем результате, который отвечает на поставленный в начале этого параграфа вопрос.

Теорема 2.10.2. *Некоторая матрица из $\mathscr{C}_{n\times n}$ унитарно подобна диагональной матрице из своих собственных значений в том и только в том случае, когда она нормальна.*

Доказательство. Предположим сначала, что существуют унитарная матрица U и диагональная матрица D, для которых $A = UDU^*$. Тогда

$$AA^* = (UDU^*)(U\bar{D}U^*) = UD\bar{D}U^* = U\bar{D}DU^* = (U\bar{D}U^*)(UDU^*) = A^*A,$$

так что A обязательно нормальна.

Наоборот, предположим, что A нормальна, и применим предыдущую теорему. Итак, существуют унитарная матрица U и верхняя треугольная матрица T, для которых $A = UTU^*$. Без труда убеждаемся в том, что $AA^* = A^*A \Longleftrightarrow$ *) $TT^* = T^*T$. Использование элементов $(1, 1)$ последнего равенства дает

$$|t_{11}|^2 + |t_{12}|^2 + \ldots + |t_{1n}|^2 = |t_{11}|^2.$$

Следовательно, $t_{1j} = 0$ для $j = 2, 3, \ldots, n$. Элементы (2.2) равенства $TT^* = T^*T$ приводят к

$$|t_{22}|^2 + |t_{23}|^2 + \ldots + |t_{2n}|^2 = |t_{12}|^2 + |t_{22}|^2.$$

Так как уже доказано, что $t_{12} = 0$, отсюда следует $t_{2j} = 0$ для $j = 3, 4, \ldots, n$. Продолжая это рассуждение, находим, что $t_{jk} = 0$ всегда, когда $j \neq k$ и $1 \leqslant j, k \leqslant n$. Таким образом, T должна быть диагональной матрицей и $A = UTU^*$. ◀

Заметим теперь, что нормальные матрицы простые. В частности, мы доказали, что эрмитовы матрицы простые и тем самым дублировали результат теоремы 2.9.4. Следует заметить, что обобщена также теорема 2.9.2. В самом деле, мы можем теперь сказать, что некоторая матрица из $\mathscr{C}_{n\times n}$ имеет множество из n ортонормированных правых собственных векторов $\Longleftrightarrow$ эта матрица нормальная.

Упр. 1. Показать, что если A нормальная, то такова же и $\mu I - A$.

Упр. 2. Доказать, что $\boldsymbol{x}$ — правый собственный вектор нормальной матрицы $A \Longleftrightarrow \boldsymbol{x}$ — правый собственный вектор для A^*.

Упр. 3. Построить пример:

(i) нормальной матрицы, которая не унитарна, не эрмитова, не косоэрмитова и не диагональна;

(ii) простой матрицы, которая не нормальна.

---

*) Знак $\Longleftrightarrow$ используется как сокращение для выражения «в том и только в том случае, когда». — *Прим. ред.*

Упр. 4. Доказать, что нормальные матрицы унитарно подобны $\Longleftrightarrow$ они имеют одинаковые собственные значения (с соответствующими кратностями).

2.11. Идемпотентные матрицы и проекции

Прежде чем детально исследовать строение идемпотентных матриц, полезно ввести некоторые геометрические понятия в изучение линейных пространств. Мы будем говорить, что линейное пространство $\mathscr{L}$ представляется в виде *прямой суммы* подпространств $\mathscr{L}_1$ и $\mathscr{L}_2$, и писать $\mathscr{L} = \mathscr{L}_1 \oplus \mathscr{L}_2$, если удовлетворяются следующие условия:

(i) $\mathscr{L}_1 \subseteq \mathscr{L}$ и $\mathscr{L}_2 \subseteq \mathscr{L}$;

(ii) для каждого $\boldsymbol{x} \in \mathscr{L}$ существуют такие $\boldsymbol{x}_1 \in \mathscr{L}_1$ и $\boldsymbol{x}_2 \in \mathscr{L}_2$, что $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$;

(iii) если $\boldsymbol{x} \in \mathscr{L}_1$ и $\boldsymbol{x} \in \mathscr{L}_2$, то $\boldsymbol{x} = \mathbf{0}$.

Следует заметить, что в условии (iii) утверждается, что единственным вектором из $\mathscr{L}$, принадлежащим одновременно как $\mathscr{L}_1$, так и $\mathscr{L}_2$, будет нулевой вектор, и, в свою очередь, это означает, что описанное в (ii) разложение единственно. Ибо если $\boldsymbol{x} \in \mathscr{L}$ и

$$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2 = \boldsymbol{y}_1 + \boldsymbol{y}_2,$$

где $\boldsymbol{x}_1$, $\boldsymbol{y}_1 \in \mathscr{L}_1$ и $\boldsymbol{x}_2$, $\boldsymbol{y}_2 \in \mathscr{L}_2$, то $\boldsymbol{x}_1 - \boldsymbol{y}_1 = \boldsymbol{y}_2 - \boldsymbol{x}_2$. Но $\boldsymbol{x}_1 - \boldsymbol{y}_1 \in \mathscr{L}_1$ и $\boldsymbol{y}_2 - \boldsymbol{x}_2 \in \mathscr{L}_2$, так что (iii) означает, что $\boldsymbol{y}_1 = \boldsymbol{x}_1$, $\boldsymbol{y}_2 = \boldsymbol{x}_2$.

При $\mathscr{L} = \mathscr{L}_1 \oplus \mathscr{L}_2$ мы будем говорить про $\mathscr{L}_1$ и $\mathscr{L}_2$, что они *дополнительные*, или что они являются *дополнениями* друг к другу. Если, в частности, мы имеем также, что $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$ для каждого $\boldsymbol{x} \in \mathscr{L}_1$ и каждого $\boldsymbol{y} \in \mathscr{L}_2$, то мы говорим, что $\mathscr{L}_1$ и $\mathscr{L}_2$ — *ортогональные дополнения* друг к другу.

Упр. 1. Доказать, что если $\mathscr{L} = \mathscr{L}_1 \oplus \mathscr{L}_2$, то размерность $\mathscr{L}$ равна сумме размерностей $\mathscr{L}_1$ и $\mathscr{L}_2$.

Теорема 2.11.1. *Если P идемпотентна, то*

(а) *$I - P$ идемпотентна*;

б) $\mathscr{R}(I - P) = \mathscr{N}(P)$;

(в) $\mathscr{N}(I - P) = \mathscr{R}(P)$.

Доказательство. Для части (а) заметим, что $(I - P)^2 = I - 2P + P^2$ и $P^2 = P$ означают, что $(I - P)^2 = I - P$.

Если $\boldsymbol{x} \in \mathscr{R}(I - P)$, то существует такой $\boldsymbol{y} \in \mathscr{C}_n$, что $\boldsymbol{x} = (I - P)\boldsymbol{y}$. Следовательно, $P\boldsymbol{x} = P(I - P)\boldsymbol{y} = (P - P^2)\boldsymbol{y} = \mathbf{0}$. Поэтому $\boldsymbol{x} \in \mathscr{N}(P)$, так что $\mathscr{R}(I - P) \subseteq \mathscr{N}(P)$.

Наоборот, если $\boldsymbol{x} \in \mathscr{N}(P)$, то $P\boldsymbol{x} = \mathbf{0}$, так что $(I - P)\boldsymbol{x} = \boldsymbol{x}$. Следовательно, $\boldsymbol{x} \in \mathscr{R}(I - P)$, так что $\mathscr{N}(P) \subseteq \mathscr{R}(I - P)$. Поэтому часть (б) доказана. Подобное же рассуждение приводит к части (в). ◀

Лемма. *Если P идемпотентна, то $\mathscr{C}_n = \mathscr{N}(P) \oplus \mathscr{R}(P)$.*

Доказательство. Аксиома (i) для прямой суммы, очевидно, удовлетворяется. Для любого $x \in \mathscr{C}_n$ можно записать $x = x_1 + x_2$, где $x_1 = (I - P)x$ и $x_2 = Px$. Таким образом, $x_1 \in \mathscr{R}(I - P) = \mathscr{N}(P)$ по части (б) предыдущей теоремы и $x_2 \in \mathscr{R}(P)$. Поэтому выполняется аксиома (ii).

Наконец, если $x \in \mathscr{R}(P)$ и $x \in \mathscr{N}(P)$, то по части (в) последней теоремы имеем $(I - P)x = 0$ и $Px = 0$. Следовательно, $x = 0$, так что все аксиомы для прямой суммы выполняются. ◀

Теорема 2.11.2. *Идемпотентная матрица проста.*

Доказательство. Как мы видели, собственные значения идемпотентной матрицы все равны либо 1, либо 0 (упр. 4 из § 2.5). Отсюда следует, что правые собственные векторы матрицы P все принадлежат либо $\mathscr{N}(P)$, либо $\mathscr{N}(I - P)$. Так как $\mathscr{N}(I - P) = \mathscr{R}(P)$, результат леммы означает, что $\mathscr{N}(P) \oplus \oplus \mathscr{N}(I - P) = \mathscr{C}_n$ и, следовательно, правые собственные векторы матрицы R должны порождать $\mathscr{C}_n$. Поэтому P проста. ◀

Следствие. *P — идемпотентная матрица ранга $r \Longleftrightarrow$ существуют такие квазибиортогональные системы $y_1, y_2, \ldots, y_r$ и $x_1, x_2, \ldots, x_r$, что*

$$P = \sum_{j=1}^{r} x_j y_j'.$$

Доказательство. Если P имеет этот вид, то ранг по столбцам P равен, очевидно, r, и, используя свойство квазибиортогональности, получаем

$$P^2 = \left(\sum_{j=1}^{r} x_j y_j'\right)\left(\sum_{k=1}^{r} x_k y_k'\right) = \sum_{j=1}^{r} x_j \sum_{k=1}^{r} (y_j' x_k) y_k' = \sum_{j=1}^{r} x_j y_j' = P.$$

Наоборот, так как идемпотентная матрица проста, то применение теоремы 2.5.1 с $f(P) = P$ показывает, что P должна иметь написанный выше вид. Эта теорема также показывает, что элементы x и y — это соответственно правые и левые собственные векторы с собственным значением 1. ◀

Важным является случай, когда матрица P идемпотентна и эрмитова. В этом случае можно выбрать *ортонормированное* множество правых собственных векторов для P и соответствующие левые собственные векторы будут просто комплексно сопряженными к правым собственным векторам. Таким образом, получаем

$$P = \sum_{j=1}^{r} x_j x_j^*$$

и для любого $x \in \mathscr{C}_n$

$$Px = \sum_{j=1}^{r} x_j x_j^* x = \sum_{j=1}^{r} \langle x_j, x \rangle x_j,$$

причем каждый член в этой сумме является проекцией вектора $\boldsymbol{x}$ на правый собственный вектор. Мы поэтому называем матрицу P ***проекцией***, если P идемпотентна и эрмитова, и будем говорить, что P — проекция на пространство, порожденное $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$, т. е. на $\mathscr{R}(P) = \mathscr{N}(I - P)$.

Будет поучительно изложить следующий более прямой подход к проекциям. Предположим, что задано произвольное подпространство $\mathscr{L}$ с базисом $\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_k$ и мы хотим найти проекцию на $\mathscr{L}$. Определим $n \times k$-матрицу $F = \|\boldsymbol{f}_1\boldsymbol{f}_2 \ldots \boldsymbol{f}_k\|$; тогда все элементы в $\mathscr{L}$ записываются в виде $F\boldsymbol{x}$ при некотором $\boldsymbol{x} \in \mathscr{C}_k$. Для данного $\boldsymbol{a} \in \mathscr{C}_n$ нам следует найти такой вектор $F\boldsymbol{x} \in \mathscr{L}$, который с наибольшим правом может быть назван проекцией вектора $\boldsymbol{a}$ на $\mathscr{L}$.

Мы определяем $F\boldsymbol{x}$ требованием, чтобы $\boldsymbol{a} - F\boldsymbol{x}$ был ортогонален к каждому элементу из $\mathscr{L}$. Это будет так, если $\boldsymbol{a} - F\boldsymbol{x}$ ортогонален к каждому базисному вектору. Таким образом, должно быть

$$\langle \boldsymbol{f}_j, \boldsymbol{a} - F\boldsymbol{x} \rangle = \boldsymbol{f}_j^* (\boldsymbol{a} - F\boldsymbol{x}) = 0, \qquad j = 1, 2, \ldots, k,$$

что эквивалентно матричному уравнению

$$F^* (\boldsymbol{a} - F\boldsymbol{x}) = \boldsymbol{0}.$$

Решая это уравнение относительно $\boldsymbol{x}$, получаем $\boldsymbol{x} = (F^*F)^{-1}F^*\boldsymbol{a}$, заметив, что обратная матрица существует, так как F^*F — грамиан линейно независимых векторов (теорема 2.7.2). Если P — матрица этого проектирования, то мы хотим, чтобы было

$$P\boldsymbol{a} = F\boldsymbol{x} = F(FF^*)^{-1} F^*\boldsymbol{a},$$

т. е.

$$P = F(FF^*)^{-1} F^*. \tag{2.11.1}$$

Таким образом, мы определили проектирующую матрицу в терминах произвольного базиса для $\mathscr{L}$. Читатель без труда проверит, что P идемпотентна и эрмитова.

Однако, что особенно интересно, так это то, что P *не зависит* от выбора базисных векторов для $\mathscr{L}$. Чтобы убедиться в этом, заметим, что если X — матрица другого множества базисных векторов, то существует такая неособая $k \times k$-матрица M, что $F = XM$, откуда

$$\begin{aligned} P &= XM(M^*X^*XM)^{-1} M^*X^* = \\ &= XM\left(M^{-1}(X^*X)^{-1}(M^*)^{-1}\right) M^*X^* = \\ &= X(X^*X)^{-1} X^*. \end{aligned}$$

Следовательно, мы получаем ту же самую проекцию P, каков бы ни был выбор базисных векторов для $\mathscr{L}$. В частности, если выбрать ортонормированный базис для $\mathscr{L}$, то $X^*X = I$ и

$$P = XX^* = \sum_{j=1}^{k} x_j x_j^*,$$

и мы возвращаемся к виду для матрицы P, который мы имели ранее.

Упр. 2. Показать, что сопутствующие матрицы для эрмитовой матрицы являются проекциями.

Упр. 3. (i) Показать, что если $\mathscr{L} = \mathscr{L}_1 \oplus \mathscr{L}_2$ и $\mathscr{L} \oplus \mathscr{L}_3$ определены, то можно записать $\mathscr{L} \oplus \mathscr{L}_3 = \mathscr{L}_1 \oplus \mathscr{L}_2 \oplus \mathscr{L}_3$, не опасаясь неопределенности.

(ii) Если $\mathscr{L}_1, \mathscr{L}_2, \ldots, \mathscr{L}_k$ — правые собственные пространства, соответствующие различным собственным значениям простой $n \times n$-матрицы, то показать, что $\mathscr{C}_n = \mathscr{L}_1 \oplus \mathscr{L}_2 \oplus \ldots \oplus \mathscr{L}_k$. Показать также, что если матрица эрмитова, то $\mathscr{L}_j$ и $\mathscr{L}_k$ $(j \neq k)$ — ортогональные дополнения.

***Упр. 4.** Пусть $\mathscr{L}$ — линейное пространство размерности n и подпространства $\mathscr{L}_1, \mathscr{L}_2$ в $\mathscr{L}$ имеют размерности n_1, n_2 соответственно, где $n_1 + n_2 > n$. Показать, что существует ненулевой элемент в $\mathscr{L}$, принадлежащий как $\mathscr{L}_1$, так и $\mathscr{L}_2$.

Упр. 5. Если P — проекция, то показать, что в лемме к теореме 2.11.2 $\mathscr{N}(P)$ и $\mathscr{R}(P)$ — ортогональные дополнения.

2.12. Эрмитовы и квадратичные формы

Предположим, что A — эрмитова $n \times n$-матрица, и рассмотрим функцию $h(x)$, определенную для всех $x \in \mathscr{C}_n$ по формуле

$$h(x) = x^*Ax = \sum_{i,\, j=1}^{n} a_{ij}\bar{x}_i x_j.$$

На первый взгляд h появляется как функция со значениями в области комплексных чисел, но мы утверждаем, что h на самом деле — вещественнозначная функция. Чтобы убедиться в этом, заметим, что $h^* = \bar{h}$, так что, так как A эрмитова,

$$\bar{h} = (x^*Ax)^* = x^*A^*x = x^*Ax = h.$$

Таким образом, для любого x имеем $h = \bar{h}$, а поэтому h вещественна. Введенная функция h называется *эрмитовой формой* от $x_1, x_2, \ldots, x_n$ и A называется матрицей эрмитовой формы (см. ниже упр. 1).

Очень важным является частный случай, когда A — вещественная и симметрическая матрица. В этом случае мы предполагаем x принадлежащим $\mathscr{R}_n$ и называем

$$q(x) = x'Ax = \sum_{i,\, j=1}^{n} a_{ij}x_i x_j$$

квадратичной формой от $x_1, x_2, \ldots, x_n$. Таким образом, квадратичная форма — вещественнозначная функция, определенная на $\mathscr{R}_n$.

Предположим, что мы рассматриваем $x_1, x_2, \ldots, x_n$ как координаты в n-мерном пространстве, и путь T — некоторая неособая матрица порядка n. Можно сказать, что уравнение $\boldsymbol{x} = T\boldsymbol{y}$, или $\boldsymbol{y} = T^{-1}\boldsymbol{x}$, определяет новую систему координат $y_1, y_2, \ldots, y_n$ в том же пространстве. Имеем тогда

$$h(\boldsymbol{x}) = h(T\boldsymbol{y}) = \boldsymbol{y}^*(T^*AT)$$

и, так как T^*AT эрмитова, $h(T\boldsymbol{y})$ будет эрмитовой формой от $y_1, y_2, \ldots, y_n$. Матрица новой формы T^*AT называется *конгруэнтным преобразованием* матрицы A. Нашей ближайшей целью будет исследование таких систем координат, для которых $h(T\boldsymbol{y})$ имеет вид $\sum_{i=1}^{n} c_i |y_i|^2$. Другими словами, мы должны исследовать те конгруэнтные преобразования T^*AT матрицы A (если они вообще существуют), для которых T^*AT приводится к диагональной матрице.

В случае квадратичных форм мы разыскиваем вещественные конгруэнтные преобразования $T'AT$ матрицы A, для которых $T'AT$ — диагональная матрица.

В нашем распоряжении уже имеется способ проведения необходимой редукции, хотя он не предлагается как практически полезный. Он содержится в следующей теореме.

Теорема 2.12.1. *Если h — эрмитова форма, то существуют такие вещественная матрица $U = \operatorname{diag}\{\mu_1, \ldots, \mu_n\}$ и унитарная матрица X, что, записывая $\boldsymbol{y} = X^*\boldsymbol{x}$, будем иметь*

$$h(\boldsymbol{x}) = h(X\boldsymbol{y}) = \sum_{j=1}^{n} \mu_j |y_j|^2.$$

Доказательство. Пусть A — матрица эрмитовой формы h. Пусть X — унитарная матрица правых собственных векторов матрицы A, как она построена в теореме 2.9.5, и $\mu_1, \mu_2, \ldots, \mu_n$ — собственные значения матрицы, которые, как мы знаем, вещественны. Если тогда $\boldsymbol{x} = X\boldsymbol{y}$, то мы имеем $X^*AX = U$ и

$$h(\boldsymbol{x}) = \boldsymbol{x}^*A\boldsymbol{x} = \boldsymbol{y}^*X^*AX\boldsymbol{y} = \boldsymbol{y}^*U\boldsymbol{y} = \sum_{j=1}^{n} \mu_j |y_j|^2. \quad ◀$$

Таким образом, эрмитову форму можно привести к сумме квадратов при помощи унитарного конгруэнтного преобразования и, соответственно, квадратичную форму можно всегда привести к сумме квадратов при помощи вещественного ортогонального конгруэнтного преобразования. Однако эти преобразования довольно специальные, и мы сейчас покажем, что существует намного более широкий класс конгруэнтных преобразований, который может быть использован в приведении эрмитовой (или квадратичной) формы к сумме квадратов.

Прежде чем приступить к исследованию этого вопроса, полезно ввести еще одно множество матриц. Напомним, что единичный вектор e_j в $\mathscr{C}_n$ — это вектор с 1 на j-м месте и 0 на всех других местах. *Перестановочная матрица* — это $n \times n$-матрица вида

$$P = \| e_{i_1} e_{i_2} \ \ldots \ e_{i_n} \|,$$

где $i_1, i_2 \ldots i_n$ — перестановка чисел $1, 2, \ldots, n$. Если $A \in \mathscr{C}_{n\times n}$, то отметим результат умножения справа матрицы A на P. Для j-го столбца произведения имеем

$$(AP)_{*j} = AP_{*j} = Ae_{i_j} = A_{*i_j}.$$

Таким образом,

$$AP = \| A_{*i_1} A_{*i_2} \ \ldots \ A_{*i_n} \|$$

и умножение матрицы A справа на P просто переставляет столбцы в A. Аналогичный результат имеет место и для строк матрицы A при умножении A слева на P', ибо

$$P'A = \left\| \begin{matrix} e'_{i_1} \\ \cdot \\ \cdot \\ \cdot \\ e'_{i_n} \end{matrix} \right\| A = \left\| \begin{matrix} A_{i_1 *} \\ \cdot \\ \cdot \\ \cdot \\ A_{i_n *} \end{matrix} \right\|.$$

Упр. 1. Доказать, что (эрмитова) матрица эрмитовой формы единственна. (Указание. Использовать упр. 9 из § 2.9.)

***Упр. 2.** Показать, что если P — перестановочная матрица и $D = \operatorname{diag}\{d_1, d_2, \ldots, d_n\}$, то $P'P = I$ и

$$P'DP = \operatorname{diag}\{d_{i_1}, d_{i_2}, \ldots, d_{i_n}\}.$$

Мы покажем теперь, как построить некоторое семейство матриц, каждая из которых может быть использована в приведении заданной эрмитовой формы. Если A — матрица этой формы, то предположим, что собственные значения A расположены в порядке их невозрастания и что различными собственными значениями будут $\mu_1, \mu_2, \ldots, \mu_s$ с кратностями $\alpha_1, \alpha_2, \ldots, \alpha_s$ соответственно. Мы можем записать диагональную матрицу из собственных значений матрицы A в виде

$$U = \left\| \begin{matrix} \mu_1 I_1 & & 0 \\ & \ddots & \\ 0 & & \mu_s I_s \end{matrix} \right\|.$$

Пусть теперь X — унитарная матрица правых собственных векторов, расположение которых соответствует зафиксированному выше порядку собственных векторов матрицы A. Наше семейство преобразующих матриц будет построено при помощи рассмотрения преобразований матрицы X, которые сохраняют диагональный вид конгруэнтного преобразования X^*AX. Рассмотрим следующий план.

Пусть $V_1, V_2, \ldots, V_s$ — произвольные унитарные матрицы порядков $\alpha_1, \alpha_2, \ldots, \alpha_s$ соответственно; $D_1, D_2, \ldots, D_s$ — произвольные неособые диагональные матрицы тех же порядков и P — произвольная перестановочная матрица. Если определить

$$Y = X \begin{Vmatrix} V_1D_1 & & & 0 \\ & V_2D_2 & & \\ & & \ddots & \\ 0 & & & V_sD_s \end{Vmatrix} P,$$

то конгруэнтное преобразование Y^*AY приводит A к диагональному виду.

Для доказательства этого заметим, что

$$Y^*AY = P' \begin{Vmatrix} \bar{D}_1V_1^* & & 0 \\ & \ddots & \\ 0 & & \bar{D}_sV_s^* \end{Vmatrix} X^*AX \begin{Vmatrix} V_1D_1 & & 0 \\ & \ddots & \\ 0 & & V_sD_s \end{Vmatrix} P =$$

$$= P' \begin{Vmatrix} \bar{D}_1V_1^* & & 0 \\ & \ddots & \\ 0 & & \bar{D}_sV_s^* \end{Vmatrix} \begin{Vmatrix} \mu_1 I_{\alpha_1} & & 0 \\ & \ddots & \\ 0 & & \mu_s I_{\alpha_s} \end{Vmatrix} \begin{Vmatrix} V_1D_1 & & 0 \\ & \ddots & \\ 0 & & V_sD_s \end{Vmatrix} P =$$

$$= P' \begin{Vmatrix} \mu_1\bar{D}_1D_1 & & 0 \\ & \ddots & \\ 0 & & \mu_s\bar{D}_sD_s \end{Vmatrix} P \text{ — диагональная матрица,}$$

причем на последнем шаге использовано упр. 2.

Таким образом, имеется широкий класс конгруэнтных преобразований, которые могут быть использованы при приведении эрмитовой формы. Возникает вопрос о том, существуют ли инварианты, общие множеству всех приведенных форм. Первый результат подсказывает вид только что найденных диагональных матриц Y^*AY. Мы формулируем результат для матрицы некоторой эрмитовой формы и надеемся, что значение этого результата для самих форм очевидно.

Теорема 2.12.2. *Пусть A — некоторая эрмитова матрица, имеющая p и только p положительных собственных значений и ранг r. Если тогда V — неособая матрица, для которой* $V^*AV = \operatorname{diag}\{d_1, \ldots, d_n\} = D$, *то*:

(i) *существует $n-r$ и только $n-r$ чисел d_j, равных нулю*;

(ii) *существует p и только p чисел d_j, которые положительны.*

Доказательство. 1. Так как V неособая и $V^*AV = D$, мы убеждаемся в том (по упр. 3 из § 1.15), что ранг матрицы D равен r, что дает результат (i).

2. Для части (ii) мы сначала строим перестановочную матрицу P, для которой $P'DP = D_1 = \operatorname{diag}\{\delta_1, \delta_2, \ldots, \delta_n\}$, где диагональные элементы будут числами $d_1, \ldots, d_n$, переставленными так, чтобы они шли в порядке невозрастания. Таким образом, $\delta_1 \geqslant \delta_2 \geqslant \ldots \geqslant \delta_n$. Если определить теперь $W = VP$, то $\det W \neq 0$ и $V^*AV = D$ означает, что

$$W^*AW = D_1. \tag{2.12.1}$$

Если предположить, что $\delta_1, \delta_2, \ldots, \delta_q > 0$ и $\delta_{q+1} \leqslant 0$, то мы должны доказать, что $q = p$.

Нам известно также, что если $\mu_1 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_n$ — собственные значения матрицы A и $U = \operatorname{diag}\{\mu_1, \ldots, \mu_n\}$, то существует унитарная матрица X, для которой

$$X^*AX = U \tag{2.12.2}$$

и, кроме того, $\mu_1, \mu_2, \ldots, \mu_p > 0$ и $\mu_{p+1} \leqslant 0$.

3. Пусть $\boldsymbol{x}_j = X_{*j}$ и $\boldsymbol{w}_j = W_{*j}$, $j = 1, 2, \ldots, n$. Если $\mathscr{L}$ — пространство, порожденное $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p$, то его размерность $d(\mathscr{L})$ равна p. Если $\boldsymbol{x} \in \mathscr{L}$ и $\boldsymbol{x} \neq \boldsymbol{0}$, то можно записать $\boldsymbol{x} = \sum_{j=1}^{p} \alpha_j \boldsymbol{x}_j$ и

$$\boldsymbol{x}^*A\boldsymbol{x} = \left(\sum_{j=1}^{p} \bar{\alpha}_j \boldsymbol{x}_j^*\right) A \left(\sum_{k=1}^{p} \alpha_k \boldsymbol{x}_k\right) = \sum_{j=1}^{p} |\alpha_j|^2 \mu_j > 0,$$

так как равенство (2.12.2) означает, что $\boldsymbol{x}_j^*A\boldsymbol{x}_k = \mu_k \boldsymbol{x}_j^* \boldsymbol{x}_k = \mu_k \delta_{jk}$.

Пусть теперь $\mathscr{R}$ — пространство, порожденное $\boldsymbol{w}_{q+1}, \ldots, \boldsymbol{w}_n$. Имеем $d(\mathscr{R}) = n - q$. Если $\boldsymbol{y} \neq \boldsymbol{0}$ и $\boldsymbol{y} \in R$, то, используя равенство (2.12.1), получаем подобно предыдущему, что

$$\boldsymbol{y}^*A\boldsymbol{y} \leqslant 0.$$

Но $d(\mathscr{R}) + d(\mathscr{L}) = n + (p - q)$, так что при $p > q$ должен существовать вектор $\boldsymbol{z} \neq \boldsymbol{0}$, который принадлежит как $\mathscr{R}$, так и $\mathscr{L}$ (упр. 4 из § 2.11). Для такого вектора мы имели бы $\boldsymbol{z}^*A\boldsymbol{z} > 0$ и $\boldsymbol{z}^*A\boldsymbol{z} \leqslant 0$, что невозможно. Следовательно, $p \leqslant q$.

4. Если $\mathscr{L}_1$, $\mathscr{R}_1$ — пространства, порожденные $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_q$ и $\boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_n$ соответственно, то можно показать подобно предыдущему, что $p \geqslant q$. Следовательно, $q = p$. ◀

Легко видеть, что то же самое рассуждение и те же заключения выполняются в случае, когда A вещественная и симметрическая и V вещественная и неособая. Таким образом, числа r и p будут инвариантами эрмитовой (квадратичной) формы при неособых (вещественных неособых) преобразованиях координат. Этот факт известен как *закон инерции Сильвестра* и инварианты r, p и $2p - r$ (разность между числом положительных и числом отрицательных членов в приведенной форме) известны как ранг, индекс и сигнатура формы соответственно.

2.13. Метод приведения Лагранжа

В последнем параграфе было доказано существование некоторых инвариантов, соответствующих данной эрмитовой или квадратичной форме, но в нашем распоряжении нет в общем случае полезного способа для нахождения этих инвариантов или приведенных форм, кроме как вычислять все собственные значения и собственные векторы матрицы формы. Мы опишем теперь принадлежащий Лагранжу метод для вычисления приведенной формы.

Начиная с этого места, мы для простоты ограничим наше рассмотрение квадратичными формами. Необходимые для распространения результатов на эрмитовы формы изменения предоставляется провести в качестве упражнения читателю.

Заметим прежде всего, что так как A симметрическая, то можем записать

$$q(\boldsymbol{x}) = \sum_{i,\,j=1}^{n} a_{ij}x_i x_j =$$
$$= a_{11}x_1^2 + \ldots + a_{nn}x_n^2 + 2a_{12}x_1x_2 + \ldots + 2a_{n-1,\,n}x_{n-1}x_n.$$

Случай 1. Предположим, что не все из коэффициентов $a_{11}, a_{22}, \ldots, a_{nn}$ равны нулю. Тогда, производя перенумерацию переменных и коэффициентов, если это необходимо, можем считать, что $a_{11} \neq 0$, и записать

$$q(\boldsymbol{x}) = a_{11}x_1^2 + 2a_{12}x_1x_2 + \ldots + 2a_{1n}x_1x_n + \sum_{i,\,j=2}^{n} a_{ij}x_ix_j =$$
$$= a_{11}\left(x_1^2 + 2\frac{a_{12}}{a_{11}}x_1x_2 + \ldots + 2\frac{a_{1n}}{a_{11}}x_1x_n\right) + q_1(\boldsymbol{\xi}),$$

где $\xi' = \| x_2 x_3 \ldots x_n \|$ и q_1 — квадратичная форма. Продолжая «дополнение до квадрата», получаем

$$q(\boldsymbol{x}) = a_{11} \left\{ \left(x_1 + \frac{a_{12}}{a_{11}} x_2 + \ldots + \frac{a_{1n}}{a_{11}} x_n \right)^2 - \left(\frac{a_{12}}{a_{11}} \right)^2 x_2^2 - \ldots \right.$$
$$\left. \ldots - \left(\frac{a_{1n}}{a_{11}} \right)^2 x_n^2 - 2 \frac{a_{12} a_{13}}{a_{11}^2} x_2 x_3 - \ldots \right\} + q_1(\xi) =$$
$$= a_{11} \left(x_1 + \frac{a_{12}}{a_{11}} x_2 + \ldots + \frac{a_{1n}}{a_{11}} x_n \right)^2 + q_2(\xi),$$

где q_2 — другая квадратичная форма от $x_2, \ldots, x_n$. Таким образом, неособое преобразование координат, определенное как

$$y_1 = x_1 + \frac{a_{12}}{a_{11}} x_2 + \ldots + \frac{a_{1n}}{a_{11}} x_n, \quad y_2 = x_2, \ \ldots, \ y_n = x_n,$$

приводит q к виду

$$d_1 y_1^2 + q_2(y_2, \ldots, y_n) \tag{2.13.1}$$

(где $d_1 = a_{11}$).

Случай 2. Если $a_{11} = \ldots = a_{nn} = 0$, то рассмотрим коэффициент $a_{ij} \neq 0$ с $i \neq j$. Опять, без потери общности, можем предположить, что a_{12} не равен нулю. В этом случае произведем конгруэнтное преобразование матрицы A, задаваемое преобразованием координат

$$\xi_1 = x_1, \quad \xi_2 = x_2 - x_1, \quad \xi_3 = x_3, \ \ldots, \ \xi_n = x_n.$$

Легко видеть, что определяемая преобразованной матрицей форма содержит член ξ_1^2. Теперь можно применить способ, описанный в случае 1, для приведения q к виду (2.13.1).

Таким образом, мы доказали, что любую данную квадратичную форму q, определенную на $\mathcal{R}_n$, можно привести к виду (2.13.1). Применяя тот же самый прием к $q_2(y_2, \ldots, y_n)$, приводим ее к виду $d_2 z_2^2 + q_3(z_3, \ldots, z_n)$. Тогда $q = d_1 z_1^2 + d_2 z_2^2 + q_3(z_3, \ldots, z_n)$, где $z_1 = y_1$.

Мы можем, очевидно, закончить приведение квадратичной формы q самое большее за $n-1$ таких шагов. Конечный эффект последовательных неособых преобразований эквивалентен некоторому одному неособому преобразованию первоначальных координат $x_1, x_2, \ldots, x_n$.

Упр. 1. Найти ранг, индекс и сигнатуру квадратичной формы

$$q(\boldsymbol{x}) = 2x_1 x_2 + 2x_2 x_3 + x_3^2;$$

найти также преобразование координат, которое приводит q к сумме квадратов.

Решение. Производя сначала преобразование

(i) $$y_1 = x_3, \quad y_2 = x_2, \quad y_3 = x_1,$$

получаем

$$q = (y_1^2 + 2y_1y_2) + 2y_2y_3 = (y_1 + y_2)^2 - y_2^2 + 2y_2y_3.$$

Положив теперь

(ii) $$z_1 = y_1 + y_2, \quad z_2 = y_2, \quad z_3 = y_3,$$

будем иметь

$$q = z_1^2 - z_2^2 + 2z_2z_3 = z_1^2 - (z_2 - z_3)^2 + z_3 = w_1^2 - w_2^2 + w_3^2,$$

где

(iii) $$w_1 = z_1, \quad w_2 = z_2 - z_3, \quad w_3 = z_3.$$

Таким образом, ранг $r = 3$, индекс $p = 2$ и сигнатура $2p - r = 1$. Приводящее к сумме квадратов преобразование получается комбинированием (i), (ii) и (iii) в виде

$$\begin{aligned} x_1 &= y_3, & x_2 &= y_2, & x_3 &= y_1, \\ y_1 &= z_1 - z_2, & y_2 &= z_2, & y_3 &= z_3, \\ z_1 &= w_1, & z_2 &= w_2 + w_3, & z_3 &= w_3. \end{aligned}$$

Таким образом, $x_1 = w_3$, $x_2 = w_2 + w_3$ и $x_3 = w_1 - w_2 - w_3$, или

$$\boldsymbol{x} = \begin{Vmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & -1 & -1 \end{Vmatrix} \boldsymbol{w}.$$

Упр. 2. Найти ранг, индекс и сигнатуру следующих квадратичных форм и преобразования, приводящие их к сумме квадратов:

(а) $$2x_1^2 - x_3^2 + x_1x_2 + x_1x_3;$$

(б) $$x_1^2 + 2(x_1x_2 + x_2x_3 + x_3x_4 + x_4x_5 + x_5x_1);$$

(в) $$2x_1x_2 - x_1x_3 + x_1x_4 - x_2x_3 + x_2x_4 - 2x_3x_4.$$

***Упр. 3.** Рассмотрим квадратичную форму q, определенную на $\mathcal{R}_3$.

(i) Показать, что преобразованию координат $\boldsymbol{x} = T\boldsymbol{y}$, где T — вещественная ортогональная матрица, соответствует некоторый поворот осей координат в $\mathcal{R}_3$.

(ii) Множество точек в $\mathcal{R}_3$ (если оно существует), удовлетворяющих уравнению $q(x_1, x_2, x_3) = 1$, называется *центральной квадрикой*. Показать, что при подходящих новых координатных осях, получаемых из прежних поворотом, уравнение квадрики приводится к виду

$$\mu_1\xi_1^2+\mu_2\xi_2^2+\mu_3\xi_3^2=1.$$

Новые координатные оси называются *главными осями* квадрики. (Заметим, что вид квадрики определяется относительными величинами и знаками чисел μ_1, μ_2, μ_3.)

Упр. 4. Исследовать вид центральной квадрики, уравнение которой

$$3x_2^2+3x_3^2-2x_2x_3-4x_3x_1-4x_1x_2=1.$$

2.14. Определенные матрицы

Как и в предыдущем параграфе, мы ограничимся рассмотрением квадратичных форм, хотя вводимые определения и получаемые результаты опять без труда могут быть распространены на эрмитовы формы.

Нас интересуют здесь два семейства матриц, определяемых следующим образом. Вещественная симметрическая матрица A порядка n называется *положительно определенной*, если $\boldsymbol{x}'A\boldsymbol{x}>0$ для всех ненулевых векторов $\boldsymbol{x}\in\mathcal{R}_n$. Если $\boldsymbol{x}'A\boldsymbol{x}\geqslant 0$ для всех $\boldsymbol{x}\in\mathcal{R}_n$ и $\boldsymbol{x}'A\boldsymbol{x}=0$ для некоторого ненулевого $\boldsymbol{x}\in\mathcal{R}_n$, то A называется *неотрицательно определенной*. В обоих случаях мы можем также сказать, что квадратичная форма, задаваемая матрицей A, положительно определенная или неотрицательно определенная соответственно. Заметим, что под *определенной* матрицей понимается любая из матриц двух указанных выше семейств. Мы могли бы также дать определения, включающие случаи, когда $\boldsymbol{x}'A\boldsymbol{x}\leqslant 0$ для всех $\boldsymbol{x}\in\mathcal{R}_n$, но эту возможность нам нет необходимости рассматривать.

Следует также заметить, что положительно определенная матрица обязательно неособая, ибо если матрица особая, то существует вектор $\boldsymbol{x}\neq\boldsymbol{0}$, для которого $A\boldsymbol{x}=\boldsymbol{0}$ и, следовательно, $\langle\boldsymbol{x}, A\boldsymbol{x}\rangle=0$. Эта возможность исключена определением положительно определенной матрицы.

Теорема 2.14.1*). *Вещественная симметрическая $n\times n$-матрица A будет определенной ранга $r\leqslant n$ в том и только в том случае, когда A имеет r положительных собственных значений и $n-r$ собственных значений, равных нулю.*

---

*) В оригинале эта теорема сформулирована следующим образом: «$n\times n$-матрица A будет определенной ранга $r\leqslant n$ в том и только в том случае, когда A вещественная и симметрическая и имеет r положительных собственных значений и $n-r$ собственных значений, равных нулю». Здесь приводится ее исправленная формулировка. — *Прим. перев.*

Доказательство. Всюду в этой теореме и в следующей U будет обозначать вещественную диагональную матрицу из собственных значений матрицы A, и X будет обозначать вещественную ортогональную матрицу, для которой $A = XUX'$ (следствие теоремы 2.9.5). Так как ортогональная матрица обязательно неособая, то сразу же получаем, что ранг U равен r, рангу A. Отсюда следует, что $n-r$ и только $n-r$ диагональных элементов U равны нулю и, следовательно, что A имеет ровно $n-r$ собственных значений, равных нулю.

Если μ — некоторое собственное значение матрицы A, то, как известно, существует такой вектор $\boldsymbol{x} \neq \boldsymbol{0}$, что $A\boldsymbol{x} = \mu\boldsymbol{x}$, откуда получаем, что

$$\mu = \frac{\boldsymbol{x}'A\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}},$$

так как $\boldsymbol{x}'\boldsymbol{x} = \langle \boldsymbol{x}, \boldsymbol{x}\rangle > 0$. Если теперь A — определенная матрица, то, очевидно, $\mu \geqslant 0$ и r ненулевых собственных значений должны быть все положительны.

Наоборот, предположим, что A имеет собственные значения $\mu_1, \ldots, \mu_r > 0$ и $\mu_{r+1} = \ldots = \mu_n = 0$. Тогда равенство $A = XAX'$ означает, что ранг матрицы A равен r.

Если столбцы в X — это собственные векторы $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, то можно написать $A = \sum\limits_{j=1}^{r} \mu_j G_j$, где $G_j = \boldsymbol{x}_j\boldsymbol{x}_j'$ (теорема 2.5.1). Поэтому для любого ненулевого $\boldsymbol{x} \in \mathcal{R}_n$

$$\boldsymbol{x}'A\boldsymbol{x} = \sum_{j=1}^{r} \mu_j (\boldsymbol{x}'\boldsymbol{x}_j)(\boldsymbol{x}_j'\boldsymbol{x}) = \sum_{j=1}^{r} \mu_j (\boldsymbol{x}_j'\boldsymbol{x})^2 \geqslant 0.$$

Кроме того, если $r < n$, то $\boldsymbol{x}_{r+1}'A\boldsymbol{x}_{r+1} = 0$ и A поэтому неотрицательно определенная. Если $r = n$, то $\boldsymbol{x}_j'\boldsymbol{x} \neq 0$ по крайней мере для одного j, так что $\boldsymbol{x}'A\boldsymbol{x} > 0$ и A положительно определенная, чем доказательство завершается.

В частности, следует заметить, что вещественная симметрическая матрица A *положительно* определенная в том и только в том случае, когда все ее собственные значения положительны.

Упр. 1. Для каких вещественных значений λ следующие матрицы положительно определены:

$$\begin{Vmatrix} 1 & \lambda & \lambda \\ \lambda & 1 & \lambda \\ \lambda & \lambda & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 1 & 1 & \lambda \\ 1 & 1 & 1 \\ \lambda & 1 & 1 \end{Vmatrix}?$$

Упр. 2. Доказать, что вещественная проекция, не являющаяся единичной матрицей, должна быть неотрицательно определенной.

Упр. 3. Если M — вещественная проекция и k_1, k_2 не равны нулю, то доказать, что матрица

$$V = k_1^2 M + k_2^2 (I - M)$$

положительно определена.

Упр. 4. Сформулировать условие на a, b, c, при выполнении которого квадратичная форма $ax^2 + bxy + cy^2$ положительно определенная.

Упр. 5. Доказать, что если A — вещественная симметрическая матрица, то матрица $I + \varepsilon A$ положительно определена для достаточно малых вещественных чисел ε.

Теорема 2.14.2. *Матрица A — определенная матрица ранга $r \iff$ существует такая определенная матрица $A^{1/2}$ ранга r, что $(A^{1/2})^2 = A$.*

Доказательство. Пусть сначала A определенная. Тогда по предыдущей теореме собственные значения $\mu_1, \ldots, \mu_n$ матрицы A неотрицательны. Определим матрицу $U^{1/2}$ как $\operatorname{diag}\{\mu_1^{1/2}, \mu_2^{1/2}, \ldots, \mu_n^{1/2}\}$, где в каждом случае берется неотрицательный квадратный корень. Положив $A^{1/2} = XU^{1/2}X'$, сразу же получаем, что $A^{1/2}$ — вещественная симметрическая матрица ранга r. Кроме того,

$$(A^{1/2})^2 = XU^{1/2}X'XU^{1/2}X' = XUX' = A.$$

Наоборот, если задана определенная матрица $A^{1/2}$ ранга r, для которой $(A^{1/2})^2 = A$, то $A^{1/2}$ имеет r положительных собственных значений и $n - r$ нулевых собственных значений. Теорема 2.5.2 означает, что собственные значения матрицы A имеют те же самые свойства и результат следует из теоремы 2.14.1. ◀

Мы можем изложить теперь новый подход к определенным матрицам и характеризовать их свойствами определителей. Прежде всего имеет место

Теорема 2.14.3. *Вещественная симметрическая матрица A положительно определена $\iff$ все ее главные миноры положительны.*

Доказательство. Предположим, что A положительно определена. Тогда A неособая и матрица $A^{1/2}$ из предыдущей теоремы тоже неособая, так как мы, очевидно, имеем $(\det A^{1/2})^2 = \det A$. Пусть столбцы матрицы $A^{1/2}$ — это $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$. Так как $A^{1/2}$ вещественная и симметрическая, мы имеем тогда для любого главного минора матрицы A

$$A\begin{pmatrix} i_1 & i_2 & \ldots & i_p \\ i_1 & i_2 & \ldots & i_p \end{pmatrix} = G(\boldsymbol{a}_{i_1}, \boldsymbol{a}_{i_2}, \ldots, \boldsymbol{a}_{i_p}) > 0$$

по теореме 2.7.2. Это то, что мы должны были доказать.

Наоборот, предположим, что A вещественная и симметрическая и все ее главные миноры положительны. Заметим, что одним из таких миноров будет и $\det A$. Пусть характеристический многочлен матрицы A — это

$$c(\mu) = \mu^n - c_1\mu^{n-1} + \ldots + (-1)^n c_n,$$

где, согласно теореме 2.1.2, c_r есть сумма всех главных миноров порядка r для $r = 1, 2, \ldots, n$. Таким образом, $c_r > 0$ для каждого r. В частности, $c_n > 0$, а поэтому все собственные значения должны быть отличны от нуля. Если теперь $\mu < 0$, то ясно, что каждый член в $c(\mu)$ одного и того же знака, а поэтому $c(\mu) \neq 0$. Следовательно, собственные значения матрицы должны быть положительны. Из теоремы 2.14.1 теперь следует, что A положительно определена.

На практике нам не хотелось бы испробовать все главные миноры для того, чтобы узнать, будет ли некоторая матрица положительно определенной, и этого можно избежать при помощи следующего более сильного результата, который с вычислительной точки зрения применять не так уже сложно.

Теорема 2.14.4. *Вещественная симметрическая $n \times n$-матрица A будет положительно определенной $\Leftrightarrow$ ведущие главные миноры*

$$A\begin{pmatrix} 1 \\ 1 \end{pmatrix},\ A\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \ldots,\ A\begin{pmatrix} 1 & 2 \ldots n \\ 1 & 2 \ldots n \end{pmatrix}$$

положительны.

Используя предыдущую теорему, мы должны лишь доказать, что если ведущие главные миноры положительны, то A положительно определенная. Прямое доказательство этого результата представляет некоторые трудности. Читателю предоставляется возможность либо найти такое доказательство, либо развить доказательство по индукции, намечаемое в следующей последовательности из четырех упражнений.

Упр. 6. Пусть $c_k(\mu)$, $k = 1, 2, \ldots, n$, — характеристический многочлен блока

$$A_k = \begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ \cdot & \cdot & & \cdot \\ \vdots & \vdots & & \vdots \\ a_{1k} & a_{2k} & \ldots & a_{kk} \end{Vmatrix}$$

матрицы A и

$$A_{k+1} = \begin{Vmatrix} A_k & a_{k+1} \\ a'_{k+1} & a_{k+1,\,k+1} \end{Vmatrix}.$$

Используя упр. 13 на стр. 53, доказать, что если μ не является собственным значением для A_k, то

$$c_{k+1}(\mu) = c_k(\mu)\left\{(\mu - a_{k+1,\,k+1}) - \boldsymbol{a}'_{k+1}(\mu I - A_k)^{-1}\boldsymbol{a}_{k+1}\right\}.$$

Упр. 7. (i) Пусть $\mu_1, \ldots, \mu_k$ — собственные значения блока A_k; если тогда $c'_k(\mu_i)$ обозначает $dc_k/d\mu$, вычисленную при $\mu = \mu_i$, то показать, что

$$c'_k(\mu_i) = \prod_{\substack{j=1\\ j\neq i}}^{k} (\mu_i - \mu_j).$$

(ii) Используя упр. 3 из § 2.5, доказать, что существуют такие вещественные числа $\alpha_1, \ldots, \alpha_k$, что

$$c_{k+1}(\mu) = (\mu - a_{k+1,\,k+1})\,c_k(\mu) - \sum_{i=1}^{k}\alpha_i^2 \prod_{\substack{j=1\\ j\neq i}}^{k} (\mu - \mu_j)$$

и, следовательно,

$$c_{k+1}(\mu_i) = -\alpha_i^2 c'_k(\mu_i), \qquad i = 1, 2, \ldots, k.$$

Упр. 8. Предположим, что A_k положительно определенная, что

$$c_{k+1}(\mu_i) \neq 0, \qquad i = 1, 2, \ldots, r,$$
$$c_{k+1}(\mu_i) = 0, \qquad i = r+1, \ldots, k,$$

и что $0 < \mu_1 < \mu_2 < \ldots < \mu_r$. Пусть

$$\pi(\mu) = \sum_{i=r+1}^{k} (\mu - \mu_i) \quad \text{и} \quad c_{k+1}(\mu) = d(\mu)\,\pi(\mu)\ ^{*)}.$$

---

*) Для положительно определенной матрицы

$$A_3 = \begin{Vmatrix} 1 & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 2 \end{Vmatrix}$$

имеем $A_2 = I$ и $\mu_1 = 1$, $\mu_2 = 1$. Как нетрудно видеть, тогда $c_3(\mu_1) = c_3(\mu_2) = 0$ и $\pi(\mu) = (\mu - 1)^2$. Но в то же время $c_3(\mu) = (\mu - 1)\,[(\mu - 1)(\mu - 2) - 1]$ не делится на $(\mu - 1)^2$, т. е. на $\pi(\mu)$. Приводимая цепочка рассуждений проходит, тем не менее, для случая, когда матрица A_k не имеет кратных корней. Поэтому следует в упр. 8 считать, что матрица A_k не имеет кратных собственных значений. Затем добавить упр. 8′: если положительно определенная матрица A_k имеет кратные собственные значения, то, добавляя к A_{k+1} матрицу $\operatorname{diag}\{\varepsilon_1, \ldots, \varepsilon_{k+1}\}$ с достаточно малыми $\varepsilon_i > 0$ и такими, что $A_k + \operatorname{diag}\{\varepsilon_1, \ldots, \varepsilon_k\}$ не имеет кратных собственных значений, доказать, что матрица $A_{k+1} + \operatorname{diag}\{\varepsilon_1, \ldots, \varepsilon_{k+1}\}$ будет иметь положительные собственные значения при $\det A_{k+1} > 0$. Устремляя затем ε_i к нулю, убедиться в том, что A_{k+1} будет иметь положительные собственные значения при $\det A_{k+1} > 0$ и положительно определенной матрице A_k. — *Прим. перев.*

Доказать, что

$$d(\mu_i) = -\alpha_i^2 \prod_{\substack{j=1 \\ j \neq i}}^{r} (\mu_i - \mu_j), \qquad i = 1, 2, \ldots, r,$$

и при помощи исследования $\operatorname{sign}[d(\mu_i)]$ доказать, что если $\det A_{k+1} > 0$, то $d(\mu)$ имеет $r+1$ положительных корней.

Упр. 9. Доказать теорему 2.14.4.

Упр. 10. Доказать, что A положительно определена $\Longleftrightarrow$ $\Longleftrightarrow$ A^{-1} существует и положительно определена.

2.15. Теория малых колебаний и одновременное приведение квадратичных форм

Начиная с обсуждения классической задачи упругих колебаний механических систем, мы вводим в этом параграфе алгебраическую задачу одновременного приведения двух квадратичных форм с помощью одного и того же конгруэнтного преобразования. Вводимая алгебраическая задача возникает и в других интересных случаях, мы же ограничимся рассмотрением задачи о механических системах. Основные результаты в чисто алгебраических терминах сформулированы в теоремах 2.15.1 и 2.15.2.

Рассмотрим движение упругой механической системы, отклонение которой от некоторого устойчивого положения равновесия может быть задано n координатами. Пусть этими координатами будут $p_1, p_2, \ldots, p_n$. Если система допускает малые колебания около положения равновесия, то ее кинетическая энергия T задается квадратичной формой от скоростей изменения во времени координат, т. е. от $\dot{p}_1, \ldots, \dot{p}_n$. Таким образом, существует такая вещественная симметрическая матрица A, что

$$T = \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk} \dot{p}_j \dot{p}_k = \frac{1}{2} \dot{\boldsymbol{p}}' A \dot{\boldsymbol{p}},$$

где $\boldsymbol{p}' = \|p_1 p_2 \ldots p_n\|$ и $\dot{\boldsymbol{p}}' = \|\dot{p}_1 \dot{p}_2 \ldots \dot{p}_n\|$. Кроме того, при любом движении с ненулевым $\dot{\boldsymbol{p}} \in \mathcal{R}_n$ по физическим соображениям T будет положительной. Таким образом, T — положительно определенная матрица.

Когда система отклоняется от своего положения равновесия, в ней, вообще говоря, накапливается потенциальная энергия. Можно показать, что если координаты определены так, что $\boldsymbol{p} = \boldsymbol{0}$ в положении равновесия, и потенциальная энергия предполагается равной нулю в этом положении, то для малых колебаний она будет выражаться в виде

$$V=\frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}c_{jk}p_jp_k=\frac{1}{2}\,\boldsymbol{p}'C\boldsymbol{p},$$

где C — вещественная симметрическая неотрицательно (или положительно) определенная матрица, т. е. $V\geqslant 0$ для любого $\boldsymbol{p}\in\mathcal{R}_n$.

Мы определяем производную от некоторой матрицы (или вектора) $\dot{P}$ как матрицу, элементы которой — производные от элементов P. Для любых соответствующих друг другу матриц P и Q, элементы которых — дифференцируемые функции от x, как легко видеть,

$$\frac{d}{dx}(PQ)=\frac{dP}{dx}Q+P\,\frac{dQ}{dx}\,. \tag{2.15.1}$$

Рассматривая теперь $p_1, p_2, \ldots, p_n, \dot{p}_1, \dot{p}_2, \ldots, \dot{p}_n$ как множество $2n$ независимых переменных, получаем

$$\frac{\partial T}{\partial \dot{p}_i}=\frac{1}{2}\left(\frac{\partial \dot{\boldsymbol{p}}'}{\partial \dot{p}_i}A\dot{\boldsymbol{p}}+\dot{\boldsymbol{p}}'A\,\frac{\partial \dot{\boldsymbol{p}}}{\partial \dot{p}_i}\right)=\frac{1}{2}\left(\boldsymbol{e}_i'A\dot{\boldsymbol{p}}+\dot{\boldsymbol{p}}'A\boldsymbol{e}_i\right)=A_{i*}\dot{\boldsymbol{p}}.$$

Подобно этому

$$\frac{\partial V}{\partial p_i}=C_{i*}\boldsymbol{p}.$$

Уравнения движения для системы (при отсутствии действия внешних сил) получаются по методу Лагранжа и записываются в следующем виде:

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{p}_i}\right)-\frac{\partial T}{\partial p_i}=-\frac{\partial V}{\partial p_i},\qquad i=1,\ 2,\ \ldots,\ n.$$

Из приведенных выше выражений для производных находим, что уравнениями движения будут

$$A_{i*}\ddot{\boldsymbol{p}}=-C_{i*}\boldsymbol{p},\qquad i=1,\ 2,\ \ldots,\ n,$$

или в матричном виде

$$A\ddot{\boldsymbol{p}}+C\boldsymbol{p}=\boldsymbol{0}. \tag{2.15.2}$$

Для этого уравнения мы ищем синусоидальное решение $\boldsymbol{p}(t)$ в виде $\boldsymbol{p}(t)=\boldsymbol{q}e^{i\omega t}$, где $\boldsymbol{q}$ не зависит от t и ω вещественно. Тогда $\ddot{\boldsymbol{p}}=-\omega^2\boldsymbol{q}e^{i\omega t}$ и, записывая $\lambda=\omega^2$, будем иметь

$$(-\lambda A+C)\,\boldsymbol{q}=\boldsymbol{0},$$

или

$$\left(\lambda I-A^{-1}C\right)\boldsymbol{q}=\boldsymbol{0}.$$

Допустимые значения λ (определяющие *собственные частоты* колебания) — это n собственных значений матрицы $A^{-1}C$. *Вид колебания* определяется соответствующими правыми собственными векторами. Но прежде всего мы должны убедиться в том, что эти собственные значения *вещественны* и *неотрицательны*, ибо в противном случае наши решения не будут иметь требуемого вида с вещественным значением ω. На самом деле можно доказать, что все возможные решения уравнения (2.15.2) имеют такой вид, но эту сторону рассматриваемой задачи мы обсуждать не будем. Ясно, что собственные значения матрицы $A^{-1}C$ будут также корнями для $\det(A\lambda - C)$.

Теорема 2.15.1. *Если A и C — положительно и неотрицательно определенные матрицы соответственно, то корни для $\det(A\lambda - C)$ вещественны и неотрицательны.*

Доказательство. Как мы знаем, существует положительно определенная матрица $A^{1/2}$, для которой $(A^{1/2})^2 = A$ (теорема 2.14.2). Запишем $(A^{1/2})^{-1} = A^{-1/2}$ и

$$A\lambda - C = A^{1/2}(\lambda I - A^{-1/2}CA^{-1/2})A^{1/2}.$$

Полагая тогда $B = A^{-1/2}CA^{-1/2}$ и $\boldsymbol{r} = A^{1/2}\boldsymbol{q}$, замечаем, что $(A\lambda - C)\boldsymbol{q} = \mathbf{0}$ в том и только в том случае, когда

$$(\lambda I - B)\boldsymbol{r} = \mathbf{0}.$$

Собственные значения матрицы $A^{-1}C$ будут поэтому собственными значениями вещественной симметрической матрицы B, и наоборот. Следовательно, они вещественны.

Пусть теперь λ — собственное значение матрицы $A^{-1}C$ с вещественным правым собственным вектором $\boldsymbol{q}$. Тогда $\lambda A\boldsymbol{q} = C\boldsymbol{q}$. Умножая это равенство слева на $\boldsymbol{q}'$, получаем

$$\lambda = \frac{\boldsymbol{q}'C\boldsymbol{q}}{\boldsymbol{q}'A\boldsymbol{q}}.$$

Так как $\boldsymbol{q}'A\boldsymbol{q} > 0$ и $\boldsymbol{q}'C\boldsymbol{q} \geqslant 0$, отсюда следует, что $\lambda \geqslant 0$. ◀

Пусть теперь $\lambda_1, \ldots, \lambda_n$ — собственные значения матрицы $A^{-1}C$ и $\lambda_i = \omega_i^2$, $i = 1, 2, \ldots, n$. Так как λ_i являются собственными значениями вещественной симметрической матрицы B, использованной выше в доказательстве, то, как мы знаем (следствие теоремы 2.9.5), существует такая вещественная ортогональная матрица T, что

$$TBT' = W^2 = \operatorname{diag}\{\omega_1^2, \ldots, \omega_n^2\}.$$

Тогда $B = T'W^2T$ и равенство $(\lambda I - B)\boldsymbol{r} = \mathbf{0}$ означает, что

$$(\lambda T'T - T'W^2T)\boldsymbol{r} = \mathbf{0},$$

или

$$(\lambda I - W^2)\boldsymbol{s} = \mathbf{0},$$

где $\boldsymbol{s} = T\boldsymbol{r} = TA^{1/2}\boldsymbol{q}$, или $\boldsymbol{q} = A^{-1/2}T'\boldsymbol{s}$.

Этот результат показывает, что, производя преобразование координат, определенное при помощи $\boldsymbol{q} = A^{-1/2}T'\boldsymbol{s}$, мы одновременно приводим A и C к диагональному виду: A приводится к I и C — к W^2.

Теорема 2.15.2. *Матрицы A и C из теоремы 2.15.1 могут быть приведены к диагональным матрицам при помощи одного и того же конгруэнтного преобразования.*

Возвратимся теперь к дифференциальному уравнению (2.15.2). Положив

$$\boldsymbol{p} = A^{-1/2}T'\xi \quad \text{и} \quad \ddot{\boldsymbol{p}} = A^{-1/2}T'\ddot{\xi},$$

получим

$$A^{1/2}T'\ddot{\xi} + CA^{-1/2}T'\xi = \boldsymbol{0}.$$

Умножая это равенство на $TA^{-1/2}$ слева, найдем, что $\ddot{\xi} + W^2\xi = \boldsymbol{0}$ или

$$\ddot{\xi}_i + \omega_i^2\xi_i = 0, \qquad i = 1,\ 2,\ \ldots,\ n;$$

это означает, что

$$\xi_i = \alpha_i \cos(\omega_i t + \varepsilon_i), \quad \text{если} \quad \omega_i \neq 0.$$

Числа α и ε в этих равенствах — произвольные постоянные, которые обычно определяются заданием начальных условий. Если ξ_i изменяется по этому закону, в то время как все остальные ξ равны нулю, то мы говорим, что система колеблется по i-й *нормальной гармонике*. Координаты $\xi_1, \xi_2, \ldots, \xi_n$ называются *нормальными координатами* для системы. Заметим, что

$$T = \frac{1}{2}\dot{\boldsymbol{p}}'A\dot{\boldsymbol{p}} = \frac{1}{2}\dot{\xi}'\dot{\xi} = \frac{1}{2}(\dot{\xi}_1^2 + \dot{\xi}_2^2 + \ldots + \dot{\xi}_n^2)$$

и

$$V = \frac{1}{2}\boldsymbol{p}'C\boldsymbol{p} = \frac{1}{2}\xi'W^2\xi = \frac{1}{2}(\omega_1^2\xi_1^2 + \ldots + \omega_n^2\xi_n^2).$$

Некоторую информацию о нормальных колебаниях мы можем получить, заметив, что когда система колеблется по нормальной гармонике частоты ω_i, мы имеем $\boldsymbol{s} = \boldsymbol{e}_i$, так что решение для *вектора нормального вида* будет

$$\boldsymbol{q}_i = (A^{-1/2}T') = \boldsymbol{e}_i - i\text{-й столбец матрицы } (A^{-1/2}T').$$

Следовательно, матрица векторов нормальных гармоник $Q = \|\boldsymbol{q}_1\boldsymbol{q}_2 \ldots \boldsymbol{q}_n\|$ задается в виде $Q = A^{-1/2}T'$. Кроме того,

$$Q'AQ = I, \quad Q'CQ = W^2, \tag{2.15.3}$$

причем эти соотношения можно считать выражающими свойства биортогональности векторов нормальных гармоник.

2.16. Колебания с внешними силами

Мы предполагаем, что, как и выше, упругая система производит лишь малые отклонения от положения равновесия, но что дополнительно имеются предписанные силы $f_1(t), \ldots, f_n(t)$, прилагаемые к координатам $p_1, p_2, \ldots, p_n$. Определяя $\boldsymbol{f}(t) = \|f_1(t), \ldots, f_n(t)\|$, будем иметь теперь такое уравнение движения:

$$A\ddot{\boldsymbol{p}} + C\boldsymbol{p} = \boldsymbol{f}(t). \qquad (2.16.1)$$

Полагая $p = A^{-1/2}T'\xi = Q\xi$ и умножая это уравнение на Q' слева, приводим его к виду

$$\ddot{\xi} + W^2\xi = Q'\boldsymbol{f}(t),$$

или

$$\ddot{\xi}_j + \omega_j^2\xi_j = \boldsymbol{q}_j'\boldsymbol{f}(t), \qquad j = 1, 2, \ldots, n.$$

При $\omega_j \neq 0$ частный интеграл такого уравнения может быть получен в таком виде:

$$\xi_j(t) = \frac{1}{\omega_j}\int_0^t \{q_j'f(\tau)\} \sin\omega_j(t-\tau)\, d\tau$$

и при $\omega_j = 0$ — в виде

$$\xi_j(t) = \int_0^t \{q_j'f(\tau)\}(t-\tau)\, d\tau.$$

Если $\det W \neq 0$, то можно написать

$$\xi(t) = W^{-1}\int_0^t \{\sin W(t-\tau)\}\, Q'\boldsymbol{f}(\tau)\, d\tau,$$

где интегрирование векторного выражения понимается как почленное интегрирование и j-й диагональный член диагональной матрицы $\sin W(t-\tau)$ есть $\sin\omega_j(t-\tau)$.

Важный для практики случай возникает при $\boldsymbol{f}(t) = \boldsymbol{f}_0 e^{i\omega t}$, где $\boldsymbol{f}_0$ не зависит от t. Это случай задания синусоидальной внешней (или возбуждающей) силы. В качестве упражнения читателю предоставляется доказать, что если $\omega \neq \omega_j$ $(j = 1, 2, \ldots, n)$, то

$$\boldsymbol{p}(t) = Q\xi(t) = e^{i\omega t}\sum_{i=1}^{n} \boldsymbol{q}_j \frac{\omega_j(\boldsymbol{q}_j'\boldsymbol{f}_0)}{\omega_j^2 - \omega^2}$$

будет решением уравнения (2.16.1).

Этот результат можно использовать для обсуждения явления резонанса. Если представить себе ω, приложенную частоту, непрерывно меняющейся в окрестности некоторой собственной частоты ω_j, то вид решения указывает на то, что величина вектора-решения (или результирующей амплитуды), вообще говоря, стремится к бесконечности при $\omega \to \omega_j$. Однако особенности можно избежать, если $\boldsymbol{q}'_j \boldsymbol{f}_0 = 0$ для всех $\boldsymbol{q}_j$, соответствующих ω_j.

Смешанные упражнения

(i) Если $A \in \mathscr{C}_{2\times 2}$, $A \neq 0$ и A имеет два собственных значения, равных нулю, то доказать, что A дефектная.

(ii) Если $A \in \mathscr{C}_{2\times 2}$, то показать, что A дефектная с собственным значением $\alpha \Longleftrightarrow A$ имеет вид

$$\begin{Vmatrix} \alpha+\beta & \gamma \\ -\delta & \alpha-\beta \end{Vmatrix},$$

где $\beta^2 = \gamma\delta$ и одно из чисел γ, δ не равно нулю.

2. Пусть α, β — вещественные числа. При каких условиях следующая матрица унитарна:

$$\begin{Vmatrix} 0 & \alpha & 0 & i\beta \\ \alpha & 0 & i\beta & 0 \\ 0 & i\beta & 0 & \alpha \\ i\beta & 0 & \alpha & 0 \end{Vmatrix}?$$

3. Найти собственные значения матрицы $f(A) = A^2 - 2A + 3I$, если

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

4. Если $\boldsymbol{a} \in \mathscr{C}_n$, то найти собственные значения матрицы в $\mathscr{C}_{n\times n}$, каждый столбец которой равен $\boldsymbol{a}$.

5. Если $A \in \mathscr{C}_{m\times n}$, то показать, что $A'A$ и AA' имеют одинаковые ненулевые собственные значения.

6. Для идемпотентной матрицы A ранга r показать, что $\operatorname{tr}(A) = r$.

7. Если $A \in \mathscr{R}_{n\times n}$ и симметрическая, то показать, что A положительно определенная $\Longleftrightarrow B'AB$ положительно определенная для всех неособых матриц $B \in \mathscr{R}_n$.

8. При $\boldsymbol{a} \in \mathscr{C}_n$ и $\boldsymbol{a} \neq \boldsymbol{0}$ построить эрмитову унитарную матрицу A, первая строка которой кратна $\boldsymbol{a}'$.

9. Найти ортогональную матрицу, которая представляет поворот на 45° вокруг прямой в $\mathscr{R}_3$, соединяющей точки (0, 0, 0) и (1, 1, 1).

10. (а) Доказать, что если A_n — $n \times n$-матрица

$$\begin{Vmatrix} 2c & 1 & & \dots & 0 \\ 1 & 2c & 1 & & \\ & 1 & 2c & 1 & \\ & & \ddots & \ddots & \ddots \\ & & & \ddots & 1 \\ 0 & \dots & & 1 & 2c \end{Vmatrix},$$

$D_n = \det A_n$ и $c = \cos\theta$, то

$$D_n = \frac{\sin(n+1)\theta}{\sin\theta},$$

и показать, как D_n может быть вычислено при $|c| > 1$. [Указание. $D_n(c)$ — многочлены Чебышева второго рода $U_n(c)$.]

(б) Найти собственные значения и собственные векторы матрицы A_n. (Матрица A_n с $c = -1$ встречается в релеевском конечномерном приближении к задаче о колеблющейся струне.)

11. Если A, B и C положительно определенные, то доказать, что корни многочлена

$$\det(\lambda^2 A + \lambda B + C)$$

имеют отрицательные вещественные части.

12. Для $A, B \in \mathscr{C}_{n\times n}$ показать, что AB и BA имеют одинаковые характеристические многочлены и, следовательно, одни и те же собственные значения.

[Указания. Предположить сначала, что A неособая, и доказать, что

$$\det(\mu I - AB) = (\det A)(\det(\mu I - BA))(\det A^{-1}),$$

откуда получается результат в этом случае.

В общем случае использовать тождество многочленов от λ.

$$\det(\mu I - (\lambda I - A)B) = \det(\mu I - B(\lambda I - A)).]$$

13. Если A и B эрмитовы, то доказать, что:

(i) собственные значения матрицы $AB + BA$ вещественны;

(ii) собственные значения матрицы $AB - BA$ чисто мнимые.

14. Если A и B — неотрицательно определенные матрицы, то доказать, что:

(i) собственные значения матрицы AB вещественны и неотрицательны;

(ii) $AB = 0 \Leftrightarrow \operatorname{tr}(AB) = 0$.

15. Пусть $A \in \mathscr{C}_{n\times n}$. Показать, что:

(i) $\mathscr{N}(A^2) \supseteq \mathscr{N}(A)$; если $\mathscr{N}(A) = \mathscr{N}(A^2)$, то $\mathscr{N}(A^p) = \mathscr{N}(A)$, $p = 1, 2, \dots$;

(ii) если нулевое собственное значение матрицы A имеет одинаковые кратность и геометрическую кратность, то $\mathcal{N}(A) = \mathcal{N}(A^2)$;

(iii) $\mathcal{R}(A^2) \subseteq \mathcal{R}(A)$; если A простая, то $\mathcal{R}(A^2) = \mathcal{R}(A)$.

16. Если U — унитарная матрица, то доказать, что $\mathscr{C}_n = \mathcal{R}(U) \oplus \mathcal{N}(U)$ и что $\mathcal{R}(U)$ и $\mathcal{N}(U)$ — ортогональные дополнения.

17. Пусть S вещественная и кососимметрическая. Доказать, что тогда $I + S$ неособая и что преобразование Кели

$$T = (I - S)(I + S)^{-1}$$

ортогонально.

18. Пусть A — вещественная ортогональная матрица и $A + I$ неособая. Доказать, что можно записать

$$A = (I - S)(I + S)^{-1},$$

где S — вещественная кососимметрическая матрица.

19. Определим свойство A для эрмитовой матрицы W порядка n следующим образом: существуют подпространства $\mathscr{C}_+$ и $\mathscr{C}_-$ такие, что $\mathscr{C}_n = \mathscr{C}_+ \oplus \mathscr{C}_-$, $\langle W\boldsymbol{x}, \boldsymbol{x}\rangle > 0$ для всех ненулевых $\boldsymbol{x} \in \mathscr{C}_+$ и $\langle W\boldsymbol{x}, \boldsymbol{x}\rangle < 0$ для всех ненулевых $\boldsymbol{x} \in \mathscr{C}_-$. Доказать, что W имеет свойство $A \Longleftrightarrow \det W \neq 0$.

Глава 3

ВАРИАЦИОННЫЙ МЕТОД

3.1. Введение

В этой главе мы будем иметь дело главным образом с вещественными симметрическими матрицами, действующими в $\mathscr{R}_n$, хотя развиваемые идеи без труда обобщаются на эрмитовы матрицы, действующие в $\mathscr{C}_n$. Применяемая техника может быть представлена как обобщение геометрического подхода к задаче о собственных значениях для вещественных симметрических матриц из $\mathscr{R}_{3\times 3}$. Мы видели (упр. 3 из § 2.13), что квадрике может быть поставлена в соответствие такая матрица при помощи уравнения $x'Ax = 1$, или $\langle x, Ax\rangle = 1$. Точка (ξ_1, ξ_2, ξ_3) на поверхности, которая одновременно лежит на главной оси, обладает тем свойством, что длина задающего ее связанного вектора $(\xi'\xi)^{1/2} = \langle\xi, \xi\rangle^{1/2}$ является стационарной величиной относительно вариаций координат на поверхности.

Это свойство приводит нас к задаче исследования стационарных значений $\langle x, x\rangle$ при условии, что x удовлетворяет уравнению $\langle x, Ax\rangle = 1$. Эта задача имеет другую формулировку, в которой исключается ограничивающее условие: для этого следует заняться поиском стационарных значений отношения $\langle x, x\rangle/\langle x, Ax\rangle$, где x пробегает все ненулевые векторы в $\mathscr{R}_3$; или, что по существу то же самое, мы разыскиваем стационарные значения для *отношения Релея*

$$\frac{\langle x, Ax\rangle}{\langle x, x\rangle}.$$

Последняя формулировка дает определение отношения Релея и ставит важную задачу для *любой* эрмитовой матрицы A.

3.2. Экстремальные собственные значения и отношение Релея

Пусть A — вещественная симметрическая $n\times n$-матрица. Определим *единичную сферу* $\mathscr{O}$ в $\mathscr{R}_n$ (или в $\mathscr{C}_n$) как множество всех векторов в $\mathscr{R}_n(\mathscr{C}_n)$, для которых $\langle x, x\rangle = 1$. Рассмотрим отношение Релея для A

$$R(\boldsymbol{x}) = \frac{\langle \boldsymbol{x}, A\boldsymbol{x}\rangle}{\langle \boldsymbol{x}\ \boldsymbol{x}\rangle}. \tag{3.2.1}$$

Очевидно, если вместо $\boldsymbol{x}$ подставить $k\boldsymbol{x}$, то отношение Релея останется без изменения. С геометрической точки зрения можно сказать, что $R(\boldsymbol{x})$ принимает одно и то же значение для каждой точки (связанного вектора) на прямой, проходящей через начало координат в $\mathscr{R}_n$ (исключая само начало координат, где $R(\boldsymbol{0})$ не определено). Таким образом, если ограничиться рассмотрением векторов $\boldsymbol{x}$, соответствующих точкам на единичной сфере, то для $R(\boldsymbol{x})$ получим ту же самую совокупность значений. Мы поэтому рассмотрим задачу нахождения максимума и минимума значений *) для $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle$ при условии, что $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = 1$. Запишем эти величины:

$$\max_{\boldsymbol{x}\in\mathcal{O}} \langle \boldsymbol{x}, A\boldsymbol{x}\rangle, \quad \min_{\boldsymbol{x}\in\mathcal{O}} \langle \boldsymbol{x}, A\boldsymbol{x}\rangle.$$

Как мы видели,

$$\max_{\boldsymbol{x}\in\mathcal{O}} R(\boldsymbol{x}) = \max_{\boldsymbol{x}\in\mathcal{O}} \langle \boldsymbol{x}, A\boldsymbol{x}\rangle$$

и подобно этому для минимума.

Предположим, что собственными значениями матрицы A будут $\mu_1 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_n$, что $U = \operatorname{diag}\{\mu_1, \ldots, \mu_n\}$ и что T — вещественная ортогональная матрица, для которой

$$T'AT = U. \tag{3.2. 2}$$

Отметим, что если $\boldsymbol{x} = T\boldsymbol{y}$, то

$$\langle \boldsymbol{x}, \boldsymbol{x}\rangle = \boldsymbol{x}'\boldsymbol{x} = \boldsymbol{y}'T'T\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{y} = \langle \boldsymbol{y}, \boldsymbol{y}\rangle.$$

Кроме того, так как T обязательно неособая, $R(T') = \mathscr{R}_n$. Отсюда следует, что если $\boldsymbol{x}$ пробегает все векторы в $\mathcal{O}$, то и $\boldsymbol{y}$ пробегает все векторы в $\mathcal{O}$ и

$$\max_{\boldsymbol{x}\in\mathcal{O}} \langle \boldsymbol{x}, A\boldsymbol{x}\rangle = \max_{\boldsymbol{y}\in\mathcal{O}} \langle T\boldsymbol{y}, AT\boldsymbol{y}\rangle = \max_{\boldsymbol{y}\in\mathcal{O}} (\boldsymbol{y}'T'AT\boldsymbol{y}) =$$

$$= \max_{\boldsymbol{y}\in\mathcal{O}} (\boldsymbol{y}'U\boldsymbol{y}) = \max_{\boldsymbol{y}\in\mathcal{O}} \left(\sum_{j=1}^{k} \mu_j y_j^2\right).$$

Подобно этому

$$\min_{\boldsymbol{x}\in\mathcal{O}} \langle \boldsymbol{x}, A\boldsymbol{x}\rangle = \min_{\boldsymbol{y}\in\mathcal{O}} \left(\sum_{j=1}^{n} \mu_j y_j^2\right).$$

*) См. дополнение 1, где обсуждается вопрос о существовании максимума и минимума.

Но согласно упорядочению собственных значений имеем далее

$$\sum_{j=1}^{n} \mu_j y_j^2 \leqslant \mu_1 \sum_{j=1}^{n} y_j^2 = \mu_1 \langle \boldsymbol{y}, \boldsymbol{y} \rangle = \mu_1$$

и

$$\sum_{j=1}^{n} \mu_j y_j^2 \geqslant \mu_n \sum_{j=1}^{n} y_j^2 = \mu_n \langle \boldsymbol{y}, \boldsymbol{y} \rangle = \mu_n,$$

если $\boldsymbol{y} \in \mathcal{O}$. Кроме того, при $\boldsymbol{y} = \boldsymbol{e}_1$ и $\boldsymbol{y} = \boldsymbol{e}_n$ получаем $\sum \mu_j y_j^2 = \mu_1$ и $\sum \mu_j y_j^2 = \mu_n$ соответственно. Следовательно, так как $\boldsymbol{e}_1$ и $\boldsymbol{e}_n \in \mathcal{O}$,

$$\max_{\boldsymbol{x} \in \mathcal{O}} \langle \boldsymbol{x}, A\boldsymbol{x} \rangle = \mu_1 \quad \text{и} \quad \min_{\boldsymbol{x} \in \mathcal{O}} \langle \boldsymbol{x}, A\boldsymbol{x} \rangle = \mu_n.$$

Этот результат может быть сформулирован следующим образом в терминах отношения Релея, определенного по (3.2.1).

Теорема 3.2.1. *Если A — некоторая вещественная симметричная матрица, то для любого ненулевого $\boldsymbol{x} \in \mathcal{R}_n$*

$$\mu_n \leqslant R(\boldsymbol{x}) \leqslant \mu_1$$

и

$$\mu_1 = \max_{\boldsymbol{x} \in \mathcal{O}} R(\boldsymbol{x}), \quad \mu_n = \min_{\boldsymbol{x} \in \mathcal{O}} R(\boldsymbol{x}).$$

Этот результат, очевидно, очень легко может быть применен к нахождению границ для экстремальных собственных значений матрицы A.

***Упр. 1.** Для собственного вектора $\boldsymbol{x}_j$ с собственным значением μ_j показать, что $R(\boldsymbol{x}_j) = \mu_j$.

Упр. 2. Доказать, что если $\alpha = n^{-1} \sum_{j,k} a_{jk}$, то

$$\mu_n \leqslant \alpha \leqslant \mu_1, \quad \mu_n \leqslant a_{jj} \leqslant \mu_1$$

для $j = 1, 2, \ldots, n$.

***Упр. 3.** При рассмотрении $R(\boldsymbol{x})$ как функции на $\mathcal{R}_n$ теорема 3.2.1 утверждает, что область значений этой функции содержится в отрезке $[\mu_n, \mu_1]$. Существуют ли точки отрезка $[\mu_n, \mu_1]$, которые не принадлежат области значений R?

3.3. Свойство стационарности отношения Релея

Во вводных замечаниях мы упоминали о стационарных значениях отношения Релея. Более точно эти понятия мы формулируем в следующей теореме, в которой $R(\boldsymbol{x})$ рассматривается как вещественнозначная функция независимых вещественных переменных $x_1, x_2, \ldots, x_n$ — компонент вектора $\boldsymbol{x}$. Мы пишем $R(\boldsymbol{x}) = R(x_1, x_2, \ldots, x_n)$.

Теорема 3.3.1. *Если A — вещественная и симметрическая матрица с собственным значением μ и соответствующим собственным вектором ξ, то $R(x_1, \ldots, x_n)$ имеет стационарное значение относительно $x_1, \ldots, x_n$ при $\boldsymbol{x} = \xi$ и $R(\xi) = \mu$.*

Доказательство. То, что $R(\xi) = \mu$, получается без труда и было отмечено в упр. 1 выше.

Для получения свойства стационарности мы должны показать, что $\partial R/\partial x_j = 0$ при $\boldsymbol{x} = \xi$ для $j = 1, 2, \ldots, n$. Но как в § 2.15, находим, что

$$\frac{\partial (\boldsymbol{x}'A\boldsymbol{x})}{\partial x_j} = 2A_{j*}\boldsymbol{x} \quad \text{и} \quad \frac{\partial (\boldsymbol{x}'\boldsymbol{x})}{\partial x_j} = 2x_j.$$

Следовательно,

$$\frac{\partial R}{\partial x_j} = \frac{(\boldsymbol{x}'\boldsymbol{x})(2A_{j*}\boldsymbol{x}) - (\boldsymbol{x}'A\boldsymbol{x})\,2x_j}{(\boldsymbol{x}'\boldsymbol{x})^2}.$$

Вычисляя это выражение при $\boldsymbol{x} = \xi$, положив $A\xi = \mu\xi$ и $A_{j*}\xi = \mu\xi_j$, находим, что числитель равен нулю, так что $\partial R/\partial x_j = 0$ при $\boldsymbol{x} = \xi$ для каждого j. ◀

Упр. 1. Сформулировать и доказать обращение теоремы 3.3.1.

Имеется интересное обобщение теоремы 3.3.1 и неэрмитовы матрицы. Мы рассмотрим некоторую простую матрицу A с собственным значением μ и соответствующей сопутствующей матрицей $G = \xi\eta'$. Таким образом, имеем $\eta'\xi = 1$ и

$$A\xi = \mu\xi, \quad \eta'A = \mu\eta' \tag{3.3.1}$$

(теорема 2.4.3). Для простоты это предположение берется даже в более сильном виде, чем необходимо; существенным условием является то, чтобы μ имело одинаковые кратность и геометрическую кратность.

Для матрицы A мы определяем обобщенное отношение Релея как

$$R(\boldsymbol{x}, \boldsymbol{y}) = \frac{\boldsymbol{y}'A\boldsymbol{x}}{\boldsymbol{y}'\boldsymbol{x}} \tag{3.3.2}$$

и рассматриваем R как функцию от $2n$ скалярных переменных $x_1, x_2, \ldots, x_n, y_1, \ldots, y_n$. Получаем тогда, что

$$\frac{\partial R}{\partial x_j} = \frac{(\boldsymbol{y}'\boldsymbol{x})(\boldsymbol{y}'A_{*j}) - (\boldsymbol{y}'A\boldsymbol{x})\,y_j}{(\boldsymbol{y}'\boldsymbol{x})^2}$$

и

$$\frac{\partial R}{\partial y_j} = \frac{(\boldsymbol{y}'\boldsymbol{x})(A_{j*}\boldsymbol{x}) - (\boldsymbol{y}'A\boldsymbol{x})\,x_j}{(\boldsymbol{y}'\boldsymbol{x})^2}.$$

Если вычислить эти производные при $\boldsymbol{x} = \boldsymbol{\xi}$, $\boldsymbol{y} = \boldsymbol{\eta}$, используя (3.3.1) и $\boldsymbol{\eta}'\boldsymbol{\xi} = 1$, то легко убеждаемся в том, что в этих точках имеем $\partial R/\partial x_j = 0$ и $\partial R/\partial y_j = 0$ для каждого j. Имеем также $R(\boldsymbol{\xi}, \boldsymbol{\eta}) = \mu$. Эти результаты являются очень естественным расширением теоремы 3.3.1 на неэрмитовы матрицы, но следует ясно себе представлять, что существуют матрицы, для которых $R(\boldsymbol{x}, \boldsymbol{y})$ не определено при $\boldsymbol{x} = \boldsymbol{\xi}$, $\boldsymbol{y} = \boldsymbol{\eta}$. В частности, если $\boldsymbol{\eta}'\boldsymbol{\xi} = 0$ для всех левых и правых собственных векторов, соответствующих данному собственному значению, то отношение (3.3.2) не определено при $\boldsymbol{x} = \boldsymbol{\xi}$, $\boldsymbol{y} = \boldsymbol{\eta}$. Свойство стационарности (если оно выполняется) очень полезно при формулировке методов вычисления собственных значений.

Можно также заметить, что мы не можем ожидать прямого обобщения теоремы 3.2.1, так как собственные значения неэрмитовых матриц могут быть комплексными.

3.4. Вариационное описание собственных значений

Возвратимся к теме § 3.2. Нам удалось там характеризовать экстремальные собственные значения вещественной симметрической матрицы вариационными методами. Мы попытаемся теперь этим способом характеризовать все собственные значения. Свойство стационарности из теоремы 3.3.1 локальное и не предоставляет прямой возможности характеризовать собственные значения.

Направление поиска можно предложить опять при помощи обдумывания классификации поверхностей квадрик. В частности, рассмотрим эллипсоид, главная ось которого уже известна (возможно, при помощи теоремы 3.2.1) и направлена вдоль вектора $\boldsymbol{\xi}_1$. Свойства эллипсоида подсказывают мысль попытаться определить вторую главную ось при помощи максимизации $R(\boldsymbol{x})$, где $\boldsymbol{x}$ теперь пробегает все ненулевые векторы в R_n, *ортогональные* к $\boldsymbol{\xi}_1$. Если бы это нам удалось, то мы попытались бы найти третью главную ось при помощи максимизации $R(\boldsymbol{x})$, где $\boldsymbol{x}$ пробегает все ненулевые векторы, ортогональные к порожденному $\boldsymbol{\xi}_1$ и $\boldsymbol{\xi}_2$ пространству. Оказывается, что этот простой подход есть все, что нам нужно.

Как и прежде, мы предположим, что $\mu_1 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_n$ — собственные значения вещественной симметрической матрицы A и $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ — соответствующие собственные векторы, образующие ортонормированную систему. Пусть $\mathcal{Y}_{n-p+1}$ — подпространство в $\mathcal{R}_n$, порожденное $\boldsymbol{x}_p, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_n$ для $p = 1, 2, \ldots, n$. Тогда, как очевидно,

$$\mathcal{R}_n = \mathcal{Y}_n \supset \mathcal{Y}_{n-1} \supset \ldots \supset \mathcal{Y}_1$$

и $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p$ ортогональны к каждому вектору из $\mathcal{Y}_{n-p}$.

Теорема 3.4.1. *Для* $j = 1, 2, \ldots, n$ *имеем*

$$\mu_j = \max R(\boldsymbol{x}),$$

где максимум берется по всем ненулевым векторам $\boldsymbol{x}$ *из* $\mathcal{Y}_{n-j+1}$.

Доказательство. Прежде всего заметим, что случай $j = 1$ в точности совпадает с теоремой 3.2.1 и что $\boldsymbol{x}_j$ — это столбцы матрицы T в равенстве (3.2.2). Кроме того, $\mathcal{Y}_{n-j+1}$ порождается $\boldsymbol{x}_j, \ldots, \boldsymbol{x}_n$. Для заданного $\boldsymbol{x} \in \mathcal{Y}_{n-j+1}$ существуют такие вещественные числа $y_j, \ldots, y_n$, что $\boldsymbol{x} = \sum_{k=j}^{n} y_k \boldsymbol{x}_k = T\boldsymbol{y}$, где $\boldsymbol{y}' = \| 0, \ldots, 0, y_j, \ldots, y_n \|$. Имеем тогда

$$\langle \boldsymbol{x}, A\boldsymbol{x} \rangle = \boldsymbol{x}' A \boldsymbol{x} = \boldsymbol{y}' T' A T \boldsymbol{y} = \boldsymbol{y}' U \boldsymbol{y} = \sum_{k=j}^{n} \mu_k y_k^2$$

и подобно этому

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \sum_{k=j}^{n} y_k^2.$$

Опять следует рассматривать лишь те векторы из $\mathcal{Y}_{n-j+1}$, для которых $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 1$. Но, как очевидно, на этом множестве имеем

$$\max \langle \boldsymbol{x}, A\boldsymbol{x} \rangle = \mu_j$$

и этот максимум достигается при $\boldsymbol{x}$ с $\boldsymbol{y}_j = 1$, $y_{j+1} = \ldots = y_n = 0$, т. е. при $\boldsymbol{x} = \boldsymbol{x}_j$. Это завершает доказательство теоремы.

Этот результат имеет тот недостаток, что μ_j $(j > 1)$ характеризуется с помощью собственных векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{j-1}$. Таким образом, при применении теоремы 3.4.1 мы должны вычислять собственные значения в порядке их невозрастания вместе с ортонормированными базисами для их собственных подпространств. Следующая теорема — это замечательный результат, позволяющий характеризовать каждое собственное значение без обращения к другим собственным значениям и собственным векторам. Это знаменитая теорема Куранта — Фишера, начало которой было положено Э. Фишером в 1905 г. и которая была переоткрыта Р. Курантом в 1920 г. При этом именно Курант увидел и сделал ясным для всех большую важность этого результата для задач математической физики.

3.5. Задачи со связями

В этом месте будет удобно ввести операцию вычитания подпространств. Если $\mathcal{C}$, $\mathcal{A}$ — подпространства и $\mathcal{A} \subseteq \mathcal{C}$, то через $\mathcal{C} \ominus \mathcal{A}$ мы обозначаем множество всех векторов из $\mathcal{C}$, которые ортогональны к каждому элементу из $\mathcal{A}$. Мы утверждаем, что $\mathcal{C} \ominus \mathcal{A}$ будет подпространством в $\mathcal{C}$ и, как нетрудно видеть, если $\mathscr{C} \ominus \mathscr{A} = \mathscr{B}$, то $\mathscr{C} = \mathscr{A} \oplus \mathscr{B}$, хотя обратное утверждение не обязательно верно. Это происходит по той причине, что $\mathscr{C} \ominus \mathscr{A} = \mathscr{B}$ означает, что $\mathscr{A}$ и $\mathscr{B}$ — *ортогональные* дополнения а $\mathscr{A} \oplus \mathscr{B}$ определено для пространств $\mathscr{A}$ и $\mathscr{B}$, которые лишь дополнительные. Определяя $\mathscr{C} \ominus \mathscr{A}$ как ортогональное дополнение к $\mathscr{A}$, мы получаем единственную интерпретацию операции вычитания.

Упр. 1. Доказать, что $\mathscr{C} \ominus \mathscr{A}$ — подпространство в $\mathscr{C}$.

Упр. 2. Если $\mathscr{A}$ — подпространство в $\mathscr{R}_3$, порожденное $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$, то указать все подпространства $\mathscr{B}$ в $\mathscr{R}_3$, для которых $\mathscr{A} \oplus \mathscr{B}$ определено. Какое из них совпадает с $\mathscr{R}_3 \ominus \mathscr{A}$?

Упр. 3. Пусть P — проекция на $\mathscr{A}$. Доказать, что $\boldsymbol{x} \in \mathscr{C} \ominus \mathscr{A} \Leftrightarrow P\boldsymbol{x} = 0$.

Пусть $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ — ортонормированные собственные векторы вещественной симметрической матрицы A, $\mathscr{X}_j$ — подпространство в $\mathscr{R}_n$, порожденное $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_j$ для $j = 1, 2, \ldots, n$. Напомним, что $\mathscr{Y}_{n-j+1}$ обозначало подпространство, порожденное $\boldsymbol{x}_j, \ldots, \boldsymbol{x}_n$. Имеем тогда

$$\mathscr{X}_j \oplus \mathscr{Y}_{n-j} = \mathscr{C}_n \quad \text{и} \quad \mathscr{C}_n \ominus \mathscr{X}_j = \mathscr{Y}_{n-j}, \qquad j = 1, 2, \ldots, n-1. \tag{3.5.1}$$

Задолго до формулировки и доказательства теоремы Куранта — Фишера математики встретились с вычислениями собственных значений, которые возникают из задач максимизации при некоторой *связи*. С вариационной точки зрения это означает, что мы ищем стационарные значения для $\boldsymbol{x}'A\boldsymbol{x}/\boldsymbol{x}'\boldsymbol{x}$, где $\boldsymbol{x}$ пробегает те ненулевые элементы из $\mathscr{R}_n$, которые удовлетворяют некоторому уравнению (или связи) вида

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0$$

с не всеми a_j, равными нулю. Очевидно, если $a_j \neq 0$, то мы можем использовать наложенную связь для выражения x_j в виде линейной комбинации остальных x_i, подставить в $R(\boldsymbol{x})$ и получить новое отношение от $n-1$ переменных, стационарные значения которого (и $n-1$ соответствующих собственных значений) можно исследовать непосредственно без обращения к связям. Оказывается, это очевидное направление наступления не самое выгодное, и если мы уже знаем свойства системы без связей, то мы очень мало можем сказать о свойствах системы со связями. Мы будем называть собственные значения и собственные векторы задачи со связями *ограниченными собственными значениями и ограниченными собственными векторами.*

В общем виде рассмотрим задачу нахождения собственного значения при r связях:

$$\boldsymbol{a}_j'\boldsymbol{x} = 0, \qquad j = 1, 2, \ldots, r, \tag{3.5.2}$$

где $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$ линейно независимы. Пусть $\mathscr{A}_r$ — подпространство, порожденное $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$, и $\mathscr{B}_{n-r} = \mathscr{B}_n \ominus \mathscr{A}_r$. Если $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n-r}$ — ортонормированный базис для $\mathscr{B}_{n-r}$ и $B = \|\boldsymbol{b}_1 \boldsymbol{b}_2 \ldots \boldsymbol{b}_{n-r}\|$, то $B'B = I$ и $P = BB'$ — проекция на $\mathscr{B}_{n-r}$ (§ 2.11).

Мы должны теперь найти стационарные значения для $\boldsymbol{x}'A\boldsymbol{x}/\boldsymbol{x}'\boldsymbol{x}$, где $\boldsymbol{x}$ пробегает те ненулевые векторы из $\mathscr{R}_n$, которые удовлетворяют наложенным связям, т. е. для $\boldsymbol{x} \neq 0$ и $\boldsymbol{x} \in \mathscr{B}_{n-r}$. Но $x \in \mathscr{B}_{n-r} \Longleftrightarrow$ существует такой $\xi \in \mathscr{R}_{n-r}$, что $\boldsymbol{x} = B\xi$. Таким образом, так как $B'B = I$, мы имеем

$$\frac{\boldsymbol{x}'A\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \frac{\xi'B'AB\xi}{\xi'B'B\xi} = \frac{\xi'B'AB\xi}{\xi'\xi},$$

и задача сводится к стандартному виду нахождения стационарных значений отношения Релея для $(n-r)\times(n-r)$ вещественной симметрической матрицы $B'AB$ (§ 3.3). Как мы видели, эти стационарные значения достигаются лишь в собственных значениях матрицы $B'AB$. Таким образом, μ — ограниченное собственное значение для A с соответствующим собственным вектором $\boldsymbol{x} \Longleftrightarrow \mu$ — собственное значение для $B'AB$ с собственным вектором ξ, где $x = B\xi$.

Пусть ограниченными собственными значениями будут $\tilde{\mu}_1 \geqslant \tilde{\mu}_2 \geqslant \ldots \geqslant \tilde{\mu}_{n-r}$ и $\xi_1, \ldots, \xi_{n-r}$ — соответствующие ортонормированные собственные векторы для $B'AB$. Если $\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_{n-r}$ — соответствующие ограниченные собственные векторы, то они тоже будут ортонормированными, потому что

$$\tilde{\boldsymbol{x}}_j' \tilde{\boldsymbol{x}}_k = \xi_j' B'B\xi_k = \xi_j'\xi_k = \delta_{jk}.$$

Для $j = 1, 2, \ldots, n-r$ пусть $\tilde{\mathscr{X}}_j$ и $\tilde{\mathscr{Y}}_{n-r-j+1}$ — подпространства в $\mathscr{R}_n$, порожденные $\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_j$ и $\tilde{\boldsymbol{x}}_j, \tilde{\boldsymbol{x}}_{j+1}, \ldots, \tilde{\boldsymbol{x}}_{n-r}$ соответственно. Применяя сначала теорему 3.4.1 к $B'AB$ в $\mathscr{R}_{n-r}$ и преобразуя затем результат в таковой для $\mathscr{R}_n$, читателю не представит труда доказать следующую лемму.

Лемма. *Если $\tilde{\mu}_1 \geqslant \tilde{\mu}_2 \geqslant \ldots \geqslant \tilde{\mu}_{n-r}$ — собственные значения в сформулированной выше задаче нахождения собственных значений со связями с соответствующими ортонормированными собственными векторами $\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \ldots, \tilde{\boldsymbol{x}}_{n-r}$, то*

$$\tilde{\mu}_j = \max R(\boldsymbol{x}), \qquad j = 1, 2, \ldots, n-r,$$

где максимум берется по всем ненулевым векторам в пространстве $\tilde{\mathscr{Y}}_{n-r-j+1}$, порожденном $\tilde{\boldsymbol{x}}_j, \ldots, \tilde{\boldsymbol{x}}_{n-r}$.

3.6. Теорема Куранта — Фишера

Для иллюстрации метода, применяемого в доказательстве этой теоремы, докажем сначала в рамках § 3.5, что если имеется одна связь, то, каков бы ни был ограничивающий вектор $\boldsymbol{a}_1$, имеем

$$\mu_2 \leqslant \tilde{\mu}_1 \leqslant \mu_1. \tag{3.6.1}$$

Прежде всего, так как $\tilde{\mathcal{Y}}_{n-1} \subset \mathcal{R}_n$, лемма и теорема 3.2.1 означают, что

$$\tilde{\mu}_1 = \max_{x \in \tilde{\mathcal{Y}}_{n-1}} R(\boldsymbol{x}) \leqslant \max_{x \in \mathcal{R}_n} R(\boldsymbol{x}) = \mu_1.$$

Таким образом, верхняя граница для $\tilde{\mu}_1$ легко получена.

Мы докажем теперь, что существует $\xi \in \mathcal{B}_{n-1}$, для которого $R(\xi) \geqslant \mu_2$. Так как $\tilde{\mu}_1 = \max R(\boldsymbol{x})$ в $\mathcal{B}_{n-1}$, отсюда будет следовать, что $\tilde{\mu}_1 \geqslant \mu_2$.

Так как сумма размерностей $\mathcal{X}_2$ и $\mathcal{B}_{n-1}$ равна $n+1$, то существует ненулевой вектор, принадлежащий как $\mathcal{X}_2$, так и $\mathcal{B}_{n-1}$ (см. упр. 4 из § 2.11). Пусть ξ — такой вектор с $\xi'\xi = 1$. Тогда, так как $\xi \in \mathcal{X}_2$, существуют такие вещественные числа α_1 и α_2, что $\xi = \alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2$ и $\xi'\xi = (\alpha_1 \boldsymbol{x}_1' + \alpha_2 \boldsymbol{x}_2')(\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2) = \alpha_1^2 + \alpha_2^2 = 1$.

Так как $\xi \in \mathcal{B}_{n-1}$, имеем также

$$R(\xi) \leqslant \max_{x \in \mathcal{B}_{n-1}} R(\boldsymbol{x}) = \tilde{\mu}_1.$$

Следовательно,

$$\tilde{\mu}_1 \geqslant R(\xi) = \xi' A \xi = (\alpha_1 \boldsymbol{x}_1' + \alpha_2 \boldsymbol{x}_2') A (\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2) =$$
$$= \alpha_1^2 (\boldsymbol{x}_1' A \boldsymbol{x}_1) + 2\alpha_1\alpha_2 (\boldsymbol{x}_1' A \boldsymbol{x}_2) + \alpha_2^2 (\boldsymbol{x}_2' A \boldsymbol{x}_2).$$

Так как $A\boldsymbol{x}_j = \mu_j \boldsymbol{x}_j$ для $j = 1, 2$ и $\boldsymbol{x}_j' \boldsymbol{x}_k = \delta_{jk}$, то $\boldsymbol{x}_1' A \boldsymbol{x}_2 = 0$ и

$$\tilde{\mu}_1 \geqslant \alpha_1^2 \mu_1 + \alpha_2^2 \mu_2 \geqslant (\alpha_1^2 + \alpha_2^2) \mu_2 = \mu_2,$$

так как $\mu_1 \geqslant \mu_2$ и $\alpha_1^2 + \alpha_2^2 = 1$. Это устанавливает неравенства (3.6.1), но можно сказать больше, а именно, доказать, что существует некоторое $\boldsymbol{a}_1$, для которого $\tilde{\mu}_1 = \mu_2$. Мы берем просто $\boldsymbol{a}_1 = \boldsymbol{x}_1$. Тогда $\mathcal{B}_{n-1} = \mathcal{Y}_{n-1}$ и $\boldsymbol{x}_2 \in \mathcal{Y}_{n-1}$. В таком случае по теореме 3.4.1 получаем

$$\tilde{\mu}_1 = \max_{x \in \mathcal{Y}_{n-1}} R(\boldsymbol{x}) = \mu_2.$$

Таким образом, связи можно использовать для характеризации μ_2 и записать

$$\mu_2 = \min_{\mathcal{B}_{n-1}} \max_{x \in \mathcal{B}_{n-1}} R(\boldsymbol{x}),$$

т. е. μ_2 — наименьшее значение, которое может быть достигнуто максимумом величины $R(\boldsymbol{x})$ при добавлении одной связи. В противоположность результату теоремы 3.4.1 это описание μ_2 не зависит от какого-либо знания μ_1 или его собственного подпространства. Теорема Куранта — Фишера является прямым обобщением этого результата.

Теорема 3.6.1. *Пусть заданы любые линейно независимые векторы $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_r$ в $\mathcal{R}_n$, и пусть $\mathcal{B}_{n-r}$ — подпространство всех векторов в $\mathcal{R}_n$, которые ортогональны к $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$. Тогда*

$$\mu_{r+1} = \min_{\mathcal{B}_{n-r}} \max_{\boldsymbol{x} \in \mathcal{B}_{n-r}} R(\boldsymbol{x}), \qquad r = 0, 1, 2, \ldots, n-1.$$

Таким образом, минимум определен по всем множествам линейно независимых векторов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$ в $\mathcal{R}_n$. Этот минимум достигается при $\boldsymbol{a}_j = \boldsymbol{x}_j$, $j = 1, \ldots, r$.

Доказательство. Мы дадим набросок доказательства. Рассмотрим некоторый вектор ξ, принадлежащий как $\mathcal{X}_{r+1}$, так и $\mathcal{B}_{n-r}$, с $\xi'\xi = 1$.

Прежде всего, так как $\xi \in \mathcal{X}_{r+1}$, то $\xi = \sum_{j=1}^{r+1} \alpha_j \boldsymbol{x}_j$ и

$$\xi'\xi = \sum_{j=1}^{r+1} \alpha_j^2 = 1.$$

Затем, так как $\xi \in \mathcal{B}_{n-r}$, то

$$\max_{\boldsymbol{x} \in \mathcal{B}_{n-r}} R(\boldsymbol{x}) \geqslant R(\xi) = \xi' A \xi = \alpha_1^2 \mu_1 + \ldots + \alpha_{r+1}^2 \mu_{r+1} \geqslant$$
$$\geqslant (\alpha_1^2 + \ldots + \alpha_{r+1}^2)\mu_{r+1} = \mu_{r+1}.$$

Кроме того, при $\boldsymbol{a}_j = \boldsymbol{x}_j$, $j = 1, 2, \ldots, r$, имеем $\mathcal{B}_{n-r} = \mathcal{Y}_{n-r}$ так что по теореме 3.4.1

$$\max_{\boldsymbol{x} \in \mathcal{Y}_{n-r}} R(\boldsymbol{x}) = \mu_{r+1}.$$

Таким образом, существует $\mathcal{B}_{n-r}$, для которого нижняя граница μ_{r+1} достигается, так что

$$\mu_{r+1} = \min_{\mathcal{B}_{n-r}} \max_{\boldsymbol{x} \in \mathcal{B}_{n-r}} R(\boldsymbol{x})$$

и минимум достигается при $\boldsymbol{a}_j = \boldsymbol{x}_j$, $j = 1, 2, \ldots, r$. ◀

В качестве первоначального следствия теоремы Куранта — Фишера мы докажем, что задача нахождения собственного значения, у которой порядок n и наложена одна связь, имеет $n-1$ собственных значений и наибольшее собственное значение $\tilde{\mu}_1$ с $\mu_2 \leqslant \tilde{\mu}_1 \leqslant \mu_1$. В действительности верно также, что

$$\mu_{j+1} \leqslant \tilde{\mu}_j \leqslant \mu_j, \qquad j = 2, 3, \ldots, n-1,$$

и даже, более того, если имеется r связей, определенных векторами $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$, то ограниченная задача имеет $n-r$ собственных значений и

$$\mu_{j+r} \leqslant \tilde{\mu}_j \leqslant \mu_j, \qquad j = 1, 2, \ldots, n-r.$$

Этот результат называют иногда теоремой Релея, хотя сам Релей приписывает ее Раусу и даже еще ранее она использовалась Коши. Эта теорема легко доказывается при помощи теоремы Куранта — Фишера.

Теорема 3.6.2. *Если* $\tilde{\mu}_1 \geqslant \tilde{\mu}_2 \geqslant \ldots \geqslant \tilde{\mu}_{n-r}$ — *собственные значения задачи* (3.5.2), *то*

$$\mu_{j+r} \leqslant \tilde{\mu}_j \leqslant \mu_j, \qquad j = 1, 2, \ldots, n-r.$$

Доказательство. Вспомним сначала определения пространств $\mathscr{X}_j$ и $\mathscr{Y}_{n-r-j+1}$, данные в § 3.5 для $j = 1, 2, \ldots, n-r$. Так как размерность пространства $\mathscr{Y}_{n-r-j+1}$ равна $n-(r+j)+1$, из теоремы 3.6.1 следует, что

$$\mu_{r+j} \leqslant \max R(\boldsymbol{x}),$$

где максимум берется по всем ненулевым векторам в $\mathscr{Y}_{n-r-j+1}$. Но лемма из § 3.5 утверждает, что $\max R(\boldsymbol{x}) = \tilde{\mu}_j$. Отсюда получаем неравенство: $\mu_{r+j} \leqslant \tilde{\mu}_j$.

Для нахождения второго неравенства рассмотрим собственные значения матрицы $-A$: $\omega_1 \geqslant \omega_2 \geqslant \ldots \geqslant \omega_n$, где $\omega_p = -\mu_{n+1-p}$, $p = 1, 2, \ldots, n$. Пусть собственными значениями матрицы $-A$ при связях $\boldsymbol{a}_j'\boldsymbol{x} = 0$, $j = 1, 2, \ldots, r$, будут $\tilde{\omega}_1 \geqslant \ldots \ldots \geqslant \tilde{\omega}_{n-r}$. Тогда

$$\tilde{\omega}_p = -\tilde{\mu}_{n+1-r-p}, \qquad p = 1, 2, \ldots, n-r.$$

Применяя только что доказанный для A результат, получаем

$$\omega_{r+k} \leqslant \tilde{\omega}_k, \qquad k = 1, 2, \ldots, n-r,$$

откуда

$$-\mu_{n+1-r-k} \leqslant -\tilde{\mu}_{n+1-r-k}.$$

Записывая $j = n+1-r-k$, получаем

$$\tilde{\mu}_j \leqslant \mu_j, \qquad j = 1, 2, \ldots, n-r. \quad ◀$$

Доказанные теоремы можно использовать для получения еще одного интересного результата. Если вещественная симметрическая матрица A имеет собственные значения $\mu_1 \geqslant \mu_2 \geqslant \ldots \ldots \geqslant \mu_n$ и B — неотрицательно определенная матрица, то что можно сказать о собственных значениях $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$ матрицы $A+B$? Интуитивно мы ожидаем, что собственные значения матрицы A могут лишь возрасти (или остаться без изменений) при таком возмущении. Так оно и есть на самом деле.

Мы можем также ожидать, что собственные значения могут возрасти на сколь угодно много, если возмущающая матрица будет равна kB, где k принимает большие положительные значения. Но, по-видимому, немного удивит то, что это верно лишь для r собственных значений матрицы A, где r — ранг B. Мы найдем верхнюю границу для изменений $\mu_{r+1}, \ldots, \mu_n$, которые независимы от B.

Теорема 3.6.3. *Если A — вещественная симметрическая матрица с собственными значениями $\mu_1 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_n$, B — неотрицательно определенная матрица ранга r $(1 \geqslant r \geqslant n)$ и $A + B$ имеет собственные значения $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$, то*

(i) $$\lambda_i \geqslant \mu_i, \quad i = 1, 2, \ldots, n,$$

и

(ii) $$\mu_{j-r} \geqslant \lambda_j, \quad j = r + 1, \ldots, n.$$

Доказательство. Заметим, что, так как B неотрицательно определенная, то

$$\boldsymbol{x}'(A + B)\boldsymbol{x} = \boldsymbol{x}'A\boldsymbol{x} + \boldsymbol{x}'B\boldsymbol{x} \geqslant \boldsymbol{x}'A\boldsymbol{x}$$

для любого $\boldsymbol{x} \in \mathscr{R}_n$. Таким образом, если $\mathscr{L}$ — любое подпространство в $\mathscr{R}_n$, то

$$\max \boldsymbol{x}'(A + B)\boldsymbol{x} \geqslant \max \boldsymbol{x}'A\boldsymbol{x},$$

где максимум берется по всем векторам $\boldsymbol{x} \in \mathscr{L}$ с $\boldsymbol{x}'\boldsymbol{x} = 1$. В частности, если рассмотреть все подпространства $\mathscr{B}_{n-i+1}$ размерности $n - i + 1$, где $1 \leqslant i \leqslant n$, то из теоремы 3.6.1 получим

$$\lambda_i = \min_{\mathscr{B}_{n-i+1}} \max_{\substack{\boldsymbol{x} \in \mathscr{B}_{n-i+1} \\ \boldsymbol{x}'\boldsymbol{x}=1}} (A + B)\boldsymbol{x} \geqslant \min_{\mathscr{B}_{n-i+1}} \max_{\substack{\boldsymbol{x} \in \mathscr{B}_{n-i+1} \\ \boldsymbol{x}'\boldsymbol{x}=1}} \boldsymbol{x}'A\boldsymbol{x} = \mu_i.$$

Это доказывает часть (i) теоремы.

Применяя теорему 2.12.1 к квадратичной форме $\boldsymbol{x}'B\boldsymbol{x}$ и предполагая, что B имеет ненулевые собственные значения $v_1, \ldots, v_r$, находим, что

$$\langle \boldsymbol{x}, B\boldsymbol{x} \rangle = \sum_{j=1}^{r} v_j y_j^2 = \sum_{j=1}^{r} (\boldsymbol{d}_j'\boldsymbol{x})^2,$$

где $\boldsymbol{d}_j' = v_j^{1/2}$ (j-я строка матрицы X^*), $j = 1, 2, \ldots, r$ и $\boldsymbol{d}_1, \ldots$ $\ldots, \boldsymbol{d}_r$, очевидно, линейно независимы. Мы можем записать теперь

$$\boldsymbol{x}'(A + B)\boldsymbol{x} = \boldsymbol{x}'Ax + \sum_{j=1}^{r} (\boldsymbol{d}_j'\boldsymbol{x})^2.$$

Рассмотрим задачи нахождения собственного значения для A и $A + B$, обе при связях $\boldsymbol{d}_j'\boldsymbol{x} = 0$, $j = 1, 2, \ldots, r$. Очевидно, согласно написанному выше равенству, эти две задачи совпадают и имеют одинаковые собственные значения $\tilde{\mu}_1 \geqslant \ldots$ $\ldots \geqslant \tilde{\mu}_{n-r}$. Применяя теорему 3.6.2 к обеим задачам, получаем также

$$\begin{aligned} \mu_{k+r} &\leqslant \tilde{\mu}_k \leqslant \mu_k, \\ \lambda_{k+r} &\leqslant \tilde{\mu}_k \leqslant \lambda_k, \end{aligned} \qquad k = 1, 2, \ldots, n-r.$$

Следовательно, $\lambda_{k+r} \leqslant \mu_k$, и, записывая $j = r + k$, мы получаем часть (ii) теоремы. ◀

Упр. 1. Рассмотреть матрицу A из упр. 2 § 2.9. Исследовать ограниченные собственные значения при наложении одной связи для каждого из следующих ограничивающих уравнений и проверить теорему 3.6.2 в каждом случае:

$$\text{(i)}\quad x_1 = 0; \qquad \text{(ii)}\quad x_1 - x_2 = 0; \qquad \text{(iii)}\quad x_1 - x_3 = 0.$$

Упр. 2. Рассмотреть ограниченную задачу нахождения собственного значения с r связями, сформулированную в § 3.5. В обозначениях, использованных там, P — проекция на $\mathscr{B}_{n-r}$. Доказать, что матрица $P'AP$ имеет собственные значения $\tilde{\mu}_1, \ldots$ $\ldots, \tilde{\mu}_{n-1}$ вместе с r нулевыми собственными значениями.

Упр. 3. Рассмотреть предположение: «Если $A, B \in \mathscr{C}_{n \times n}$ и B положительно определенная, то собственные значения $\mu_1, \ldots, \mu_n$ матрицы A и $\lambda_1, \ldots, \lambda_n$ матрицы $A + B$ можно перенумеровать так, что $\operatorname{Re}(\lambda_j) \geqslant \operatorname{Re}(\mu_j)$, $j = 1, \ldots, n$». Опровергнуть это предположение, рассмотрев

$$A = \begin{Vmatrix} 6 & 1 \\ -36 & -6 \end{Vmatrix}, \qquad B = \begin{Vmatrix} 2 & 0 \\ 0 & 1 \end{Vmatrix}.$$

3.7. Приложения к теории малых колебаний

Мы возвратимся теперь к задаче, обсуждавшейся в § 2.15. В этом параграфе физическая задача о колебании была сведена к задаче решения матричного дифференциального уравнения (2.15.2):

$$A\ddot{\mathbf{p}} + C\mathbf{p} = \mathbf{0},$$

где A положительно определенная и C либо положительно, либо неотрицательно определенная. Последняя задача была затем сведена к чисто алгебраической задаче нахождения конгруэнтного преобразования, одновременно приводящего A и C к диагональным матрицам. Это, как мы видели, в свою очередь, зависит от приведения матрицы $B = A^{1/2}CA^{-1/2}$ к диагональному виду посредством ортогонального конгруэнтного преобразования.

Собственные частоты колебания системы задаются неотрицательными корнями из корней многочлена $\det(A\lambda - C)$, которые являются также собственными значениями матрицы B. Мы знаем теперь поэтому, что наибольшая собственная частота $\omega_1 = \lambda_1^{1/2}$ задается как $\lambda_1 = \max\limits_{x \in \mathcal{R}_n} R_1(x)$, где R_1 — отношение Релея для B. Теоремы 3.4.1 и 3.6.1 означают, что все собственные частоты могут быть определены в терминах экстремальных значений отношения R_1. Заметим, однако, что если записать $x = A^{1/2}q$, то

$$R_1(x) = \frac{x'A^{-1/2}CA^{-1/2}x}{x'x} = \frac{q'Cq}{q'Aq}.$$

Отношение Релея, соответствующее рассматриваемой задаче, совпадает поэтому с

$$R(q) = \frac{q'Cq}{q'Aq}. \tag{3.7.1}$$

Заметим также, что если x_j — собственный вектор матрицы B, то $q_j = A^{-1/2}x_j$ — вектор нормального вида для задачи колебания и, как в равенствах (2.15.3),

$$q_j'Aq_k = \delta_{jk}, \qquad q_j'Cq_k = \omega_j^2\delta_{jk}.$$

Применяя каждую теорему этой главы к B, без труда убеждаемся в том, что получаются обобщения, применимые к корням многочлена $\det(A\lambda - C)$. Как пример мы сформулируем лишь обобщенный вид теоремы Куранта — Фишера.

Теорема 3.7.1. *Для данных любых линейно независимых векторов $a_1, a_2, \ldots, a_r$ в $\mathcal{R}_n$ пусть $\mathcal{B}_{n-r}$ — подпространство, состоящее из всех векторов в $\mathcal{R}_n$, которые ортогональны к $a_1, \ldots, a_r$. Тогда*

$$\lambda_{r+1} = \min_{\mathcal{B}_{n-r}} \max_{x \in \mathcal{B}_{n-r}} \frac{x'Cx}{x'Ax}.$$

Минимум достигается при $a_j = q_j$, $j = 1, 2, \ldots, r$.

Физические интерпретации теорем 3.6.2 и 3.6.3 более очевидны в рамках данного изложения. Добавление одной связи может соответствовать закреплению некоторой точки системы, так что в этой точке колебания нет. Возмущение, рассмотренное в теореме 3.6.3, должно соответствовать увеличению жесткости системы. Теорема утверждает, что это может сказаться лишь в увеличении собственных частот.

Наконец, следует заметить, что все теоремы этой главы и результаты, о которых говорилось в этом параграфе, могут быть расширены заменой предположения «вещественная симметрическая» всюду на «эрмитова».

Смешанные упражнения

1. Сформулировать и доказать обобщения теорем 3.2.1, 3.3.1 и 3.4.1 на эрмитовы матрицы. (Обобщение теоремы 3.3.1 требует некоторого знания теории функций комплексной переменной.)

2. Если A, $C \in \mathscr{C}_{n\times n}$ и A неособая, то предложить подходящее обобщение определения (3.7.1), которое будет включать (3.3.2) как частный случай. Исследовать вопрос существования стационарных свойств у данного обобщения, как намечено в § 3.3.

3. Доказать теорему 2.15.2 при более слабом предположении, что A положительно определенная и C вещественная и симметрическая. Доказать, что теорема 3.7.1 также выполняется при этих предположениях. (Заметим, что, полагая $A = I$ в теореме 3.7.1, сводим результат к теореме 3.6.1.)

4. Предположим, что A — эрмитова матрица и по заданному вещественному числу ν_0 и заданному вектору $\xi \in \mathscr{C}_n$ мы строим последовательности $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots$ и $\nu_0, \nu_1, \ldots$ при помощи формул

$$\left.\begin{aligned} \boldsymbol{x}_r &= (I\nu_r - A)^{-1}\xi, \\ \nu_{r+1} &= R(\boldsymbol{x}_r), \end{aligned}\right. \qquad r = 0,\ 1,\ 2,\ \ldots$$

Доказать, что если μ_j — собственное значение матрицы A и $\nu_r \to \mu_j$ при $r \to \infty$, то

$$\frac{\nu_{r+1} - \mu_j}{(\nu_r - \mu_j)^2} \to (\text{константа}) \text{ при } r \to \infty.$$

[Указание. Использовать спектральное разложение для $(I\nu_r - A)^{-1}$; упр. 3 из § 2.5.]

5. (i) Пусть $A \in \mathscr{C}_{n\times n}$, $\mathscr{N} = \mathscr{N}(A)$, $\mathscr{R} = \mathscr{R}(A)$, $\mathscr{N}^* = \mathscr{N}(A^*)$ и $\mathscr{R}^* = \mathscr{R}(A^*)$. Доказать, что $\mathscr{R} = \mathscr{C}_n \ominus \mathscr{N}^*$ и $\mathscr{R}^* = \mathscr{C}_n \ominus \mathscr{N}$.

(ii) Для нормальной матрицы A доказать, что $\mathscr{N} = \mathscr{N}^*$.

6. В обозначениях упр. 5 доказать, что существует единственное решение $\xi \in \mathscr{R}^*$ уравнения $A\boldsymbol{x} = \boldsymbol{b} \Longleftrightarrow \boldsymbol{b} \in \mathscr{R}$.

Глава 4

МИНИМАЛЬНЫЙ МНОГОЧЛЕН И НОРМАЛЬНЫЕ ФОРМЫ

4.1. Введение

В главе 2 нас особенно интересовали те матрицы, или классы матриц, которые могут быть приведены к диагональным матрицам применением подходящего подобного преобразования. Мы описали любую такую матрицу как простую матрицу (теорема 2.4.2). В этой главе нашей основной целью будет приведение любой матрицы из $\mathscr{C}_{n\times n}$ подобными преобразованиями. Мы знаем, что, вообще говоря, не всегда некоторую матрицу из $\mathscr{C}_{n\times n}$ можно привести к диагональному виду, но какой бы простой (нормальный) вид мы ни получили, он должен превращаться в диагональный, если рассматриваемая матрица простая.

В дополнение к основным результатам мы получим несколько других важных и интересных теорем. Наш подход к рассматриваемому вопросу покажется, может быть, не прямым, но, с нашей точки зрения, он выгоден тем, что при этом получаются другие полезные результаты и требуется минимум математической абстракции. Мы подходим к задаче с алгебраической точки зрения, при которой исследуются природа характеристического многочлена и его разложение на множители в том случае, когда имеются собственные значения, у которых (алгебраические) кратности и геометрические кратности не совпадают.

Читателю следует по крайней мере отдавать себе отчет в том, что существует другой подход к задаче, который, вероятно, более популярен среди математиков и который может быть описан как геометрический. При этом подходе сначала отмечается, что собственные векторы некоторой матрицы в $\mathscr{C}_{n\times n}$ могут не порождать $\mathscr{C}_n$, и внимание сосредоточивается на способе добавления дополнительных векторов, которые вместе с собственными векторами порождают $\mathscr{C}_n$. Как только эта задача решена, нормальные формы получаются без труда.

4.2. Алгебра λ-матриц

Мы рассматриваем $n \times n$-матрицу $A(\lambda)$, элементы которой — многочлены от λ. Предполагается, что значения для λ и коэффициенты многочленов берутся из некоторого поля $\mathcal{F}$, так что если элементы матрицы A вычисляются для частного значения λ, скажем $\lambda = \lambda_0$, то $A(\lambda_0) \in \mathcal{F}_{n \times n}$. Если имеется некоторый элемент в $A(\lambda)$, который является многочленом от λ степени l, и нет элементов в $A(\lambda)$ степени, большей l, то мы будем говорить, что l — *степень* матрицы $A(\lambda)$. Матрица $A(\lambda)$ известна как *матрица многочленов*, или λ-*матрица*. Если нужно будет выделить основное поле, то мы будем называть $A(\lambda)$ λ-матрицей над $\mathcal{F}$.

Если $A(\lambda)$ имеет степень l, то (i, j)-элемент в $A(\lambda)$ может быть записан в виде

$$a_{ij}(\lambda) = a_{ij}^{(0)}\lambda^l + a_{ij}^{(1)}\lambda^{l-1} + \ldots + a_{ij}^{(l-1)}\lambda + a_{ij}^{(l)}$$

и имеется по крайней мере один элемент, для которого $a_{ij}^{(0)} \neq 0$. Если определить матрицу A_r $(r = 0, 1, \ldots, l)$ как такую, у которой (i, j)-элемент равен $a_{ij}^{(r)}$, то будем иметь $A_0 \neq 0$ и

$$A(\lambda) = A_0\lambda^l + A_1\lambda^{l-1} + \ldots + A_{l-1}\lambda + A_l.$$

λ-матрица $A(\lambda)$ называется *регулярной*, если $\det(A_0) \neq 0$. Для иллюстрации этих понятий предположим, что λ_0 — вещественное число и

$$A(\lambda) = \begin{Vmatrix} \lambda^2 + \lambda + 1 & \lambda^3 - \lambda + 2 \\ 2\lambda & \lambda^2 - 3\lambda - 1 \end{Vmatrix}.$$

Тогда $A(\lambda_0) \in \mathcal{R}_{2\times 2}$ и $A(\lambda) = A_0\lambda^3 + A_1\lambda^2 + A_2\lambda + A_3$, где

$$A_0 = \begin{Vmatrix} 0 & 1 \\ 0 & 0 \end{Vmatrix}, \quad A_1 = \begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix}, \quad A_2 = \begin{Vmatrix} 1 & -1 \\ 2 & -3 \end{Vmatrix}, \quad A_3 = \begin{Vmatrix} 1 & 2 \\ 0 & -1 \end{Vmatrix}.$$

Отметим, что $A(\lambda)$ не регулярна.

Матрицы многочленов одного и того же порядка можно складывать и перемножать между собой обычным образом и в каждом случае в результате получится другая матрица многочленов. Что можно сказать о степени суммы и произведения?

Пусть $A(\lambda)$, $B(\lambda)$ — λ-матрицы одного и того же порядка, имеющие степени l, m соответственно. Пусть $k = \max(l, m)$. Тогда (с очевидным изменением данных выше определений) можем написать

$$A(\lambda) = A_0\lambda^k + A_1\lambda^{k-1} + \ldots + A_k,$$
$$B(\lambda) = B_0\lambda^k + B_1\lambda^{k-1} + \ldots + B_k$$

и либо $A_0 \neq 0$, либо $B_0 \neq 0$. Как очевидно,

$$A(\lambda) + B(\lambda) = (A_0 + B_0)\lambda^k + (A_1 + B_1)\lambda^{k-1} + \ldots + (A_k + B_k).$$

Таким образом, степень λ-матрицы $A(\lambda)+B(\lambda)$ не превосходит k.

Предположим теперь, что

$$A(\lambda)=A_0\lambda^l+A_1\lambda^{l-1}+\ldots+A_l,$$
$$B(\lambda)=B_0\lambda^m+B_1\lambda^{m-1}+\ldots+B_m,$$

где $A_0\neq 0$ и $B_0\neq 0$. Тогда

$$A(\lambda)B(\lambda)=A_0B_0\lambda^{l+m}+(A_0B_1+A_1B_0)\lambda^{l+m-1}+\ldots+A_lB_m.$$

Таким образом, произведение будет λ-матрицей, степень которой не превосходит $l+m$. Заметим, что можем иметь $A_0B_0=0$, но если одна из $A(\lambda)$ и $B(\lambda)$ — регулярная λ-матрица, то $A_0B_0\neq 0$ и $A(\lambda)B(\lambda)$ имеет степень $l+m$.

Предположим, что $B(\lambda)$ — регулярная λ-матрица и что существуют такие λ-матрицы $Q(\lambda)$, $R(\lambda)$ с $R(\lambda)\equiv 0$ или со степенью $R(\lambda)$, меньшей m, что

$$A(\lambda)=Q(\lambda)B(\lambda)+R(\lambda).$$

В этом случае мы будем называть $Q(\lambda)$ *правым частным* $A(\lambda)$ при делении на $B(\lambda)$ и $R(\lambda)$ — *правым остатком* $A(\lambda)$ при делении на $B(\lambda)$. Подобно этому $\hat{Q}(\lambda)$, $\hat{R}(\lambda)$ — *левое частное и левый остаток* $A(\lambda)$ при делении на $B(\lambda)$, если

$$A(\lambda)=B(\lambda)\hat{Q}(\lambda)+\hat{R}(\lambda)$$

и $\hat{R}(\lambda)\equiv 0$ или степень $\hat{R}(\lambda)$ меньше m.

Если правый (левый) остаток $A(\lambda)$ при делении на $B(\lambda)$ равен нулю, то $Q(\lambda)$ ($\hat{Q}(\lambda)$) называется *правым* (*левым*) *делителем* $A(\lambda)$ при делении на $B(\lambda)$.

Чтобы придать смысл этим формальным определениям, мы должны доказать, что для данных $A(\lambda)$ и $B(\lambda)$, причем $B(\lambda)$ регулярная, обязательно существуют частные и остатки, как они были определены выше. После того как это будет сделано. мы докажем, что они также единственны. Доказательство следующей теоремы является обобщением алгоритма деления для скалярных многочленов.

Теорема 4.2.1. *Пусть $A(\lambda)$, $B(\lambda)$ — λ-матрицы степеней l, m соответственно и $B(\lambda)$ регулярная. Тогда существуют правое частное и правый остаток $A(\lambda)$ при делении на $B(\lambda)$ и подобно этому существуют левое частное и левый остаток.*

Доказательство. При $l<m$ мы должны лишь положить $Q(\lambda)=0$ и $R(\lambda)=A(\lambda)$ для получения результата.

При $l\geqslant m$ мы сначала произведем «деление» на ведущий член матрицы $B(\lambda)$, а именно $B_0\lambda^m$. Заметим, что членом наивысшей степени λ-матрицы $A_0B_0^{-1}\lambda^{l-m}B(\lambda)$ будет как раз $A_0\lambda^l$.

Следовательно,

$$A(\lambda)=A_0B_0^{-1}\lambda^{l-m}B(\lambda)+A^{(1)}(\lambda),$$

где $A^{(1)}(\lambda)$ — λ-матрица, степень которой l_1 не превосходит $l-1$. Запишем $A^{(1)}(\lambda)$ по убывающим степеням:

$$A^{(1)}(\lambda)=A_0^{(1)}\lambda^{l_1}+A_1^{(1)}\lambda^{l_1-1}+\ldots+A_{l_1}^{(1)},\qquad A_0^{(1)}\neq 0,\qquad l_1<l.$$

При $l\geqslant m$ процесс повторяем, но уже применительно к $A^{(1)}(\lambda)$ вместо $A(\lambda)$, получая в результате

$$A^{(1)}(\lambda)=A_0^{(1)}B_0^{-1}\lambda^{l_1-m}B(\lambda)+A^{(2)}(\lambda),$$

где

$$A^{(2)}(\lambda)=A_0^{(2)}\lambda^{l_2}+\ldots+A_{l_2}^{(2)},\qquad A_0^{(2)}\neq 0,\qquad l_2<l_1.$$

Таким способом мы можем построить последовательность λ-матриц $A(\lambda)$, $A^{(1)}(\lambda)$, $A^{(2)}(\lambda)$, ..., степени которых строго убывающие. После конечного числа шагов мы получим λ-матрицу $A^{(r)}(\lambda)$ степени $l_r<m$ с $l_{r-1}\geqslant m$ и (если положить $A(\lambda)=$ $=A^{(0)}(\lambda)$)

$$A^{(s-1)}(\lambda)=A_0^{(s-1)}B_0^{-1}\lambda^{l_{s-1}-m}B(\lambda)+A^{(s)}(\lambda),\qquad s=1,\ 2,\ \ldots,\ r.$$

Объединяя эти равенства, получаем

$$A(\lambda)=\left(A_0B_0^{-1}\lambda_0^{l-m}+A_0^{(1)}B_0^{-1}\lambda^{l_1-m}+\ldots\right.$$
$$\left.\ldots+A_0^{(r-1)}B_0^{-1}\lambda^{l_{r-1}-m}\right)B(\lambda)+A^{(r)}(\lambda).$$

Матрица в скобках может теперь считаться правым частным $A(\lambda)$ при делении на $B(\lambda)$ и $A^{(r)}(\lambda)$ — правым остатком.

Читатель без труда произведет изменения в доказательстве, нужные для того, чтобы убедиться в существовании левого частного и левого остатка. ◀

Теорема 4.2.2. *При предположениях теоремы* 4.2.1 *правое частное, правый остаток, левое частное и левый остаток единственны.*

Доказательство. Предположим, что существуют такие λ-матрицы $Q(\lambda)$, $R(\lambda)$ и $Q_1(\lambda)$, $R_1(\lambda)$, что

$$A(\lambda)=Q(\lambda)B(\lambda)+R(\lambda)$$

и

$$A(\lambda)=Q_1(\lambda)B(\lambda)+R_1(\lambda)$$

и из $R(\lambda)$, $R_1(\lambda)$ каждая имеет степень меньше m. Тогда

$$\{Q(\lambda)-Q_1(\lambda)\}B(\lambda)=R_1(\lambda)-R(\lambda).$$

При $Q(\lambda)\neq Q_1(\lambda)$ левая часть этого равенства будет λ-матрицей, степень которой по крайней мере m. Но в то же время правая часть — это λ-матрица степени, меньшей m. Следовательно, $Q_1(\lambda)=Q(\lambda)$, откуда также и $R_1(\lambda)=R(\lambda)$.

Подобное приведенному рассуждение может быть использовано для получения единственности левого частного и левого остатка. ◀

Упр. 1. Найти правые и левые частные и остатки $A(\lambda)$ при делении на $B(\lambda)$, где

$$A(\lambda)=\begin{Vmatrix} \lambda^4+\lambda^2+\lambda-1 & \lambda^3+\lambda^2+\lambda+2 \\ 2\lambda^3-\lambda & 2\lambda^2+2\lambda \end{Vmatrix},$$

$$B(\lambda)=\begin{Vmatrix} \lambda^2+1 & 1 \\ \lambda & \lambda^2+\lambda \end{Vmatrix}.$$

Решение. Прежде всего заметим, что $B(\lambda)$ регулярная. Находим, что

$$A(\lambda)=\begin{Vmatrix} \lambda^2-1 & \lambda-1 \\ 2\lambda & 2 \end{Vmatrix}\begin{Vmatrix} \lambda^2+1 & 1 \\ \lambda & \lambda^2+\lambda \end{Vmatrix}+\begin{Vmatrix} 2\lambda & 2\lambda+3 \\ -5\lambda & -2\lambda \end{Vmatrix}=$$
$$=Q(\lambda)B(\lambda)+R(\lambda)$$

и

$$A(\lambda)=\begin{Vmatrix} \lambda^2+1 & 1 \\ \lambda & \lambda^2+\lambda \end{Vmatrix}\begin{Vmatrix} \lambda^2 & \lambda+1 \\ \lambda-1 & 1 \end{Vmatrix}=B(\lambda)\widehat{Q}(\lambda).$$

Таким образом, $B(\lambda)$ — левый делитель $A(\lambda)$.

4.3. λ-матрицы с матричными аргументами

При рассмотрении скалярного многочлена p мы можем записать

$$p(\lambda)=a_0\lambda^l+a_1\lambda^{l-1}+\ldots+a_l=\lambda^l a_0+\lambda^{l-1}a_1+\ldots+a_l.$$

Для λ-матрицы с матричным аргументом в общем случае это невозможно. Если $A(\lambda)$ — λ-матрица над $\mathcal{F}$ и $B\in\mathcal{F}_{n\times n}$, мы определяем *правое значение* $A(B)$ матрицы $A(\lambda)$ в B как

$$A(B)=A_0B^l+A_1B^{l-1}+\ldots+A_l$$

и *левое значение* $\widehat{A}(B)$ матрицы $A(\lambda)$ в B как

$$\widehat{A}(B)=B^lA_0+B^{l-1}A_1+\ldots+A_l.$$

Читатель, по-видимому, знаком с классической теоремой об остатке, а именно, что остатком при делении скалярного многочлена $p(\lambda)$ на $\lambda-b$ будет $p(b)$. Мы докажем теперь замечательное расширение этого результата на λ-матрицы. Прежде всего заметим, что $I\lambda-B$ — регулярная λ-матрица степени 1.

Теорема 4.3.1. *Правые и левым остатками λ-матрицы $A(\lambda)$ при делении на $I\lambda-B$ являются $A(B)$ и $\widehat{A}(B)$ соответственно.*

Доказательство. Разложение на множители

$$I\lambda^l - B^l = (I\lambda^{l-1} + B\lambda^{l-2} + \ldots + B^{l-2}\lambda + B^{l-1})(I\lambda - B)$$

может быть проверено произведением умножения справа. Умножим обе части этого равенства на A_{l-j} слева и сложим полученные равенства при $j = 1, 2, \ldots, l$. Правая часть полученного равенства будет иметь вид $C(\lambda)(I\lambda - B)$, где $C(\lambda)$ — λ-матрица.

Левая часть равенства

$$\sum_{j=1}^{l} A_{l-j}\lambda^j - \sum_{j=1}^{l} A_{l-j}B^j = \sum_{j=0}^{l} A_{l-j}\lambda^j - \sum_{j=0}^{l} A_{l-j}B^j = A(\lambda) - A(B).$$

Таким образом,

$$A(\lambda) = C(\lambda)(I\lambda - B) + A(B).$$

Результат следует теперь из единственности правого остатка.

Утверждение для левого остатка получается обращением множителей в исходном разложении, умножением полученного равенства на A_{l-j} справа и суммированием. ◀

Следствие. *λ-матрица $A(\lambda)$ делится справа (слева) на $I\lambda - B$ с нулевым остатком* $\Longleftrightarrow A(B) = 0\ (\hat{A}(B) = 0)$.

Предположим теперь, что $A \in \mathscr{F}_{n\times n}$ и $c(\lambda) = \det(I\lambda - A)$ — характеристический многочлен матрицы A. Определим матрицу $B(\lambda) = (I\lambda - A)^{\vee}$. Из определения присоединенной матрицы видим, что $B(\lambda)$ — λ-матрица над $\mathscr{F}$ порядка n и степени $n-1$ и

$$(I\lambda - A)B(\lambda) = B(\lambda)(I\lambda - A) = c(\lambda)I.$$

Но $c(\lambda)I$ — λ-матрица степени n и эти равенства означают, что матрица $c(\lambda)I$ делится слева и справа на $I\lambda - A$ с нулевым остатком. Следовательно, согласно следствию, имеем:

Теорема 4.3.2. *Если A — квадратная матрица с характеристическим многочленом c, то $c(A) = 0$.*

Этот результат известен как теорема Кели — Гамильтона и иногда формулируется в виде: *каждая квадратная матрица удовлетворяет своему характеристическому уравнению.* Частный случай этого результата появлялся в упр. 9 из § 2.5, где был предложен совершенно другой подход, чем в этом параграфе. Гамильтон опубликовал теорему для 2×2-матриц в 1853 г. и Кели сформулировал общее утверждение вскоре после этого, но не дал доказательства. Первое полное доказательство было опубликовано Фробениусом в 1878 г.

Эта теорема имеет такое важное следствие: *Если f — любой скалярный многочлен с коэффициентами в поле $\mathscr{F}$ и $A \in \mathscr{F}_{n\times n}$, то существует многочлен p (зависящий от A) степени, меньшей n, для которого $f(A) = p(A)$.* Чтобы убедиться в этом, находим многочлены q, r, которые будут частным и остатком при делении f на c. Тогда

$$f(\lambda) = q(\lambda)\, c(\lambda) + r(\lambda),$$

где r — либо нулевой многочлен, либо имеет степень меньше n. Таким образом,

$$f(A) = q(A)\, c(A) + r(A).$$

Ввиду теоремы 4.3.2, получаем $f(A) = r(A)$.

Этот результат еще раз появляется в главе 5 в качестве следствия определения $f(A)$ для более общих классов скалярных функций f.

Упр. 1. Проверить теорему Кели — Гамильтона для матрицы

$$\begin{Vmatrix} 2 & 1 \\ -1 & 3 \end{Vmatrix}.$$

Упр. 2. Воспользоваться теоремой Кели — Гамильтона для вычисления $A^6 - 25A^2 + 112A$, где

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

Упр. 3. Имеется ошибка в следующем рассуждении: теорему Кели — Гамильтона можно доказать, подставляя A вместо λ в $c(\lambda) = \det(\lambda I - A)$, что дает в результате $c(A) = 0$. В чем она заключается?

Упр. 4. Теоремой Кели — Гамильтона можно воспользоваться для нахождения характеристического многочлена простой матрицы. Записывая характеристический многочлен $c(\mu) = \mu^n - c_1\mu^{n-1} + \ldots + (-1)^n c_n$, получаем

$$c_1 A^{n-1} - c_2 A^{n-2} + \ldots + (-1)^{n+1} c_n I = A^n.$$

Умножая это равенство на произвольный вектор $\boldsymbol{a} \in \mathscr{C}_n$ справа, получаем

$$T\boldsymbol{c} = A^n \boldsymbol{a},$$

где $\boldsymbol{c}' = \| c_1 \ c_2 \ \ldots \ c_n \|$ и $T \in \mathscr{C}_{n \times n}$,

$$T_{*j} = (-1)^{j-1} A^{n-j} \boldsymbol{a}, \qquad j = 1, 2, \ldots, n.$$

Таким образом, коэффициенты многочлена $c(\mu)$ находятся как решение (если оно существует) неоднородного уравнения $T\boldsymbol{c} = A^n\boldsymbol{a}$. Столбцы матрицы T вычисляются, начиная с первого, при помощи умножения $\boldsymbol{a}$ слева повторно на A. При каких условиях вектор-решение $\boldsymbol{c}$ единственный?

Упр. 5. Воспользоваться упр. 4 для нахождения характеристического многочлена матрицы

$$\begin{Vmatrix} 2 & 4 & 6 \\ 3 & 9 & 15 \\ 4 & 16 & 36 \end{Vmatrix} \quad \text{с} \quad \boldsymbol{a} = \begin{Vmatrix} 0 \\ 0 \\ 1 \end{Vmatrix}.$$

4.4. Аннулирующие многочлены

Скалярный многочлен f над F называется *аннулирующим многочленом* для квадратной матрицы $A \in \mathcal{F}_{n\times n}$, если $f(A) = 0$. В теореме 4.3.2 утверждается, что $c(A) = 0$, а поэтому всегда существует по крайней мере один аннулирующий многочлен степени n, порядка матрицы A. Мы, конечно, можем построить аннулирующие многочлены степени, большей n, но нас особенно интересуют аннулирующие многочлены наименьшей возможной степени. Чтобы избежать ненужных усложнений, мы будем рассматривать многочлены, у которых старший коэффициент (коэффициент при наивысшей степени переменной) равен 1. Такой многочлен называется *приведенным*. Мы определяем теперь приведенный многочлен ψ, являющийся аннулирующим многочленом для A наименьшей степени, как *минимальный многочлен* для A.

Теорема 4.4.1. *Каждый аннулирующий многочлен матрицы A делится без остатка на минимальный многочлен для A.*

Доказательство. Пусть f — некоторый аннулирующий многочлен матрицы A и ψ — минимальный многочлен для A. Пусть $f(\lambda) = \psi(\lambda) q(\lambda) + r(\lambda)$, где q, r — многочлены и либо r — нулевой многочлен, либо его степень меньше степени ψ. Вычисляя f в A и замечая, что $\psi(A) = 0$, получаем $r(A) = 0$. Но r не может быть аннулирующим многочленом для A, так как его степень меньше степени минимального многочлена ψ. Отсюда поэтому следует, что $r \equiv 0$, и теорема доказана.

Мы можем доказать теперь следующую теорему.

Теорема 4.4.2. *Минимальный многочлен для данной матрицы единствен.*

Доказательство. Пусть ψ_1, ψ_2 — минимальные многочлены для A. Тогда по определению ψ_1 и ψ_2 — приведенные многочлены одной и той же степени, каждый из которых делится на другой без остатка. Как легко видеть, это означает, что $\psi_1 = \psi_2$. ◀

Мы должны теперь исследовать связь между характеристическим и минимальным многочленами. Пока не было сказано ничего, что препятствовало бы этим многочленам быть совпадающими. Действительно, как мы увидим, для некоторых матриц это так и есть, но имеется также важный класс матриц, для которых степень минимального многочлена меньше степени характеристического многочлена.

Нам нужно будет воспользоваться некоторыми простыми понятиями из теории уравнений, задаваемых многочленами, эти понятия мы поэтому сначала и перечислим. Многочлен с коэффициентами в поле $\mathcal{F}$ будет обозначаться как многочлен над $\mathcal{F}$. Если $f_1, f_2, \ldots, f_k, g$ — многочлены над $\mathcal{F}$, то мы говорим, что g — *общий делитель* для $f_1, f_2, \ldots, f_k$, если g делит каждый из f без остатка. Многочлен h, который является общим делителем для $f_1, \ldots, f_k$, называется *наибольшим* общим делителем, или НОД, этих многочленов, если (i) h — приведенный многочлен над $\mathcal{F}$ и (ii) каждый общий делитель для $f_1, \ldots, f_k$ будет делителем h.

Заметим, что каждое множество многочленов над $\mathcal{F}$ имеет общий делитель 1 (многочлен степени 0). Если два многочлена над $\mathcal{F}$ имеют НОД 1, то они называются *взаимно простыми* в $\mathcal{F}$. Если f — многочлен над $\mathcal{F}$ и единственным приведенным многочленом над $\mathcal{F}$, делящим f (и не совпадающим с кратным f), является 1, то f называется *неприводимым* над $\mathcal{F}$.

Упр. 1. Проверить следующие утверждения (многочлены — над полем вещественных чисел):

(i) НОД многочленов $f(\lambda) = (\lambda + 1)(\lambda^2 + \lambda + 1)$ и $(\lambda^2 + 2)(\lambda^2 + \lambda + 1)$ равен $\lambda^2 + \lambda + 1$.

(ii) НОД многочленов $f(\lambda)$ и $(\lambda^2 - 1)(\lambda^2 + \lambda + 1)$ равен $(\lambda + 1)(\lambda^2 + \lambda + 1)$.

(iii) $\lambda + 1$ и $5\lambda^2 + 6$ взаимно просты.

Упр. 2. Пусть $f(\lambda) = \lambda^2 + \lambda + 1$. Будет ли f неприводимым, если его рассматривать как многочлен над (а) вещественными числами, (б) рациональными числами, (в) комплексными числами?

Упр. 3. Если f, g — многочлены над $\mathcal{F}$ и h — НОД для f и g, то доказать, что существуют такие многочлены s и t над $\mathcal{F}$, что для всех $x \in \mathcal{F}$ имеем $h(x) = s(x) f(x) + t(x) g(x)$.

Мы можем теперь продолжить исследование связи между характеристическим и минимальным многочленами. Рассмотрим еще раз (стр. 126) λ-матрицу $B(\lambda) = (I\lambda - A)^{\vee}$ степени $n - 1$ и порядка n. n^2 элементов в $B(\lambda)$ — это скалярные многочлены от λ над $\mathcal{F}$. Пусть $d_{n-1}(\lambda)$ — их НОД. Тогда существует λ-матрица $C(\lambda)$, известная как *приведенная присоединенная*, для которой

$$B(\lambda) = d_{n-1}(\lambda)\, C(\lambda).$$

Но $(I\lambda - A) B(\lambda) = c(\lambda) I$, так что

$$c(\lambda) I = d_{n-1}(\lambda)(I\lambda - A)\, C(\lambda).$$

Это равенство означает, что $c(\lambda)$ делится без остатка на $d_{n-1}(\lambda)$, а так как оба многочлена c и d_{n-1} приведенные, то существует такой приведенный многочлен ψ, что

$$c(\lambda) = d_{n-1}(\lambda)\psi(\lambda), \tag{4.4.1}$$

откуда

$$\psi(\lambda) I = (I\lambda - A) C(\lambda). \tag{4.4.2}$$

Таким образом, λ-матрица $\psi(\lambda) I$ делится на $I\lambda - A$ слева без остатка, а поэтому, согласно следствию теоремы 4.3.1, $\psi(A) = 0$, т. е. ψ — аннулирующий многочлен для A. Мы сейчас докажем, что ψ будет минимальным многочленом для A.

Пусть ψ^* — минимальный многочлен для A; тогда, согласно теореме 4.4.1, существует такой многочлен θ, что $\psi(\lambda) = \psi^*(\lambda)\theta(\lambda)$. Так как $\psi^*(A) = 0$, то из следствия теоремы 4.3.1 также получаем, что $\psi^*(\lambda) I$ делится на $I\lambda - A$ слева без остатка. Таким образом, существует такая λ-матрица $C^+(\lambda)$, что

$$\psi^*(\lambda) = (I\lambda - A) C^+(\lambda),$$

а следовательно, умножая на $\theta(\lambda)$, получаем

$$\psi(\lambda) I = (I\lambda - A) C^+(\lambda)\theta(\lambda).$$

Сравнивая это равенство с равенством (4.4.2), замечаем, что (по единственности левого частного, теорема 4.2.2) $C(\lambda) = C^+(\lambda)\theta(\lambda)$. Но это означало бы, что $\theta(\lambda)$ — общий делитель элементов матрицы $C(\lambda)$. Это противоречит предположению, что $d_{n-1}(\lambda)$ — НОД элементов матрицы $B(\lambda)$, если только $\theta(\lambda)$ — не константа, многочлен степени 0. Так как ψ и ψ^* — оба приведенные многочлены, то получаем, что $\theta(\lambda) = 1$ и, следовательно, что $\psi = \psi^*$. Таким образом, нами доказана

Теорема 4.4.3. *Пусть c, ψ — характеристический и минимальный многочлен матрицы A соответственно. Если $d_{n-1}(\lambda)$ — НОД элементов матрицы $(I\lambda - A)^{\vee}$, то $c(\lambda) = d_{n-1}(\lambda)\psi(\lambda)$.*

Упр. 4. Найти минимальный многочлен скалярной матрицы $A = aI$, $a \in \mathscr{F}$.

Решение. Если $A \in \mathscr{F}_{n\times n}$, то $B(\lambda) = (I\lambda - Ia)^{\vee} = (\lambda - a)^{n-1} I$. Следовательно, $d_{n-1}(\lambda) = (\lambda - a)^{n-1}$. Как очевидно, $c(\lambda) = \det(I\lambda - A) = (\lambda - a)^n$ и, следовательно, по теореме 4.4.3 $\psi(A) = \lambda - a$. Заметим, что $\psi(A) = A - aI = 0$.

Упр. 5. Сравнить характеристический и минимальный многочлены матрицы $A = \operatorname{diag}\{a, a, b, b\}$.

Упр. 6. Найти $(I\lambda - A)^{\vee}$, если

$$A = \begin{Vmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{Vmatrix},$$

и показать, что характеристический и минимальный многочлен совпадают.

***Упр. 7.** Доказать, что подобные матрицы имеют один и тот же минимальный многочлен.

Не представляет труда выяснение некоторых свойств корней минимального многочлена. Мы предположим сейчас, что $A \in \mathscr{C}_{n\times n}$, и будем рассматривать c, ψ и т. д. как многочлены над полем комплексных чисел. Таким образом, можно считать, что каждый многочлен степени k имеет k и только k корней. Беря определители в равенстве (4.4.2), находим, что

$$\psi(\lambda)^n = c(\lambda) \det C(\lambda).$$

Следовательно, каждый корень c будет также корнем ψ. Но мы имеем также $c(\lambda) = d_{n-1}(\lambda)\psi(\lambda)$, а поэтому каждый корень ψ будет корнем c. Более того, кратность корня ψ не может превзойти кратности соответствующего корня c. Результат можно сформулировать в следующем виде.

Теорема 4.4.4. *Если $A \in \mathscr{C}_{n\times n}$ и характеристический многочлен матрицы A имеет вид*

$$c(\mu) = (\mu - \mu_1)^{\alpha_1} (\mu - \mu_2)^{\alpha_2} \ldots (\mu - \mu_s)^{\alpha_s},$$

где никакие два из комплексных чисел $\mu_1, \ldots, \mu_s$ не равны, то минимальный многочлен матрицы A имеет вид

$$\psi(\mu) = (\mu - \mu_1)^{\beta_1} (\mu - \mu_2)^{\beta_2} \ldots (\mu - \mu_s)^{\beta_s},$$

где $1 \leqslant \beta_j \leqslant \alpha_j$, $j = 1, 2, \ldots, s$.

Следствие. *Если матрица $A \in \mathscr{C}_{n\times n}$ имеет n различных собственных значений, то ее характеристический и минимальный многочлены совпадают.*

Упр. 8. Если

$$A = \begin{Vmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 1 & -1 & 2 \end{Vmatrix}$$

и f — произвольный скалярный многочлен над полем комплексных чисел, то показать, что существуют такие комплексные числа α, β, зависящие от f, что $f(A) = \alpha I + \beta A$. Показать, как вычислить α, β.

Решение. Используя прием, описанный после теоремы 4.3.2, мы могли бы, конечно, выразить $f(A)$ в виде многочлена $r_1(A)$ степени 2 по A. Это достигается взятием в качестве r_1 остатка при делении f на c. Тот же самый способ может быть использован для нахождения остатка r при делении f на ψ. Если ψ имеет степень, меньшую степени c, то степень r может оказаться меньше степени r_1.

В рассматриваемом случае легко находим, что $c(\mu) = (\mu-2)^2(\mu-4)$ и

$$(\mu I - A)^{\vee} = \begin{Vmatrix} (\mu-3)(\mu-2) & -(\mu-2) & 0 \\ -(\mu-2) & (\mu-3)(\mu-2) & 0 \\ \mu-2 & -(\mu-2) & (\mu-4)(\mu-2) \end{Vmatrix}.$$

Следовательно, $d_2(\mu) = \mu - 2$ и $\psi(\mu) = (\mu-2)(\mu-4)$. Мы можем теперь записать

$$f(\mu) = q(\mu)\psi(\mu) + r(\mu),$$

где $r(\mu)$ — либо нулевой многочлен, либо имеет степень не больше 1. Тогда $f(A) = r(A)$, и первая часть поставленной задачи решена.

Коэффициенты α, β могут быть найдены без проведения процесса деления. Заметим, сначала, что собственные значения матриц $f(A)$ и $r(A)$ совпадают и (по теореме 2.5.2) равны $f(2)$, $f(4)$ и $\alpha + 2\beta$, $\alpha + 4\beta$ соответственно. Таким образом, при заданном многочлене f для нахождения α и β можно решить систему уравнений

$$\alpha + 2\beta = f(2),$$
$$\alpha + 4\beta = f(4).$$

4.5. Приведенная присоединенная матрица и минимальный многочлен

Мы покажем сейчас, как может быть вычислена приведенная присоединенная матрица, если уже известен минимальный многочлен. Пусть

$$\psi(\lambda) = \lambda^m + \alpha_1\lambda^{m-1} + \ldots + \alpha_m.$$

Тогда

$$\psi(\lambda) - \psi(\mu) = (\lambda^m - \mu^m) + \alpha_1(\lambda^{m-1} - \mu^{m-1}) + \ldots + \alpha_{m-1}(\lambda - \mu)$$

и правая часть делится на $\lambda - \mu$. Пусть $\Psi(\lambda, \mu)$ — многочлен степени $m-1$ по λ и μ, определенный как

$$\Psi(\lambda, \mu) = \frac{\psi(\lambda) - \psi(\mu)}{\lambda - \mu}.$$

Тогда имеем тождество

$$\psi(\lambda) - \psi(\mu) = (\lambda - \mu)\Psi(\lambda, \mu).$$

Далее, так как умножение $I\lambda$ на любую степень A коммутативно, $\Psi(I\lambda, A)$ определено однозначно и коммутирует с $I\lambda - A$. Следовательно, можно записать

$$\psi(I\lambda) - \psi(A) = (I\lambda - A)\Psi(I\lambda, A).$$

Но $\psi(A)=0$ и $\psi(I\lambda)=\psi(\lambda)I$, так что

$$\psi(\lambda)I=(I\lambda-A)\Psi(I\lambda,\ A).$$

Сравнивая это с равенством (4.4.2) и используя единственность частного при делении $\psi(\lambda)I$ на $I\lambda-A$, получаем, что

$$C(\lambda)=\Psi(I\lambda,\ A).$$

Упр. 1. Пусть $A\in\mathscr{C}_{n\times n}$ и имеет характеристический многочлен c. Определим

$$\Gamma(\lambda,\ \mu)=\frac{c(\lambda)-c(\mu)}{\lambda-\mu}.$$

Доказать, что если $B(\lambda)=(I\lambda-A)^{\vee}$, то $B(\lambda)=\Gamma(I\lambda, A)$.

Упр. 2. Пусть λ_j — собственное значение матрицы A. Доказать, что $C(\lambda_j)\neq 0$ и что ненулевые столбцы матрицы $C(\lambda_j)$ будут правыми собственными векторами A, соответствующими λ_j.

***Упр. 3.** Если λ_j — собственное значение матрицы A, то доказать, что ненулевые столбцы в $B(\lambda_j)$ (*если они имеются*) будут правыми собственными векторами для A, соответствующими λ_j.

Упр. 4. Показать, что две матрицы

$$A_1=\begin{Vmatrix} 6 & 2 & -2 \\ -2 & 2 & 2 \\ 2 & 2 & 2 \end{Vmatrix},\quad A_2=\begin{Vmatrix} 6 & 2 & 2 \\ -2 & 2 & 0 \\ 0 & 0 & 2 \end{Vmatrix}$$

имеют одинаковые характеристические многочлены, и найти их минимальные многочлены. Воспользоваться упр. 2 для нахождения правых собственных векторов A_1 и A_2.

Решение. Читатель легко проверит, что характеристическим многочленом, общим для A_1 и A_2, будет

$$c(\mu)=(\mu-4)^2(\mu-2)=\mu^3-10\mu^2+32\mu-32.$$

В теореме 4.4.4 утверждается, что минимальными многочленами должны быть либо $(\mu-4)^2(\mu-2)$, либо $(\mu-4)(\mu-2)$.

Мы вычислим $B_j(\lambda)=(I\lambda-A_j)^{\vee}$, $j=1, 2$, при помощи упр. 1 и, следовательно, воспользовавшись теоремой 4.4.3, получим $c_j(\lambda)$ и $\psi_j(\lambda)$. Таким образом, для обеих матриц имеем

$$\Gamma(\lambda,\ \mu)=\frac{(\lambda^3-\mu^3)-10(\lambda^2-\mu^2)+32(\lambda-\mu)}{\lambda-\mu}=$$
$$=\mu^2+(\lambda-10)\mu+(\lambda^2-10\lambda+32)$$

и

$$B_j(\lambda)=A_j^2+(\lambda-10)A_j+\left(\lambda^2-10\lambda+32\right)I,\qquad j=1,\ 2.$$

Вычисляя A_1^2 и A_2^2, подставляя их в эти равенства и складывая, находим, что

$$B_1(\lambda) = (\lambda - 4)\begin{Vmatrix} \lambda & 2 & -2 \\ -2 & \lambda - 4 & 2 \\ 2 & 2 & \lambda - 4 \end{Vmatrix},$$

$$B_2(\lambda) = \begin{Vmatrix} (\lambda - 2)^2 & 2(\lambda - 2) & 2(\lambda - 2) \\ -2(\lambda - 2) & (\lambda - 2)(\lambda - 6) & -4 \\ 0 & 0 & (\lambda - 4)^2 \end{Vmatrix}.$$

НОД элементов этих матриц равны $(\lambda - 4)$ и 1 соответственно. Следовательно, $\psi_1(\lambda) = (\lambda - 4)(\lambda - 2)$ и $\psi_2(\lambda) = (\lambda - 4)^2(\lambda - 2)$.

Для нахождения некоторых собственных векторов воспользуемся упр. 2. Имеем

$$C_1(4) = \begin{Vmatrix} 4 & 2 & -2 \\ -2 & 0 & 2 \\ 2 & 2 & 0 \end{Vmatrix}, \quad C_1(2) = \begin{Vmatrix} 2 & 2 & -2 \\ -2 & -2 & 2 \\ 2 & 2 & -2 \end{Vmatrix},$$

откуда получаем два линейно независимых правых собственных вектора, соответствующих собственному значению 4, например

$$\begin{Vmatrix} 1 \\ 0 \\ 1 \end{Vmatrix} \quad \text{и} \quad \begin{Vmatrix} -1 \\ 1 \\ 0 \end{Vmatrix},$$

и правый собственный вектор

$$\begin{Vmatrix} 1 \\ -1 \\ 1 \end{Vmatrix},$$

соответствующий собственному значению 2. Заметим, что A_1 — простая матрица.

Находим также, что $C_2(4)$ дает лишь один собственный вектор

$$\begin{Vmatrix} -1 \\ 1 \\ 0 \end{Vmatrix}$$

для собственного значения 4 и $C_2(2)$ дает собственный вектор

$$\begin{Vmatrix} 0 \\ -1 \\ 1 \end{Vmatrix}$$

для собственного значения 2.

4.6. Элементарные операции и эквивалентность λ-матриц

Как было отмечено в § 4.1, нашей целью в этой главе является приведение произвольной матрицы $A \in \mathscr{C}_{n \times n}$ к некоторому простому виду при помощи подобных преобразований.

Прежде чем мы сможем приступить к исследованию этого вопроса, требуется получить дополнительную информацию о разложении на множители минимального многочлена матрицы A. Эта промежуточная стадия состоит в общем приведении матрицы $I\lambda - A$ к некоторому простому виду посредством *эквивалентных преобразований*, которые мы и собираемся описать. Однако, так как почти с такой же легкостью мы можем обсудить сразу редукцию общей λ-матрицы размера $n \times n$ при помощи эквивалентных преобразований, наше внимание не ограничивается рассмотрением лишь матриц вида $I\lambda - A$. Это даст возможность получить дополнительные важные результаты относительно λ-матриц.

Мы определим сначала *правые* и *левые элементарные операции* над λ-матрицами, многочлены которых имеют коэффициенты в $\mathscr{F}$. Скобки относятся к левым элементарным операциям.

1. Умножение любого столбца (строки) на ненулевой элемент $c \in \mathscr{F}$.

2. Прибавление к любому столбцу (строке) любого другого столбца (строки), умноженного на произвольный многочлен над $\mathscr{F}$.

3. Перестановка любых двух столбцов (строк).

Читатель без труда проверит, что умножение некоторой λ-матрицы размера 5×5 на каждую из следующих матриц справа является правой элементарной операцией:

$$S_1 = \begin{Vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & c & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{Vmatrix},$$

$$S_2 = \begin{Vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & b(\lambda) & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{Vmatrix},$$

$$S_3 = \begin{Vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{Vmatrix}.$$

С помощью S_1 второй столбец умножается на c; с помощью S_2 четвертый столбец, умноженный на $b(\lambda)$, прибавляется ко второму столбцу; с помощью S_3 (перестановочная матрица) второй и четвертый столбцы меняются местами. S_1, S_2 и S_3 называются поэтому *правыми элементарными матрицами.* Для общей λ-матрицы размера $n \times n$ правые элементарные матрицы, приводящие к правым элементарным операциям над i-м и j-м столбцами, образуются так же легко и лишь их запись более громоздкая. Отметим, что правые элементарные матрицы могут быть получены произведением нужной операции над единичной матрицей I.

Для описанных выше частных случаев аналогичные левые элементарные матрицы задаются в виде

$$T_1 = S_1, \quad T_2 = S_2', \quad T_3 = S_3.$$

Однако обратим особое внимание на то, что левые элементарные операции получаются при *умножении слева* λ-матрицы на левую элементарную матрицу. И опять любая левая элементарная матрица может быть получена произведением соответствующей элементарной операции над единичной матрицей.

***Упр. 1.** Доказать, что определитель элементарной матрицы не зависит от λ.

***Упр. 2.** Доказать, что левая (правая) элементарная матрица неособая и что ее обратная будет левой (правой) элементарной матрицей.

Мы можем дать теперь формальное определение эквивалентных преобразований. Две λ-матрицы над $\mathcal{F}$, $A(\lambda)$ и $B(\lambda)$, называются *эквивалентными* или связанными *эквивалентным преобразованием*, если $B(\lambda)$ может быть получена из $A(\lambda)$ последовательностью элементарных операций. Из совпадения элементарных операций и операций, получаемых с помощью элементарных матриц, следует, что $A(\lambda)$ и $B(\lambda)$ эквивалентны в том и только в том случае, когда существуют такие правые элементарные матрицы $E_1(\lambda), \ldots, E_r(\lambda)$ и левые элементарные матрицы $F_1(\lambda), \ldots, F_s(\lambda)$ $(r, s \geqslant 1)$, что

$$B(\lambda) = F_s(\lambda) \ldots F_1(\lambda) A(\lambda) E_1(\lambda) \ldots E_r(\lambda) = Q(\lambda) A(\lambda) P(\lambda),$$

где $P(\lambda)$, $Q(\lambda)$ — λ-матрицы над $\mathcal{F}$ и

$$P(\lambda) = E_1(\lambda) \ldots E_r(\lambda), \quad Q(\lambda) = F_s(\lambda) \ldots F_1(\lambda).$$

Более того, из приводившихся выше упражнений следует, что $\det P(\lambda)$ и $\det Q(\lambda)$ ненулевые и не зависят от λ.

***Упр. 3.** Обозначая факт эквивалентности $A(\lambda)$ и $B(\lambda)$ через $A \sim B$, доказать, что:

(i) $A \sim A$;

(ii) $A \sim B$ означает $B \sim A$;

(iii) $A \sim B$ и $B \sim C$ означают $A \sim C$.

Упр. 4. Показать, что две следующие матрицы эквивалентны:

$$A(\lambda) = \begin{Vmatrix} \lambda & \lambda + 1 \\ \lambda^2 - \lambda & \lambda^2 - 1 \end{Vmatrix} \quad \text{и} \quad \begin{Vmatrix} 1 & 0 \\ 0 & 0 \end{Vmatrix}.$$

Решение. Мы приведем $A(\lambda)$ к более простой матрице последовательностью элементарных операций. Прибавляя сначала первую строку, умноженную на $-(\lambda-1)$, ко второй, получим

$$\begin{Vmatrix} 1 & 0 \\ 1-\lambda & 1 \end{Vmatrix} A(\lambda) = \begin{Vmatrix} \lambda & \lambda+1 \\ 0 & 0 \end{Vmatrix}. \tag{1}$$

Для дальнейшего упрощения прибавляем первый столбец с обратным знаком ко второму:

$$\begin{Vmatrix} \lambda & \lambda+1 \\ 0 & 0 \end{Vmatrix} \begin{Vmatrix} 1 & -1 \\ 0 & 1 \end{Vmatrix} = \begin{Vmatrix} \lambda & 1 \\ 0 & 0 \end{Vmatrix}. \tag{2}$$

Прибавляя теперь второй столбец, умноженный на $-\lambda$, к первому и переставляя, наконец, столбцы, получаем

$$\begin{Vmatrix} \lambda & 1 \\ 0 & 0 \end{Vmatrix} \begin{Vmatrix} 1 & 0 \\ -\lambda & 1 \end{Vmatrix} \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix} = \begin{Vmatrix} 1 & 0 \\ 0 & 0 \end{Vmatrix}. \tag{3}$$

Объединяя (1), (2) и (3), находим, что

$$Q(\lambda) A(\lambda) P(\lambda) = \begin{Vmatrix} 1 & 0 \\ 0 & 0 \end{Vmatrix},$$

где

$$Q(\lambda) = \begin{Vmatrix} 1 & 0 \\ 1-\lambda & 1 \end{Vmatrix}$$

и

$$P(\lambda) = \begin{Vmatrix} 1 & -1 \\ 0 & 1 \end{Vmatrix} \begin{Vmatrix} 1 & 0 \\ -\lambda & 1 \end{Vmatrix} \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix} = \begin{Vmatrix} -1 & 1+\lambda \\ 1 & -\lambda \end{Vmatrix}.$$

Отметим, что $\det Q(\lambda) = 1$, $\det P(\lambda) = -1$.

Упр. 5. Показать, что следующие пары матриц эквивалентны:

(а) $$\begin{Vmatrix} 0 & 1 & \lambda \\ \lambda & \lambda & 1 \\ \lambda^2-\lambda & \lambda^2-1 & \lambda^2-1 \end{Vmatrix} \text{ и } \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{Vmatrix};$$

(б) $$\begin{Vmatrix} \lambda^2 & \lambda+1 \\ \lambda-1 & \lambda^2 \end{Vmatrix} \text{ и } \begin{Vmatrix} 1 & 0 \\ 0 & \lambda^4-\lambda^2+1 \end{Vmatrix}.$$

4.7. Приведение λ-матриц эквивалентными преобразованиями к простейшему виду

В этом параграфе мы показываем, что λ-матрицу можно привести к простой диагональной λ-матрице последовательностью элементарных операций. Диагональная λ-матрица, обладающая свойствами, описанными в лемме, называется *канонической диагональной матрицей.*

Лемма. *Если $A(\lambda)$ — λ-матрица над $\mathcal{F}$ порядка n, то $A(\lambda)$ эквивалентна диагональной матрице* $\operatorname{diag}\{a_1(\lambda), \ldots, a_s(\lambda), 0, \ldots$

$\ldots, 0\}$, *где* $s \leqslant n$ *и* $a_1(\lambda), \ldots, a_s(\lambda)$ — *приведенные многочлены над* $\mathcal{F}$ *и* $a_j(\lambda)$ *делится на* $a_{j-1}(\lambda)$, $j = 2, 3, \ldots, s$.

Доказательство. Можно предположить, что $A(\lambda) \neq 0$, ибо если это не так, то доказывать нечего. Доказательство состоит лишь в описании последовательности элементарных преобразований, нужных для последовательного приведения строк и столбцов матрицы $A(\lambda)$ к требуемому виду.

Шаг 1. Пусть $a_{ij}(\lambda)$ — элемент матрицы $A(\lambda)$ наименьшей степени. Перестановкой строк и столбцов (элементарные операции типа 3) переводим этот элемент в положение $(1, 1)$ и обозначаем его через $a_{11}(\lambda)$. Для каждого элемента первых строки и столбца полученной матрицы находим частное и остаток при делении на $a_{11}(\lambda)$:

$$a_{1i}(\lambda) = a_{11}(\lambda)\, q_{1i}(\lambda) + r_{1i}(\lambda), \qquad i = 2, 3, \ldots, n,$$
$$a_{j1}(\lambda) = a_{11}(\lambda)\, q_{j1}(\lambda) + r_{j1}(\lambda), \qquad j = 1, 2, \ldots, n.$$

Далее, для каждого i и j вычтем из i-го столбца умноженный на $q_{1i}(\lambda)$ первый столбец и умноженную на $q_{j1}(\lambda)$ первую строку вычтем из j-й строки (элементарные операции типа 2). Тогда элементы $a_{1i}(\lambda)$, $a_{j1}(\lambda)$ заменятся на $r_{1i}(\lambda)$ и $r_{j1}(\lambda)$ соответственно $(i, j = 2, 3, \ldots, n)$, каждый из которых либо есть нулевой многочлен, либо имеет степень, меньшую степени $a_{11}(\lambda)$. Если эти многочлены не все нулевые, то используем операцию типа 3 для замены $a_{11}(\lambda)$ на элемент $r_{1i}(\lambda)$ или $r_{j1}(\lambda)$ меньшей степени. Далее повторяем процесс приведения степени внедиагональных элементов первых строки и столбца к меньшей, чем степень нового $a_{11}(\lambda)$. Так как степень $a_{11}(\lambda)$ строго убывает при каждом шаге, то, очевидно, в конечном счете мы приведем λ-матрицу к виду

$$\begin{Vmatrix} a_{11}(\lambda) & 0 & \ldots & 0 \\ 0 & a_{22}(\lambda) & \ldots & a_{2n}(\lambda) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & a_{n2}(\lambda) & \ldots & a_{nn}(\lambda) \end{Vmatrix}. \tag{4.7.1}$$

Шаг 2. Для матрицы вида (4.7.1) ненулевыми элементами, степень которых меньше степени $a_{11}(\lambda)$, могут быть теперь $a_{ij}(\lambda)$, $2 \leqslant i, j \leqslant n$. Если это так, то повторяем шаг 1 и получаем другую матрицу вида (4.7.1), но со степенью $a_{11}(\lambda)$, еще меньшей. Таким образом, повторяя шаг 1 достаточное число раз, мы можем найти матрицу вида (4.7.1), которая эквивалентна $A(\lambda)$ и для которой $a_{11}(\lambda)$ — ненулевой элемент наименьшей степени.

Шаг 3. Завершив шаг 2, мы ставим вопрос: существуют ли ненулевые элементы, которые не делятся на $a_{11}(\lambda)$? Если такой один имеется, например $a_{ij}(\lambda)$, то прибавим столбец j к столбцу 1; находя остатки и частные нового столбца 1 при делении на $a_{11}(\lambda)$, мы продолжим повторение шагов 1 и 2 и закончим опять матрицей вида (4.7.1), где $a_{11}(\lambda)$ заменен на многочлен меньшей степени

Этот процесс опять может продолжаться лишь конечное число шагов, после чего мы получим матрицу вида

$$\begin{Vmatrix} a_1(\lambda) & 0 & \dots & 0 \\ 0 & b_{22}(\lambda) & \dots & b_{2n}(\lambda) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & b_{n2}(\lambda) & \dots & b_{nn}(\lambda) \end{Vmatrix},$$

где после операции типа 1 (если необходимо) $a_1(\lambda)$ — приведенный многочлен и все ненулевые элементы $b_{ij}(\lambda)$ делятся на $a_1(\lambda)$ без остатка.

Шаг 4. Если все $b_{ij}(\lambda)$ нулевые, то лемма доказана. Если нет, то написанная выше матрица может быть приведена к виду

$$\begin{Vmatrix} a_1(\lambda) & 0 & 0 & \dots & 0 \\ 0 & a_2(\lambda) & 0 & \dots & 0 \\ 0 & 0 & c_{33}(\lambda) & \dots & c_{3n}(\lambda) \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & c_{n3}(\lambda) & \dots & c_{nn}(\lambda) \end{Vmatrix},$$

где $a_2(\lambda)$ делится на $a_1(\lambda)$ и элементы $c_{ij}(\lambda)$, $3 \leqslant i, j \leqslant n$, делятся на $a_2(\lambda)$. Продолжая этот процесс, получаем утверждение леммы. ◀

Упр. 1. Доказать, что, используя лишь левые (правые) элементарные операции, λ-матрицу можно привести к верхней (нижней) треугольной λ-матрице $B(\lambda)$ со свойствами: если степень $b_{jj}(\lambda)$ равна d_j ($j = 1, 2, \dots, n$), то (i) $d_j = 0$ означает, что $b_{kj} = 0$ ($b_{jk} = 0$), $k = 1, 2, \dots, j-1$, и (ii) $d_j > 0$ означает, что степень ненулевого элемента $b_{kj}(\lambda)$ ($b_{jk}(\lambda)$), $k = 1, 2, \dots$ $\dots, j-1$, меньше d_j.

Упр. 2. Воспользовавшись упр. 1, доказать, что если λ-матрица имеет ненулевой определитель, который не зависит от λ, то она может быть выражена в виде произведения конечного числа левых или правых элементарных матриц.

Упр. 3. Воспользоваться упр. 2 для доказательства того, что λ-матрицы $A(\lambda)$ и $B(\lambda)$ эквивалентны в том и только в том случае, когда существуют такие λ-матрицы $P(\lambda)$, $Q(\lambda)$, что $B(\lambda) = Q(\lambda)A(\lambda)P(\lambda)$ и $\det P(\lambda)$, $\det Q(\lambda)$ отличны от нуля и не зависят от λ.

4.8. Эквивалентные преобразования матриц из $\mathcal{F}_{n\times n}$

Описанное в § 4.7 приведение λ-матриц принимает простой вид в важном частном случае, когда $A(\lambda)$ явно совсем не зависит от λ, т. е. когда $A(\lambda)$ — λ-матрица степени 0. В этом случае элементы матрицы $A(\lambda)$ — элементы поля $\mathcal{F}$. Кроме того, в дополнение к процессу приведения используемые в элементарных операциях типа 2 множители — это элементы поля $\mathcal{F}$, а не многочлены над $\mathcal{F}$. Имея в виду это естественное ограничение для допустимых элементарных операций, матрицы $A, B \in \mathcal{F}_{n\times n}$ называем *эквивалентными*, если B может быть получена из A некоторой последовательностью элементарных операций. Таким образом, существуют такие неособые матрицы $P, Q \in \mathcal{F}_{n\times n}$, что $B = QAP$.

Лемма из § 4.7 означает теперь, что для любой матрицы $A \in \mathcal{F}_{n\times n}$ существуют такие неособые матрицы $P, Q \in \mathcal{F}_{n\times n}$, что $A = Q \operatorname{diag}\{1, \ldots, 1, 0, \ldots, 0\} P$. Читатель легко проверит, что число единиц в диагональной канонической форме равно в точности рангу матрицы A. Таким образом,

Теорема 4.8.1. *Матрица $A \in \mathcal{F}_{n\times n}$ эквивалентна $n\times n$-матрице* $\operatorname{diag}\{1, \ldots, 1, 0, \ldots, 0\}$, *где число единиц на диагонали равно рангу матрицы A.*

Упр. 1. Воспользовавшись упр. 1 из § 4.7, показать, что матрица $A \in \mathcal{F}_{n\times n}$ может быть приведена к верхнему (нижнему) треугольному виду при помощи лишь левых (правых) элементарных операций.

4.9. Инвариантные многочлены и каноническая форма Смита

Возвратимся к изучению λ-матриц и, в частности, к результату леммы из § 4.7. Наша ближайшая цель — показать, что многочлены $a_1(\lambda), \ldots, a_s(\lambda)$ однозначно определяются λ-матрицей $A(\lambda)$ и не зависят от выбора эквивалентных преобразований и что эти многочлены общие для класса всех λ-матриц, эквивалентных $A(\lambda)$.

Определим сначала *ранг* λ-матрицы как порядок наибольшего минора, который не равен нулевому многочлену. Предположим, что λ-матрица $A(\lambda)$ размера $n\times n$ имеет ранг r, и пусть $d_j(\lambda)$ — НОД всех миноров матрицы $A(\lambda)$ порядка j, $j = 1, 2, \ldots, r$. (Это согласуется с определением $d_{n-1}(\lambda)$, использованным в теореме 4.4.3) Очевидно, любой минор порядка $j \geqslant 2$ может быть выражен в виде линейной комбинации миноров порядка $j-1$, так что $d_{j-1}(\lambda)$ будет обязательно делителем $d_j(\lambda)$. Если определить $d_0(\lambda) \equiv 1$, то в последовательности

$$d_0(\lambda),\ d_1(\lambda),\ \ldots,\ d_r(\lambda)$$

$d_j(\lambda)$ делится на $d_{j-1}(\lambda)$, $j = 1, 2, \ldots, r$. Запишем соответствующие частные следующим образом:

$$i_r(\lambda) = \frac{d_r(\lambda)}{d_{r-1}(\lambda)}, \quad i_{r-1}(\lambda) = \frac{d_{r-1}(\lambda)}{d_{r-2}(\lambda)}, \ \ldots, \ i_1(\lambda) = \frac{d_1(\lambda)}{d_0(\lambda)}.$$

Многочлены $i_1(\lambda), \ldots, i_r(\lambda)$ называется *инвариантными многочленами* матрицы $A(\lambda)$. Употребление слова «инвариантные» будет оправдано, если мы покажем, что эти многочлены инвариантны относительно эквивалентных преобразований матрицы $A(\lambda)$. Пусть $B(\lambda)$ — произвольная λ-матрица, эквивалентная $A(\lambda)$. Тогда существуют такие λ-матрицы $P(\lambda)$, $Q(\lambda)$, что $B(\lambda) = Q(\lambda) A(\lambda) P(\lambda)$ и $\det P(\lambda)$, $\det Q(\lambda)$ — ненулевые константы. Последние условия означают, что $P^{-1}(\lambda)$ и $Q^{-1}(\lambda)$ — также λ-матрицы.

Мы можем применить к равенству $B = QAP$ дважды формулу Бине — Коши для выражения некоторого минора матрицы $B(\lambda)$ порядка j в виде линейной комбинации миноров матрицы $A(\lambda)$ того же самого порядка. Отсюда поэтому следует, что ранг $r(B)$ матрицы $B(\lambda)$ не превосходит $r(A)$. Однако, применяя то же самое рассуждение к λ-матричному равенству $A = Q^{-1}BP^{-1}$, находим, что $r(A) \leqslant r(B)$. Следовательно, $r(B) = r(A)$.

То же самое разложение равенства $B = QAP$ показывает, что $\delta_p(\lambda)$, НОД миноров матрицы B порядка p, делятся на $d_p(\lambda)$. Но опять равенство $A = Q^{-1}BP^{-1}$ означает, что $d_p(\lambda)$ делится на $\delta_p(\lambda)$, а следовательно,

$$\delta_p(\lambda) = d_p(\lambda), \qquad p = 1, 2, \ldots, r.$$

Поэтому $B(\lambda)$ и $A(\lambda)$ имеют одни и те же инвариантные многочлены.

Применим этот результат к канонической диагональной форме, полученной в лемме из § 4.7. Ясно, что

$$d_1(\lambda) = a_1(\lambda), \quad d_2(\lambda) = a_1(\lambda) a_2(\lambda), \ \ldots, \ d_r(\lambda) = a_1(\lambda) a_2(\lambda) \ldots a_r(\lambda)$$

и, следовательно, что

$$a_p(\lambda) = \frac{d_p(\lambda)}{d_{p-1}(\lambda)} = i_p(\lambda), \qquad p = 1, 2, \ldots, r.$$

Каноническая диагональная форма $\operatorname{diag}\{i_1(\lambda), \ldots, i_r(\lambda), 0, \ldots, 0\}$ известна как *каноническая форма Смита*. Г. Дж. С. Смитом этот вид был получен для матриц целых чисел (которые имеют почти ту же самую алгебраическую структуру, что и λ-матрицы) в 1861 г. Этот результат для λ-матриц был получен Фробениусом в 1878 г. Нами доказана

Теорема 4.9.1. *λ-матрица $A(\lambda)$ ранга r эквивалентна матрице* $\operatorname{diag}\{i_1(\lambda), \ldots, i_r(\lambda), 0, \ldots, 0\}$, *где* $i_1(\lambda), \ldots, i_r(\lambda)$ — *инвариантные многочлены для* $A(\lambda)$.

Следствие. *Две λ-матрицы эквивалентны $\Longleftrightarrow$ они имеют одинаковые инвариантные многочлены.*

Для получения следствия заметим, что мы уже доказали, что эквивалентные λ-матрицы имеют одинаковые инвариантные многочлены. Наоборот, если две λ-матрицы имеют одинаковые инвариантные многочлены, то теорема означает, что они имеют одинаковые канонические формы Смита. Свойство транзитивности отношений эквивалентности (§ 4.6, упр. 1 (iii)) означает тогда, что λ-матрицы эквивалентны.

Наконец, следует заметить, что если в утверждении теоремы элементы матрицы $A(\lambda)$ — это многочлены над $\mathcal{F}$, то инвариантные многочлены имеют коэффициенты в том же самом поле $\mathcal{F}$.

Упр. 1. Найти (из их определения) инвариантные многочлены следующих матриц:

$$\text{(а)}\quad \begin{Vmatrix} 0 & 1 & \lambda \\ \lambda & \lambda & 1 \\ \lambda^2-\lambda & \lambda^2-1 & \lambda^2-1 \end{Vmatrix}, \qquad \text{(б)}\quad \begin{Vmatrix} \lambda & \lambda^2 & 0 \\ \lambda^3 & \lambda^5 & 0 \\ 0 & 0 & 2\lambda \end{Vmatrix}.$$

Решение. Для матрицы (а), как легко видеть, определитель равен 0. Следовательно, ее ранг меньше 3. Для всех миноров порядка 2 составляем следующую таблицу.

Строки	Столбцы		
	1,2	1,3	2,3
1,2	$-\lambda$	$-\lambda^2$	$1-\lambda^2$
1,3	$-\lambda^2+\lambda$	$-\lambda^3+\lambda^2$	$-\lambda^3+\lambda^2+\lambda-1$
2,3	$\lambda^2-\lambda$	$\lambda^3-\lambda^2$	$\lambda^3-\lambda^2-\lambda+1$

Поэтому $d_2(\lambda)=1$ и $d_1(\lambda)=1$. Следовательно, $i_1(\lambda)=d_1(\lambda)=1$ и $i_2(\lambda)=d_2(\lambda)/d_1(\lambda)=1$ (ср. упр. 5 (а) из § 4.6).

Для матрицы (б) определитель равен $2\lambda^7-2\lambda^6$. Следовательно, $d_3(\lambda)=\lambda^7-\lambda^6$ (приведенный многочлен по определению НОД). Используя введенную выше таблицу для миноров порядка 2, записываем в этом случае:

Строки	Столбцы		
	1,2	1,3	2,3
1,2	$\lambda^6-\lambda^5$	0	0
1,3	0	$2\lambda^2$	$2\lambda^3$
2,3	0	$2\lambda^4$	$2\lambda^6$

Поэтому $d_2(\lambda)=\lambda^2$ и $d_1(\lambda)=\lambda$. Следовательно, $i_1(\lambda)=\lambda$, $i_2(\lambda)=$ $=d_2(\lambda)/d_1(\lambda)=\lambda$ и $i_3(\lambda)=d_3(\lambda)/d_2(\lambda)=\lambda^5-\lambda^4$.

***Упр. 2.** Доказать, что минимальный многочлен матрицы $A\in\mathcal{F}_{n\times n}$ совпадает с инвариантным многочленом матрицы $\lambda I-A$ наивысшей степени.

4.10. Подобие

Этот параграф посвящен установлению связи между эквивалентными и подобными преобразованиями. Мы можем теперь воспользоваться полученными для эквивалентных преобразований λ-матриц результатами для достижения объявленной цели этой главы, а именно, приведения матриц из $\mathcal{C}_{n\times n}$ к каноническому виду посредством подобных преобразований. Следующая теорема является связывающим звеном между этими двумя направлениями, при этом мы все еще можем рассматривать матрицы с элементами из произвольного поля $\mathcal{F}$.

Теорема 4.10.1. *Матрицы A, $B\in\mathcal{F}_{n\times n}$ подобны $\Longleftrightarrow$ $I\lambda-A$ и $I\lambda-B$ имеют одинаковые инвариантные многочлены.*

Доказательство. 1. Допустим сначала, что A и B подобны. Тогда существуют такая неособая матрица $T\in\mathcal{F}_{n\times n}$, что $A=TBT^{-1}$. Отсюда получаем равенство

$$I\lambda-A=T(I\lambda-B)T^{-1},$$

которое означает, что $I\lambda-A$ и $I\lambda-B$ эквивалентны. По следствию теоремы 4.9.1 получаем теперь, что $I\lambda-A$ и $I\lambda-B$ имеют одинаковые инвариантные многочлены.

2. Наоборот, предположим, что $I\lambda-A$ и $I\lambda-B$ имеют одинаковые инвариантные многочлены. Тогда они эквивалентны и существуют такие λ-матрицы $P(\lambda)$ и $Q(\lambda)$, что

$$P(\lambda)(I\lambda-A)Q(\lambda)=I\lambda-B$$

и $\det P(\lambda)$, $\det Q(\lambda)$ не равны нулю и не зависят от λ. Таким образом, полагая $M(\lambda)=[P(\lambda)]^{-1}$, замечаем, что $M(\lambda)$ — λ-матрица и

$$M(\lambda)(I\lambda-B)=(I\lambda-A)Q(\lambda).$$

Предположим теперь, что деление $M(\lambda)$ слева на $I\lambda-A$ и $Q(\lambda)$ справа на $I\lambda-B$ приводит к равенствам

$$\begin{aligned} M(\lambda)&=(I\lambda-A)S(\lambda)+M_0,\\ Q(\lambda)&=T(\lambda)(I\lambda-B)+Q_0, \end{aligned}\tag{4.10.1}$$

где M_0, Q_0 не зависят от λ. Производя тогда подстановку в предыдущее равенство, получаем

$$\{(I\lambda-A)S(\lambda)+M_0\}(I\lambda-B)=(I\lambda-A)\{T(\lambda)(I\lambda-B)+Q_0\},$$

откуда

$$(I\lambda - A)(S(\lambda) - T(\lambda))(I\lambda - B) = (I\lambda - A)Q_0 - M_0(I\lambda - B).$$

Но это есть тождество между λ-матрицами и, так как степень λ-матрицы справа равна 1, из него следует, что $S(\lambda) \equiv T(\lambda)$; если это не так, то степень λ матрицы слева равна по крайней мере 2. Следовательно,

$$M_0(I\lambda - B) = (I\lambda - A)Q_0, \tag{4.10.2}$$

так что

$$M_0 = Q_0, \quad M_0 B = AQ_0 \text{ и } M_0 B = AM_0. \tag{4.10.3}$$

Теперь следует лишь доказать, что M_0 неособая, и доказательство будет закончено.

3. Предположим, что деление $P(\lambda)$ слева на $I\lambda - B$ дает

$$P(\lambda) = (I\lambda - B)U(\lambda) + P_0,$$

где P_0 не зависит от λ. Воспользовавшись тогда равенствами (4.10.1) и (4.10.2), получим

$$\begin{aligned} I = M(\lambda)P(\lambda) &= \{(I\lambda - A)S(\lambda) + M_0\}\{(I\lambda - B)U(\lambda) + P_0\} = \\ &= (I\lambda - A)\{S(\lambda)(I\lambda - B)U(\lambda)\} + \\ &+ (I\lambda - A)Q_0U(\lambda) + (I\lambda - A)S(\lambda)P_0 + M_0P_0 = \\ &= (I\lambda - A)\{Q(\lambda)U(\lambda) + S(\lambda)P_0\} + M_0P_0. \end{aligned}$$

Следовательно, $Q(\lambda)U(\lambda) + S(\lambda)P_0 = 0$ и $M_0P_0 = I$. Таким образом, $\det M_0 \neq 0$ и равенство (4.10.3) означает, что A и B подобны. ◀

4.11. Первая естественная нормальная форма

Пусть f — приведенный скалярный многочлен над $\mathscr{F}$:

$$f(\lambda) = \lambda^m + \alpha_1\lambda^{m-1} + \ldots + \alpha_m.$$

Мы определяем матрицу $L \in \mathscr{F}_{m\times m}$ по f равенством

$$L(f) = \left\|\begin{matrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & & 0 & 1 \\ -\alpha_m & -\alpha_{m-1} & \ldots & -\alpha_2 & -\alpha_1 \end{matrix}\right\|.$$

Эта матрица называется *сопровождающей матрицей* для f и имеет некоторые интересные свойства.

Лемма 1. *Если $L(f)$ — сопровождающая матрица приведенного многочлена f над $\mathcal{F}$ и f имеет степень m, то*

(i) *характеристический многочлен матрицы $L(f)$ равен f;*

(ii) *минимальный многочлен для $L(f)$ равен f,*

(iii) *инвариантные многочлены для $I\lambda - L(f)$ равны $1, \ldots, 1, f$.*

Доказательство. 1. Имеем

$$I\lambda - L = \begin{Vmatrix} \lambda & -1 & & \ldots & 0 \\ 0 & \lambda & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & -1 & 0 \\ 0 & 0 & \ldots & \lambda & -1 \\ \alpha_m & \alpha_{m-1} & \ldots & \alpha_2 & \lambda + \alpha_1 \end{Vmatrix}.$$

Для вычисления $\det(I\lambda - L)$ прибавляем умноженный на λ^j столбец $j+1$ к первому для $j = 1, 2, \ldots, m-1$. В результате получим

$$\det(I\lambda - L) = \det \begin{Vmatrix} 0 & -1 & & \ldots & 0 \\ 0 & \lambda & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & \lambda & -1 \\ f(\lambda) & \alpha_{m-1} & \ldots & \alpha_2 & \lambda + \alpha_1 \end{Vmatrix}.$$

Разлагая этот определитель по первому столбцу, получаем утверждение (i).

2. Алгебраическое дополнение элемента α_m в $I\lambda - L$ равно 1. Это означает, что НОД элементов в $(I\lambda - L)^\vee$ равен 1. Утверждение (ii) следует теперь из теоремы 4.4.3.

3. Заметим, что $I\lambda - L$ имеет ненулевые миноры порядков $1, 2, \ldots, m-1$, которые в то же время не зависят от λ. Следовательно, в обозначениях § 4.9 имеем $d_{m-1} = d_{m-2} = \ldots$ $\ldots d_1 = 1$, а так как $d_m = f$ согласно (i), определение $i_r(\lambda) = d_r(\lambda)/d_{r-1}(\lambda)$ приводит к утверждению (iii). ◀

Лемма 2. *Если $A \in \mathcal{F}_{n\times n}$, то сумма степеней инвариантных многочленов матрицы $I\lambda - A$ равна n.*

Доказательство. Так как $I\lambda - A$ — λ-матрица ранга n, теорема 4.9.1 означает, что существуют λ-матрицы $P(\lambda)$, $Q(\lambda)$, определители которых — ненулевые константы и

$$P(\lambda)(I\lambda - A)Q(\lambda) = \operatorname{diag}\{i_1(\lambda), i_2(\lambda), \ldots, i_n(\lambda)\}.$$

При взятии определителей левая часть становится кратной характеристическому многочлену и имеет степень n, откуда и получаем нужное утверждение. ◀

Мы теперь в состоянии доказать теорему о простом нормальном виде, к которому подобными преобразованиями может быть приведена матрица из $\mathcal{F}_{n\times n}$. Эта первая нормальная форма будет также элементом из $\mathscr{F}_{n\times n}$. Будет удобно пользоваться следующим определением. Матрица A называется *квазидиагональной*, если она может быть разбита на блоки A_{qr}, где $q, r = 1, 2, \ldots, k$, $k \geqslant 2$, A_{qq} — квадратная матрица для каждого q и $A_{qr} = 0$, как только $q \neq r$. В этом случае мы пишем

$$A = \operatorname{diag}\{A_{11}, A_{22}, \ldots, A_{kk}\}.$$

Теорема 4.11.1 (первая естественная нормальная форма). *Если $A \in \mathscr{F}_{n\times n}$ и $I\lambda - A$ имеет инвариантные многочлены $i_r(\lambda)$, $i_{r+1}(\lambda), \ldots, i_n(\lambda)$ ненулевой степени, то A подобна квазидиагональной матрице*

$$L = \operatorname{diag}\{L(i_r), L(i_{r+1}), \ldots, L(i_n)\}.$$

Доказательство. Так как инвариантные многочлены определены над $\mathscr{F}$, лемма 2 обеспечивает то, что $L \in \mathscr{F}_{n\times n}$. Мы покажем, что $I\lambda - A$ и $I\lambda - L$ имеют одинаковые инвариантные многочлены, из чего утверждение будет следовать затем по теореме 4.10.1.

Используя утверждение (iii) леммы 1 и теорему 4.9.1, убеждаемся в том, что $I\lambda - L(i_s)$ эквивалентна диагональной матрице $D_s = \operatorname{diag}\{1, 1, \ldots, 1, i_s(\lambda)\}$, $s = r, r+1, \ldots, n$. А тогда, как нетрудно видеть, $I\lambda - L$ будет эквивалентна диагональной матрице $I\lambda - D$, где $D = \operatorname{diag}\{D_r, \ldots, D_n\}$. Таким образом, существуют такие λ-матрицы $Q(\lambda)$ и $R(\lambda)$ с ненулевыми константами $\det Q(\lambda)$ и $\det R(\lambda)$, что

$$I\lambda - L = Q(\lambda)(I\lambda - D)R(\lambda).$$

Определим теперь

$$\Delta = \operatorname{diag}\{1, 1, \ldots, 1, i_r(\lambda), \ldots, i_n(\lambda)\}.$$

Элементарными операциями типа 3 (§ 4.6) можно превратить $I\lambda - D$ в $I\lambda - \Delta$. Следовательно, $I\lambda - L$ и $I\lambda - \Delta$ эквивалентны, и мы опознаем $I\lambda - \Delta$ как каноническую форму Смита для $I\lambda - L$ (теорема 4.9.1). Таким образом, $1, 1, \ldots, 1, i_r, \ldots$ $\ldots, i_n(\lambda)$ — инвариантные многочлены одновременно для $I\lambda - L$ и $I\lambda - A$. ◀

Упр. 1. Найти первые естественные нормальные формы для матриц из упр. 8 (§ 4.4) и упр. 4 (§ 4.5).

Упр. 2. Если λ — собственное значение матрицы $L(f)$, то доказать, что существует соответствующий правый собственный вектор $\boldsymbol{x}$ с $\boldsymbol{x}' = \|1\ \lambda\ \lambda^2\ \ldots\ \lambda^{m-1}\|$.

4.12. Элементарные делители над полем комплексных чисел

Мы попытаемся теперь привести первую естественную нормальную форму к еще более простому виду. Очевидно, эта задача зависит от возможности приведения сопровождающих матриц инвариантных многочленов. Здесь поле $\mathscr{F}$, над которым определены эти многочлены, начинает играть важную роль, ибо дальнейшее приведение получается при рассмотрении разложения каждого инвариантного многочлена на неприводимые множители над $\mathscr{F}$. Для простоты и потому, что это — наиболее важный случай в приложениях, мы ограничимся рассмотрением матриц (и, следовательно, инвариантных многочленов), определенных над полем комплексных чисел $\mathscr{C}$.

Мы рассмотрим сначала случай λ-матрицы $A(\lambda)$ размера $n\times n$ с определенными над $\mathscr{C}$ элементами, имеющей ранг n и инвариантные многочлены $i_1(\lambda)$, $i_2(\lambda)$, ..., $i_n(\lambda)$. Определим корни $\det A(\lambda)$ как *скрытые корни* матрицы $A(\lambda)$. Так как $\det A(\lambda)$ — многочлен над $\mathscr{C}$, можем записать

$$\det A(\lambda)=k_1\prod_{j=1}^{s}(\lambda-\lambda_j)^{m_j},$$

где $k_1\neq 0$, λ_1, λ_2, ..., λ_s — различные скрытые корни $A(\lambda)$ и $m_j\geqslant 1$ для каждого j. Из канонической формы Смита получаем, что

$$\det A(\lambda)=k_2\prod_{j=1}^{n}i_j(\lambda),$$

где $k_2\neq 0$. Так как инвариантные многочлены приведенные, отсюда следует, что $k_1=k_2$ и

$$\prod_{j=1}^{n}i_j(\lambda)=\prod_{k=1}^{s}(\lambda-\lambda_k)^{m_k}.$$

Кроме того, так как $i_j(\lambda)$ является делителем $i_{j+1}(\lambda)$ для $j=1, 2, \ldots, n-1$, то существуют такие целые числа α_{jk}, $1\leqslant j\leqslant n$ и $1\leqslant k\leqslant s$, что

$$\begin{aligned}
i_1(\lambda)&=(\lambda-\lambda_1)^{\alpha_{11}}(\lambda-\lambda_2)^{\alpha_{12}}\ldots(\lambda-\lambda_s)^{\alpha_{1s}},\\
i_2(\lambda)&=(\lambda-\lambda_1)^{\alpha_{21}}(\lambda-\lambda_2)^{\alpha_{22}}\ldots(\lambda-\lambda_s)^{\alpha_{2s}},\\
&\cdot\\
&\cdot\\
&\cdot\\
i_n(\lambda)&=(\lambda-\lambda_1)^{\alpha_{n1}}(\lambda-\lambda_2)^{\alpha_{n2}}\ldots(\lambda-\lambda_s)^{\alpha_{ns}}
\end{aligned}\tag{4.12.1}$$

и

$$\left.\begin{aligned}
0\leqslant\alpha_{1k}\leqslant\alpha_{2k}\leqslant\ldots\leqslant\alpha_{nk}\leqslant m_k,\\
\sum_{j=1}^{n}\alpha_{jk}=m_k
\end{aligned}\right\}\ k=1, 2, \ldots, s.$$

Каждый множитель $(\lambda-\lambda_k)^{\alpha_{jk}}$, входящий в разложения (4.12.1) с $\alpha_{jk}>0$, называется *элементарным делителем* $A(\lambda)$. Элементарный делитель, для которого $\alpha_{jk}=1$, называется *линейным*, в противном случае — *нелинейным*. Элементарные делители $(\lambda-\lambda_k)^{\alpha_{jk}}$ можно считать соответствующими λ_k в очевидном смысле.

Перед переходом к общему приведению матриц из $\mathscr{C}_{n\times n}$ требуется выяснить еще одну деталь. Она содержится в следующей теореме.

Теорема 4.12.1. *Если $A(\lambda)$, $B(\lambda)$ — неособые λ-матрицы с определенными над $\mathscr{C}$ элементами, то множество элементарных делителей квазидиагональной матрицы*

$$D(\lambda)=\begin{Vmatrix} A(\lambda) & 0 \\ 0 & B(\lambda)\end{Vmatrix}$$

будут объединением множеств элементарных делителей матриц $A(\lambda)$ и $B(\lambda)$.

Доказательство. Так как $\det D(\lambda)=(\det A(\lambda))\times \times(\det B(\lambda))$, то некоторый скрытый корень λ_1 для $D(\lambda)$ будет скрытым корнем либо для $A(\lambda)$, либо для $B(\lambda)$, либо для обеих матриц вместе. Предположим поэтому, что $A(\lambda)$ имеет инвариантные многочлены $i_1(\lambda),\ \ldots,\ i_p(\lambda)$, причем

$$\left.\begin{array}{l} i_1(\lambda)=(\lambda-\lambda_1)^{\alpha_1}k_1(\lambda),\\ \quad\vdots\\ i_p(\lambda)=(\lambda-\lambda_1)^{\alpha_p}k_p(\lambda)\end{array}\right\}k_r(\lambda_1)\neq 0,\quad r=1,\ 2,\ \ldots,\ p,$$

и что $B(\lambda)$ имеет инвариантные многочлены $j_1(\lambda),\ \ldots,\ j_q(\lambda)$, причем

$$\left.\begin{array}{l} j_1(\lambda)=(\lambda-\lambda_1)^{\beta_1}l_1(\lambda),\\ \quad\vdots\\ j_q(\lambda)=(\lambda-\lambda_1)^{\beta_q}l_q(\lambda)\end{array}\right\}l_s(\lambda_1)\neq 0,\quad s=1,\ 2,\ \ldots,\ q.$$

Можно предположить также, что

$$0\leqslant\alpha_1\leqslant\ldots\leqslant\alpha_p\quad\text{и}\quad 0\leqslant\beta_1\leqslant\ldots\leqslant\beta_q,$$

и затем расположить ненулевые α и β в неубывающем порядке, перенумеровав их следующим образом:

$$0<\gamma_1\leqslant\gamma_2\leqslant\ldots\leqslant\gamma_\nu.$$

Далее, очевидно, матрица $D(\lambda)$ эквивалентна матрице

$$\operatorname{diag}\{i_1(\lambda),\ \ldots,\ i_p(\lambda),\ j_1(\lambda),\ \ldots,\ j_q(\lambda)\}.$$

Переставляя строки и столбцы, находим, что $D(\lambda)$ эквивалентна матрице

$$\operatorname{diag}\{(\lambda-\lambda_1)^{\gamma_1}(*),\ (\lambda-\lambda_1)^{\gamma_2}(*),\ \ldots,\ (\lambda-\lambda_1)^{\gamma_\nu}(*),\ (*),\ (*),\ \ldots,\ (*)\},$$

где $(*)$ обозначает многочлен, взаимно простой с $\lambda-\lambda_1$. Если $D(\lambda)$ имеет размерность n, то многочлены $d_u(\lambda)$, $u=1, 2, \ldots, n$, для $D(\lambda)$ имеют вид

$$d_n(\lambda)=(\lambda-\lambda_1)^{\gamma_1+\gamma_2+\ldots+\gamma_\nu}(*),$$
$$d_{n-1}(\lambda)=(\lambda-\lambda_1)^{\gamma_1+\ldots+\gamma_{\nu-1}}(*), \ldots$$

и, следовательно, инвариантные многочлены для $D(\lambda)$ имеют вид

$$i_n(\lambda)=(\lambda-\lambda_1)^{\gamma_\nu}(*), \quad i_{n-1}(\lambda)=(\lambda-\lambda_1)^{\gamma_{\nu-1}}(*), \ldots$$

Отсюда следует, что соответствующие λ_1 элементарные делители для $A(\lambda)$ и $B(\lambda)$ будут также элементарными делителями для $D(\lambda)$. Так как проведенное рассуждение применимо к каждому отдельному скрытому корню матрицы $D(\lambda)$, теорема доказана. ◀

Упр. 1. Каковы элементарные делители λ-матрицы (б) из упр. 1 (§ 4.9)?

Упр. 2. Найти элементарные делители матрицы $I\lambda-D$, где D — диагональная матрица.

Упр. 3. Если $A(\lambda)$ — регулярная λ-матрица порядка n и степени l, то показать, что она имеет ln скрытых корней (подсчитанных с учетом их кратности). Если эти ln скрытых корней различны, то показать, что элементарные делители матрицы $A(\lambda)$ линейны. (*Кратность* скрытого корня λ_1 — это число появлений множителя $\lambda-\lambda_1$ при разложении $\det A(\lambda)$ на линейные множители.)

Упр. 4. Если λ_1 — скрытый корень кратности α для $n\times n$-матрицы $A(\lambda)$, то доказать, что соответствующие λ_1 элементарные делители все линейны $\Longleftrightarrow$ нулевое подпространство матрицы $A(\lambda_1)$ имеет размерность α.

4.13. Вторая естественная нормальная форма и жорданова нормальная форма

Лемма. *Матрицы A, $B\in\mathscr{C}_{n\times n}$ подобны $\Longleftrightarrow$ $I\lambda-A$ и $I\lambda-B$ имеют одинаковые элементарные делители.*

Доказательство. По существу лемма является переформулировкой теоремы 4.10.1. Ибо, если $I\lambda-A$, $I\lambda-B$ имеют одинаковые элементарные делители, то они должны иметь одинаковые инвариантные многочлены, и поэтому, согласно теореме 4.10.1, матрицы A и B подобны.

Наоборот, если A и B подобны, то $I\lambda - A$ и $I\lambda - B$ имеют одинаковые инвариантные многочлены и, следовательно, одинаковые элементарные делители. ◀

Теорема 4.13.1 (вторая естественная нормальная форма). *Если $e_1(\lambda), \ldots, e_p(\lambda)$ обозначают элементарные делители для $I\lambda - A$, $A \in \mathscr{C}_{n\times n}$, то A подобна квазидиагональной матрице*

$$L_1 = \operatorname{diag}\{L(e_1),\ L(e_2),\ \ldots,\ L(e_p)\}.$$

Доказательство Как мы видели (лемма 1 из § 4.11), единственным инвариантным многочленом ненулевой степени для $I\lambda - L(e_j)$ будет $e_j(\lambda)$ и, следовательно, $e_j(\lambda)$ — единственный элементарный делитель для $I\lambda - L(e_j)$. Повторное применение теоремы 4.12.1 означает теперь, что элементарными делителями для $I\lambda - L_1$ будут $e_1(\lambda), \ldots, e_p(\lambda)$ и, согласно лемме, A и L_1 подобны. ◀

Теперь мы можем очень легко получить жорданову нормальную форму. Остается лишь придать подходящий вид диагональным блокам во второй естественной нормальной форме. Предположим, что

$$e_j(\lambda) = (\lambda - \lambda_j)^{p_j}.$$

Определим *жорданову клетку*, соответствующую этому элементарному делителю, как матрицу $J_j \in \mathscr{C}_{p_j \times p_j}$, задаваемую в виде

$$\left\|\begin{array}{cccccc} \lambda_j & 1 & 0 & \ldots & & 0 \\ 0 & \lambda_j & 1 & & & \cdot \\ \cdot & & \ddots & \ddots & & \cdot \\ \cdot & & & \ddots & \ddots & \cdot \\ \cdot & & & & & 1 \\ 0 & \ldots & & & & \lambda_j \end{array}\right\|.$$

Иногда удобно записывать жорданову клетку в другом виде. Мы определяем сначала H_n как $n \times n$-матрицу с единицами в верхней «наддиагонали» и нулями во всех других местах. Таким образом, элементы матрицы H_n задаются как $h_{jk} = \delta_{j+1,k}$ для $j, k = 1, 2, \ldots, n$. Имеем тогда

$$J_j = I\lambda_j + H_{p_j}, \tag{4.13.1}$$

где предполагается, что порядок I равен p_j. Следует также заметить, что при $p_j = 1$ мы берем $J_j = \lambda_j$. В этом случае элементарный делитель линейный.

Теорема 4.13.2 (жорданова нормальная форма). *Если $A \in \mathscr{C}_{n\times n}$ и $I\lambda - A$ имеет p элементарных делителей с соответствующими жордановыми клетками $J_1, J_2, \ldots, J_p$ и $J = \operatorname{diag}\{J_1, \ldots, J_p\}$, то A и J подобны.*

Доказательство. Согласно лемме и второй естественной нормальной форме мы должны лишь доказать, что элементарные делители матриц $I\lambda - J_j$ и $I\lambda - L(e_j)$ совпадают для $j = 1, 2, \ldots, p$, причем мы уже знаем, что единственным элементарным делителем для $I\lambda - L(e_j)$ будет $e_j(\lambda)$.

Рассматривая λ-матрицу $I\lambda - J_j$, убеждаемся в том, что в ней имеются миноры порядков $1, 2, \ldots, p_j - 1$, принимающие значения ± 1. Ясно, также, что $\det(I\lambda - J_j) = (\lambda - \lambda_j)^{p_j} = e_j(\lambda)$. Отсюда следует, что $e_j(\lambda)$ — единственный элементарный делитель для $I\lambda - J_j$, $j = 1, 2, \ldots, p$. ◀

Мы можем теперь получить две новые характеризации простых матриц.

Следствие 1. *Матрица $A \in \mathscr{C}_{n\times n}$ простая $\Longleftrightarrow$ все элементарные делители λ-матрицы $I\lambda - A$ линейные.*

Доказательство. Пусть сначала $A \in \mathscr{C}_{n\times n}$ простая и $U = \operatorname{diag}\{\mu_1, \ldots, \mathscr{C}_n\}$ — диагональная матрица собственных значений для A. Тогда существует такая неособая матрица $X \in \mathscr{C}_{n\times n}$, что $A = XUX^{-1}$, откуда

$$I\lambda - A = X(I\lambda - U)X^{-1}.$$

Повторным применением теоремы 4.12.1 легко убеждаемся в том, что элементарные делители диагональной матрицы $I\lambda - U$ все линейные. Так как написанное выше равенство означает, что $I\lambda - A$ и $I\lambda - U$ эквивалентны, то элементарные делители для $I\lambda - A$ совпадают с таковыми для $I\lambda - U$, а поэтому линейны.

Наоборот, если все элементарные делители линейны, то каждая жорданова клетка (4.13.1) имеет $p_j = 1$, так что матрица J из теоремы превращается в диагональную матрицу. Матрица A поэтому подобна диагональной матрице и, следовательно, A простая. ◀

Таким образом, приведение простой матрицы к диагональному виду, описанное в теореме 2.4.2, является частным случаем общего приведения к жордановой нормальной форме.

Следствие 2. *Матрица из $\mathscr{C}_{n\times n}$ простая $\Longleftrightarrow$ ее минимальный многочлен имеет лишь простые корни.*

Доказательство. Как вы видели (упр. 2 из § 4.9), минимальный многочлен ψ матрицы $A \in \mathscr{C}_{n\times n}$ совпадает с инвариантным многочленом i_n для $I\lambda - A$ наивысшей степени. Таким образом, простота A означает, что все элементарные делители для $I\lambda - A$ линейные, и поэтому

$$\psi(\lambda) = i_n(\lambda) = \prod_{j=1}^{s}(\lambda - \mu_j),$$

где $\mu_1, \ldots, \mu_s$ — все различные собственные значения матрицы A.

Наоборот, если ψ, а следовательно, и i_n имеют лишь простые корни, то определение элементарных делителей означает, что они все линейные. Следовательно, согласно первому следствию, A простая. ◀

Упр. 1. Найти жордановы нормальные формы для двух матриц из упр. 4 (§ 4.5).

Решение.

Для $I\lambda - A_1$:

$$d_3(\lambda) = (\lambda - 2)(\lambda - 4)^2,$$
$$d_2(\lambda) = (\lambda - 4),$$
$$d_1(\lambda) = 1;$$
$$i_1(\lambda) = (\lambda - 2)^0(\lambda - 4)^0,$$
$$i_2(\lambda) = (\lambda - 2)^0(\lambda - 4)^1,$$
$$i_3(\lambda) = (\lambda - 2)^1(\lambda - 4)^1.$$

Элементарные делители таковы:

$$(\lambda - 2), \quad (\lambda - 4), \quad (\lambda - 4).$$

Жорданова нормальная форма такова:

$$\begin{Vmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{Vmatrix}.$$

Для $I\lambda - A_2$:

$$d_3(\lambda) = (\lambda - 2)(\lambda - 4)^2,$$
$$d_2(\lambda) = 1;$$
$$d_1(\lambda) = 1,$$
$$i_1(\lambda) = (\lambda - 2)^0(\lambda - 4)^0,$$
$$i_2(\lambda) = (\lambda - 2)^0(\lambda - 4)^0,$$
$$i_3(\lambda) = (\lambda - 2)^1(\lambda - 4)^2.$$

Элементарные делители таковы:

$$(\lambda - 2), \quad (\lambda - 4)^2.$$

Жорданова нормальная форма такова:

$$\begin{Vmatrix} 2 & 0 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & 4 \end{Vmatrix}.$$

Упр. 2. Если $A \in \mathscr{C}_{10\times 10}$ и $I\lambda - A$ имеет инвариантные многочлены $i_1(\lambda) = \ldots = i_7(\lambda) = 1$ и $i_8(\lambda) = \lambda + 1$, $i_9(\lambda) = \lambda^3 + 1$, $i_{10}(\lambda) = (\lambda^3 + 1)^2$, то найти первую и вторую естественную нормальную формы и жорданову нормальную форму матрицы A.

Решение. Как легко видеть, первой естественной нормальной формой будет

$$L = \operatorname{diag}\left\{-1, \quad \begin{Vmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & -2 & 0 & 0 \end{Vmatrix}\right\}.$$

Если положить $\mu_1 = -1$, $\mu_{2,3} = (1/2) \pm \left(\sqrt{3}/2\right) i$, то

$$i_8(\lambda) = (\lambda - \mu_1),$$
$$i_9(\lambda) = (\lambda - \mu_1)(\lambda - \mu_2)(\lambda - \mu_3),$$
$$i_{10}(\lambda) = (\lambda - \mu_1)^2(\lambda - \mu_2)^2(\lambda - \mu_3)^2,$$

и поэтому имеются семь элементарных делителей, четыре из которых линейные. Получаем, что

$$L_1 = \operatorname{diag}\left\{-1,\ -1,\ \begin{Vmatrix} 0 & 1 \\ -1 & -2 \end{Vmatrix},\ \mu_2,\ \begin{Vmatrix} 0 & 1 \\ -\mu_2^2 & 2\mu_2 \end{Vmatrix},\ \mu_3,\ \begin{Vmatrix} 0 & 1 \\ -\mu_3^2 & 2\mu_3 \end{Vmatrix}\right\},$$

$$J = \operatorname{diag}\left\{-1,\ -1,\ \begin{Vmatrix} -1 & 1 \\ 0 & -1 \end{Vmatrix},\ \mu_2,\ \begin{Vmatrix} \mu_2 & 1 \\ 0 & \mu_2 \end{Vmatrix},\ \mu_3,\ \begin{Vmatrix} \mu_3 & 1 \\ 0 & \mu_3 \end{Vmatrix}\right\}.$$

Упр. 3. То же самое, что в упр. 2, с

(i) $i_1(\lambda) = \ldots = i_7(\lambda) = 1$ и $i_8(\lambda) = \lambda$, $i_9(\lambda) = \lambda^2(\lambda - 2)$, $i_{10}(\lambda) = \lambda^2(\lambda - 2)^2(\lambda^2 + 1)$;

(ii) A — матрица 12×12 и $i_1(\lambda) = \ldots = i_9(\lambda) = 1$, $i_{10}(\lambda) = \lambda^2 + 1$, $i_{11}(\lambda) = \lambda^4 + 5\lambda^2 + 4$, $i_{12}(\lambda) = \lambda^6 + 6\lambda^4 + 9\lambda^2 + 4$.

Упр. 4. Доказать, что геометрическая кратность некоторого собственного значения равна числу элементарных делителей, соответствующих ему.

Смешанные упражнения

1. Если $A(\lambda)$, $B(\lambda)$, $D(\lambda)$ — λ-матрицы порядка n и $A(\lambda)$, $B(\lambda)$ регулярны и имеют степени l, m соответственно, то доказать, что степень произведения $A(\lambda)D(\lambda)B(\lambda)$ не меньше $l + m$.

2. Показать, что если $A \in \mathscr{C}_{n\times n}$ — идемпотентная матрица и $A \neq I$, $A \neq 0$, то минимальным многочленом для A будет $\lambda^2 - \lambda$ и характеристический многочлен имеет вид $(\lambda - 1)^r\lambda^s$. Доказать, что A подобна матрице

$$\begin{Vmatrix} I_r & 0 \\ 0 & 0 \end{Vmatrix}.$$

3. Если ψ — минимальный многочлен матрицы $A \in \mathscr{C}_{n\times n}$ и f — некоторый многочлен над $\mathscr{C}$, то доказать, что $f(A)$ неособая $\Longleftrightarrow$ ψ и f взаимно просты. Обобщается ли результат на многочлен над произвольным полем $\mathscr{F}$ и матрицу $A \in \mathscr{F}_{n\times n}$?

4. Пусть $A(\lambda)$ — λ-матрица размера $n \times n$ со скрытыми корнями $\lambda_1, \ldots, \lambda_n$, и предположим, что существуют такие линейно независимые векторы $\mathbf{q}_1, \ldots, \mathbf{q}_n$, что $A(\lambda_j)\mathbf{q}_j = \mathbf{0}$ для $j = 1, \ldots, n$. Если $Q = \|\mathbf{q}_1 \ldots \mathbf{q}_n\|$, $\Lambda = I\lambda - \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\}$ и $S = Q\Lambda Q^{-1}$, то доказать, что S будет правым делителем для $A(\lambda)$.

5. Если $A(\lambda) = A_0\lambda^2 + A_1\lambda + A_2$ — регулярная λ-матрица, то доказать, что $A(\lambda)\mathbf{q} = \mathbf{0}$ в том и только в том случае, когда

$$\left\{\begin{Vmatrix} 0 & A_0 \\ A_0 & A_1 \end{Vmatrix}\lambda + \begin{Vmatrix} -A_0 & 0 \\ 0 & A_2 \end{Vmatrix}\right\}\begin{Vmatrix} \lambda\mathbf{q} \\ \mathbf{q} \end{Vmatrix} = \mathbf{0}.$$

Обобщить этот результат на регулярные λ-матрицы любой степени.

6. Исследовать собственные значения матрицы H_n, определенной в § 4.13. Показать, что $H_n^n = 0$ и что единичные векторы $e_1, \ldots, e_k$ порождают $\mathcal{N}(H_n^k)$ для $k = 1, 2, \ldots, n$.

7. Если $\mathcal{R}$ и $\mathcal{N}$ обозначают образ и аннулируемое подпространство соответственно, то показать, что

$$\mathcal{R}(H_n^{j-1}) = \mathcal{N}(H_n^{n-j+1}), \quad j = 1, 2, \ldots, n.$$

8. Исследовать вид матрицы $(I\lambda^n - H_n)^k$ при $k < n$ и при $k \geqslant n$.

9. Как известно из упр. 4 (§ 4.13), если $I\lambda - A$ имеет p элементарных делителей, то существуют p линейно независимых правых собственных векторов для A. Как можно дополнить эти p собственных векторов до базиса в $\mathcal{C}_n$?

[Указание. Если $A = XJX^{-1}$, то столбцы матрицы X образуют базис требуемого вида. Предположим для простоты, что J_1 — единственная клетка в J, соответствующая собственному значению μ_1, и что J_k имеет порядок n_k. Тогда для $k \leqslant n_1$ имеем

$$X^{-1}(I\mu_1 - A)^k X = (I\mu_1 - J)' = \operatorname{diag}\{(-H_{n_1})^k, \ (I(\mu_1 - \mu_2) - H_{n_2})^k, \ \ldots\}$$

и замечаем, что существуют k и лишь k столбцов из нулей в матрице справа.]

*10. Пусть матрица $A \in \mathcal{C}_{n\times n}$ имеет различные собственные значения $\mu_1, \ldots, \mu_s$, и предположим, что максимальная степень элементарного делителя, соответствующего μ_k, равна γ_k, $k = 1, 2, \ldots, s$. Тогда γ_k называется *индексом* μ_k. Если $\mathcal{N}_k$ — аннулируемое подпространство для $(I\mu_k - A)^{\gamma_k}$, то показать, что

$$\mathcal{C}_n = \mathcal{N}_1 \oplus \mathcal{N}_2 \oplus \ldots \oplus \mathcal{N}_s.$$

11. Показать, что характеристический и минимальный многочлены для матрицы

$$\begin{Vmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 \\ -\alpha_1 & \alpha_0 & -\alpha_3 & \alpha_2 \\ -\alpha_2 & \alpha_3 & \alpha_0 & -\alpha_1 \\ -\alpha_3 & -\alpha_2 & \alpha_1 & -\alpha_0 \end{Vmatrix}$$

равны соответственно $\{\lambda^2 - 2\alpha_0\lambda + (\alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2)\}^j$ с $j = 2$ и $j = 1$.

Дополнение к главе 4

Чтобы подвести итог рассуждениям этой главы, приводящим к нормальным формам, мы выделяем ниже некоторые из основных положений. Мы рассматриваем лишь те результаты, которые нужны для получения нормальных форм, и поэтому некоторые из результатов и определений главы появляются здесь в несколько ограниченном объеме. За пояснениями и обозначениями следует обратиться к основному тексту этой главы.

$d_r(\lambda)$ — НОД миноров в $I\lambda - A$ порядка r.

Эквивалентные преобразования

Инвариантные многочлены для $I\lambda - A$:
$$i_r(\lambda) = \frac{d_r(\lambda)}{d_{r-1}(\lambda)}.$$

Каноническая форма Смита: $I\lambda - A$ эквивалентна $\operatorname{diag}\{a_1(\lambda), \ldots, a_n(\lambda)\}$.

$i_r(\lambda) = a_r(\lambda)$

$A, B \in \mathscr{F}_{n\times n}$ подобны $\Longleftrightarrow$ $I\lambda - A$ и $I\lambda - B$ имеют одинаковые инвариантные многочлены.

Если $L(f)$ — сопровождающая матрица для f, то $I\lambda - L(f)$ имеет инвариантные многочлены $1, 1, \ldots, 1, f$.

Множители инвариантных многочленов для $I\lambda - A$ приводят к элементарным делителям.

Если инвариантными многочленами для $I\lambda - A$ являются $1, 1, \ldots, 1, i_s, \ldots, i_n$ и $L = \operatorname{diag}\{L(i_s), \ldots, L(i_n)\}$, то A и L подобны.

Первая естественная нормальная форма

$A, B \in \mathscr{C}_{n\times n}$ подобны $\Longleftrightarrow$ $I\lambda - A$ и $I\lambda - B$ имеют одинаковые элементарные делители.

Если e — элементарный делитель, то единственным элементарным делителем для $I\lambda - L(e)$ будет e.

Если J_i — жорданова клетка, соответствующая элементарному делителю e_i, то единственным элементарным делителем для $I\lambda - J_i$ будет e_i

Если $I\lambda - A$ имеет элементарные делители $e_1, \ldots, e_p$ и $L_1 = \operatorname{diag}\{L(e_1), \ldots, L(e_p)\}$, то A и L_1 подобны,

Вторая естественная нормальная форма

Если $J = \operatorname{diag}\{J_1, \ldots, J_p\}$, то A и J подобны

Жорданова нормальная форма

Глава 5

ФУНКЦИИ ОТ МАТРИЦ

5.1. Введение

В этой главе мы рассматриваем матрицы из $\mathscr{C}_{n\times n}$ (включая как частный случай и матрицы из $\mathscr{R}_{n\times n}$) и возможность придать смысл выражению $f(A)$, где f — комплекснозначная функция комплексной переменной и $A\in\mathscr{C}_{n\times n}$. Мы хотели бы, чтобы определение $f(A)$ охватывало настолько широкий класс функций f, насколько это возможно. Как мы видели, вопрос решается без труда в случае, когда f — многочлен над полем комплексных чисел. Более того, если матрица A имеет минимальный многочлен ψ степени m, то для многочленов f существуют такие многочлены q и r, что

$$f(\lambda)=q(\lambda)\psi(\lambda)+r(\lambda)$$

и r — либо нулевой многочлен, либо имеет степень меньше m. Таким образом, так как $\psi(A)=0$, имеем $f(A)=r(A)$.

Более общие функции f, которые будут рассматриваться, сохранят это свойство, т. е. для данных f и A будет существовать такой многочлен r (со степенью, меньшей степени минимального многочлена для A), что $f(A)=r(A)$. Нам потребуется сначала некоторые результаты о приближениях при помощи многочленов.

5.2. Интерполяционные многочлены

Пусть $\lambda_1, \lambda_2, \ldots, \lambda_s$ — различные комплексные (или вещественные) числа и $f_1, f_2, \ldots, f_s$ — произвольное множество комплексных (или вещественных) чисел. Мы хотим определить некоторый интерполяционный многочлен степени $s-1$, принимающий значения $f_1, \ldots, f_s$ при $\lambda_1, \ldots, \lambda_s$ соответственно. Можно дать простую конструкцию для такого многочлена. Определим сначала базисные многочлены $l_1, \ldots, l_s$ степени $s-1$ как

$$l_k(\lambda) = \prod_{\substack{j=1 \\ j \neq k}}^{s} (\lambda - \lambda_j) \Big/ \prod_{\substack{j=1 \\ j \neq k}}^{s} (\lambda_k - \lambda_j), \qquad k = 1, 2, \ldots, s.$$

Как нетрудно проверить, эти многочлены обладают свойством:

$$l_k(\lambda_j) = \delta_{kj}, \qquad j, k = 1, 2, \ldots, s.$$

Тогда *многочлен Лагранжа* с нужными свойствами есть

$$L(\lambda) = \sum_{k=1}^{s} f_k l_k(\lambda).$$

Более того, L — единственный многочлен степени $s-1$ с $L(\lambda_j) = f_j$, $j = 1, 2, \ldots, s$. Ибо если существует еще L_1 с теми же самыми свойствами, то $L_1 - L$ будет многочленом, степень которого не превосходит $s-1$ и имеющим корни в различных точках $\lambda_1, \ldots, \lambda_s$. Это означает, что $L_1 - L$ — нулевой многочлен и, следовательно, $L_1 = L$.

Такова основная задача лагранжева приближения. Нам потребуется нечто более общее. Мы хотим фиксировать не только значения функции в $\lambda_1, \ldots, \lambda_s$, но также некоторое число $m_k - 1$ последовательных производных интерполяционного многочлена в λ_k для $k = 1, 2, \ldots, s$. Таким образом, имеются m_k условий, наложенных на многочлен в точке λ_k, и если $m = m_1 + m_2 + \ldots$ $\ldots + m_s$, то мы разыскиваем интерполяционный многочлен степени $m-1$. Эта задача известна как задача общей интерполяции Эрмита.

Решение этой задачи более громоздкое, чем простого случая Лагранжа, но в принципе не более трудное. Мы определяем сначала

$$t_k(\lambda) = \prod_{\substack{j=1 \\ j \neq k}}^{s} (\lambda - \lambda_j)^{m_j} \Big/ \prod_{\substack{j=1 \\ j \neq k}}^{s} (\lambda_k - \lambda_j)^{m_j}, \qquad k = 1, 2, \ldots, s,$$

и замечаем, что при $m_j = 1$ для $j = 1, 2, \ldots, s$ имеем $t_k(\lambda) = l_k(\lambda)$. При $s = 1$ следует полагать $t_1(\lambda) = 1$. Тогда $t_k(\lambda)$ принимает значение 1 в λ_k и равен нулю в каждой точке $\lambda_j \neq \lambda_k$, но производные не обладают нужными свойствами. Определим $T_{k,p}$ как многочлен степени p, получаемый срезанием ряда Тейлора для $1/t_k(\lambda)$ по степеням $\lambda - \lambda_k$. Тогда мы утверждаем, что

$$\varphi_{kj}(\lambda) = \frac{(\lambda - \lambda_k)^{j-1}}{(j-1)!} t_k(\lambda) T_{k,\, m_k - j}(\lambda),$$
$$k = 1, 2, \ldots, s, \quad j = 1, 2, \ldots, m_k, \tag{5.2.1}$$

имеют базисные интерполяционные свойства:

$$\begin{array}{lll}
\varphi_{k1}(\lambda_k)=1, & \varphi_{k1}^{(1)}(\lambda_k)=0,\ \ldots, & \varphi_{k1}^{(m_k-1)}(\lambda_k)=0,\\
\varphi_{k2}(\lambda_k)=0, & \varphi_{k2}^{(1)}(\lambda_k)=1,\ \ldots, & \varphi_{k2}^{(m_k-1)}(\lambda_k)=0,\\
\vdots & \vdots & \vdots\\
\varphi_{k,\,m_k}(\lambda_k)=0, & \varphi_{k,\,m_k}^{(1)}(\lambda_k)=0,\ \ldots, & \varphi_{k,\,m_k}^{(m_k-1)}(\lambda_k)=1
\end{array}$$

и

$$\varphi_{kj}^{(r)}(\lambda_p)=0 \text{ для } p\neq k \text{ и } r=0,\ 1,\ \ldots,\ m_p-1,$$

где мы пишем $\varphi_{kj}^{(s)}(\lambda_t)$ для s-й производной от φ_{kj} в λ_t. Заметим, что если $m=\sum_{k=1}^{s} m_k$, то φ_{kj} имеет степень $(j-1)+(m-m_k)+(m_k-j)=m-1$. Теперь мы можем выписать многочлен H степени, не большей $m-1$, который удовлетворяет условиям $H^{(r)}(\lambda_k)=f_k^{(r)}$, $k=1,\ 2,\ \ldots,\ s$ и $r=0,\ 1,\ \ldots,\ m_k-1$. Он равен

$$H(\lambda)=\sum_{k=1}^{s}\sum_{j=1}^{m_k} f_k^{(j-1)}\varphi_{kj}(\lambda). \tag{5.2.2}$$

Упр. 1. Доказать, что $t_k(\lambda)T_{k,\,p}(\lambda)-1$ — многочлен от $(\lambda-\lambda_k)$, в котором каждый член имеет степень не меньше $p+1$.

Упр. 2. Доказать базисные интерполяционные свойства многочленов φ_{kj}.

Упр. 3. Доказать, что определенный выше многочлен H единствен.

***Упр. 4.** Показать, что в случае $s=1$ равенство (5.2.2) превращается в

$$H(\lambda)=f_1+\frac{(\lambda-\lambda_1)}{1!}f_1^{(1)}+\frac{(\lambda-\lambda_1)^2}{2!}f_1^{(2)}+\ \ldots$$
$$\ldots+\frac{(\lambda-\lambda_1)^{m_1-1}}{(m_1-1)!}f_1^{(m_1-1)}.$$

5.3. Определение функции от матрицы

Предположим, что матрица $A\in\mathscr{C}_{n\times n}$ имеет различные собственные значения $\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_s$ с индексами*) $m_1,\ m_2,\ \ldots,\ m_s$, так что минимальный многочлен для A равен

$$\psi(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\ldots(\lambda-\lambda_s)^{m_s}.$$

Пусть g, h — многочлены над $\mathscr{C}$, для которых $g(A)=h(A)$, и положим $d=g-h$. Тогда $d(A)=0$, так что d — аннулирующий

---

*) См. упр. 10 на стр. 154. В обозначениях равенств (4.12.1) $m_j=\alpha_j$.

многочлен и поэтому (теорема 4.4.1) делится на ψ. Таким образом, существует такой многочлен p, что

$$g(\lambda)-h(\lambda)=\psi(\lambda)\,p(\lambda).$$

Но, очевидно, для $k=1, 2, \ldots, s$

$$\psi(\lambda_k)=\psi^{(1)}(\lambda_k)=\ldots=\psi^{(m_k-1)}(\lambda_k)=0,$$

откуда получаем, что

$$g(\lambda_k)=h(\lambda_k),\ g^{(1)}(\lambda_k)=h^{(1)}(\lambda_k),\ \ldots,\ g^{(m_k-1)}(\lambda_k)=h^{(m_k-1)}(\lambda_k).$$

В общем случае m чисел

$$f(\lambda_k),\ f^{(1)}(\lambda_k),\ \ldots,\ f^{(m_k-1)}(\lambda_k),\qquad k=1,\ 2,\ \ldots,\ s,$$

называются *значениями функции f на спектре матрицы A*. Про любую функцию f, для которой эти числа существуют, мы будем говорить, что она *определена на спектре матрицы A*. Очевидно, каждый многочлен определен на спектре любой матрицы из $\mathscr{C}_{n\times n}$ и, в частности, мы показали, что многочлены g и h имеют одинаковые значения на спектре матрицы A, если $g(A)=h(A)$.

Наоборот, если g и h — любые два многочлена с одинаковыми значениями на спектре матрицы A, то $d=g-h$ имеет корень кратности m_k в λ_k для каждого k. Следовательно, d должен делиться на ψ, и поэтому $d(A)=0$, т. е. $g(A)=h(A)$. Нами доказана

Лемма. *Если g и h — многочлены на $\mathscr{C}$ и $A\in\mathscr{C}_{n\times n}$, то $g(A)=h(A) \Longleftrightarrow g$ и h имеют одинаковые значения на спектре матрицы A.*

Именно это свойство многочленов с матричными аргументами мы используем теперь для определения $f(A)$ с более общими функциями f. Таким образом, мы будем требовать, чтобы все функции, которые определены на спектре матрицы A и принимают там одинаковые значения, приводили к одной и той же матрице $f(A)$. В частности, для любой f, определенной на спектре A, мы будем иметь возможность записать $f(A)=g(A)$, где g — некоторый многочлен. Всегда можно взять g в виде общего интерполяционного многочлена (5.2.2) Эрмита. Таким образом, можно положить

$$g(\lambda)=\sum_{k=1}^{s}\sum_{j=1}^{m_k} f_k^{(j-1)}\varphi_{kj}(\lambda), \tag{5.3.1}$$

где $f_k^{(j)}=f^{(j)}(\lambda_k)$ — j-я производная от f, вычисленная в λ_k.

С первого взгляда это требование кажется очень жестким, но, как мы увидим, оно удовлетворяет всем нашим потребностям и включает в себя второй подход к задаче с использованием степенных рядов. Таким образом, если функция f определена на спектре матрицы A, то мы *по определению полагаем* $f(A) = g(A)$, где g — произвольный многочлен, принимающий те же самые значения, что и f, на спектре A. Согласно лемме выбор многочлена g не имеет значения и, как мы знаем, выбор, определяемый равенством (5.3.1), дает многочлен наименьшей возможной степени. Следовательно, имеем:

Теорема 5.3.1. *Если f — некоторая функция, определенная на спектре матрицы A, и g — интерполяционный многочлен минимальной степени, определенный значениями функции f на спектре A, то $f(A) = g(A)$.*

Если A — простая матрица, то многочлен (5.3.1) принимает особенно простой вид. В этом случае имеем (следствие 2 теоремы 4.13.2) $m_1 = m_2 = \ldots = m_k = 1$, так что

$$g(\lambda) = \sum_{k=1}^{s} f(\lambda_k)\, l_k(\lambda),$$

где l_k — базисный многочлен лагранжева типа степени $s-1$.

Упр. 1. Вычислить $f(A)$, где $f(\lambda) = e^{\lambda t}$ и $A = \begin{Vmatrix} 6 & -1 \\ 3 & 2 \end{Vmatrix}$.

Решение. Собственные значения для A равны $\lambda_1 = 3$, $\lambda_2 = 5$, так что минимальным многочленом матрицы A будет $(\lambda - 3)(\lambda - 5)$. Мы не используем в явном виде лагранжева интерполяционного многочлена g, а, зная, что он должен быть степени 1, записываем $g(\lambda) = \alpha + \beta\lambda$ с константами α, β, которые следует определить. Как мы знаем, $f(\lambda_1) = g(\lambda_1)$ и $f(\lambda_2) = g(\lambda_2)$, так что

$$\alpha + 3\beta = e^{3t}, \quad \alpha + 5\beta = e^{5t}.$$

Решая эти уравнения относительно α и β, находим

$$\alpha = \frac{1}{2}(5e^{3t} - 3e^{5t}), \quad \beta = \frac{1}{2}(e^{5t} - e^{3t}).$$

Таким образом,

$$f(A) = e^{At} = \alpha I + \beta A = \frac{1}{2} e^{5t} \begin{Vmatrix} 3 & -1 \\ 3 & -1 \end{Vmatrix} + \frac{1}{2} e^{3t} \begin{Vmatrix} -1 & 1 \\ -3 & 3 \end{Vmatrix}.$$

***Упр. 2.** Если

$$A = \frac{\pi}{2} \begin{Vmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{Vmatrix},$$

то доказать, что

$$\sin A = \frac{4}{\pi} A - \frac{4}{\pi^2} A^2 = \begin{Vmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}.$$

***Упр. 3.** Пусть H_n — $n \times n$-матрица

$$\begin{Vmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \cdot & \cdot & & \cdot & & \\ \cdot & \cdot & & & \cdot & \\ \cdot & \cdot & & & & \cdot \\ 0 & 0 & \dots & & & 1 \\ 0 & 0 & \dots & & & 0 \end{Vmatrix}$$

и $f(\lambda)$ — некоторая функция, имеющая $n-1$ производных в $\lambda=0$. Найти $f(H_n)$.

Решение (ср. упр. 6 на стр. 154). Минимальный многочлен для H_n равен λ^n и значениями f на спектре H_n будут поэтому $f(0)$, $f^{(1)}(0)$, ..., $f^{(n-1)}(0)$. Положим $f_0^{(k)}=f^{(k)}(0)$, $k=0, 1, \dots, n-1$. Интерполяционный многочлен g степени $n-1$, определенный значениями f на спектре H_n, как нетрудно видеть, равен (упр. 4 из § 5.2)

$$g(\lambda)=f_0+\frac{1}{1!}f_0^{(1)}(\lambda)+\frac{1}{2!}f_0^{(2)}\lambda^2+\dots+\frac{1}{(n-1)!}f_0^{(n-1)}\lambda^{n-1}.$$

Имеем тогда $f(H_n)=g(H_n)$. Таким образом,

$$f(H_n)=f_0I+\frac{1}{1!}f_0^{(1)}H_n+\dots+\frac{1}{(n-1)!}f_0^{(n-1)}H_n^{n-1}.$$

Вычисляя степени H_n, находим, что

$$f(H_n)=\begin{Vmatrix} f_0 & \frac{1}{1!}f_0^{(1)} & \frac{1}{2!}f_0^{(2)} & \dots & \frac{1}{(n-1)!}f_0^{(n-1)} \\ 0 & f_0 & \frac{1}{1!}f_0^{(1)} & & \vdots \\ 0 & 0 & f_0 & & \frac{1}{1!}f_0^{(1)} \\ \vdots & & \ddots & \ddots & \\ 0 & \dots & & 0 & f_0 \end{Vmatrix}.$$

***Упр. 4.** Пусть $J=I\lambda_0+H_n$ — жорданова клетка порядка n и $f(\lambda)$ — некоторая функция, имеющая $n-1$ производных в $\lambda=\lambda_0$. Если $f_0^{(k)}$ — k-я производная от f в λ_0, $k=0, 1, \dots, n-1$, то доказать, что $f(J)$ совпадает с матрицей, выписанной в решении упр. 3.

Теорема 5.3.2. *Если A — квазидиагональная матрица*

$$A=\operatorname{diag}\{A_1, A_2, \dots, A_p\}$$

и f определена на спектре матрицы A, то

$$f(A)=\operatorname{diag}\{f(A_1), f(A_2), \dots, f(A_p)\}.$$

Доказательство. Прежде всего, ясно, что для любого многочлена h

$$h(A) = \operatorname{diag}\{h(A_1), \ldots, h(A_p)\}.$$

Следовательно, если g — интерполяционный многочлен для f на спектре A, то

$$f(A) = g(A) = \operatorname{diag}\{g(A_1), \ldots, g(A_p)\}.$$

Далее, так как спектр для A_j $(j = 1, 2, \ldots, p)$, очевидно, является подмножеством спектра для A, то f определена на спектре A_j, а так как g и f совпадают на спектре A, то они должны совпадать также на спектре A_j. Следовательно, $f(A_j) = g(A_j)$, и мы получаем

$$f(A) = \operatorname{diag}\{f(A_1), \ldots, f(A_p)\}. \blacktriangleleft$$

Теорема 5.3.3. *Если $A, B, T \in \mathscr{C}_{n \times n}$, $A = T^{-1}BT$ и f определена на спектре A, то $f(A) = T^{-1}f(B)T$.*

Доказательство. Так как A и B подобны, они имеют одинаковые минимальные многочлены (упр. 7 из § 4.4). Таким образом, если g — интерполяционный многочлен для f на спектре A, то g будет также интерполяционным многочленом для f на спектре B, и мы имеем $f(A) = g(A)$, $f(B) = g(B)$. Но, очевидно, $g(A) = T^{-1}g(B)T$ (§ 2.4), откуда следует, что $f(A) = T^{-1}f(B)T$. $\blacktriangleleft$

Если есть возможность привести матрицу $A \in \mathscr{C}_{n \times n}$ к жордановой нормальной форме и f определена на спектре A, то $f(A)$ находится затем очень легко. Предположим сначала, что матрица A простая с собственными значениями $\mu_1, \mu_2, \ldots, \mu_n$. В этом случае существует такая матрица $T \in \mathscr{C}_{n \times n}$, что

$$A = T^{-1} \operatorname{diag}\{\mu_1, \ldots, \mu_n\} T.$$

Воспользовавшись теоремами 5.3.3 и 5.3.2, немедленно убеждаемся в том, что

$$f(A) = T^{-1} \operatorname{diag}\{f(\mu_1), \ldots, f(\mu_n)\} T.$$

В общем случае мы можем лишь предположить, что A подобна матрице в жордановой нормальной форме, т. е. что существует такая матрица T, что

$$A = T^{-1} \operatorname{diag}\{J_1, \ldots, J_p\} T,$$

где $J_1, \ldots, J_p$ — жордановы клетки, соответствующие элементарным делителям λ-матрицы $I\lambda - A$ (теорема 4.13.2). В этом случае получаем

$$f(A) = T^{-1} \operatorname{diag}\{f(J_1), \ldots, f(J_p)\} T,$$

и матрицы $f(J_k)$ верхние треугольные, вид которых описан выше в упр. 4. Заметим, что если J_k соответствует собственному значению μ_k, то диагональные элементы в $f(J_k)$ все равны $f(\mu_k)$, а так как собственные значения треугольной матрицы совпадают с ее диагональными элементами, то собственные значения матрицы $f(A)$, возможно кратные, равны $f(\mu_1), \ldots, f(\mu_p)$. Верна поэтому

Теорема 5.3.4. *Если $\mu_1, \ldots, \mu_n$ — собственные значения матрицы $A \in \mathscr{C}_{n\times n}$ и f определена на спектре A, то собственными значениями матрицы $f(A)$ будут $f(\mu_1), \ldots, f(\mu_n)$.*

Приведенное выше упр. 2 служит иллюстрацией того, что элементарные делители, соответствующие некоторому выделенному собственному значению, не обязательно сохраняются при преобразовании, определяемом f. В рассмотренном случае элементарный делитель собственного значения $\pi/2$ матрицы A нелинеен, а элементарные делители собственного значения $f(\pi/2) = 1$ матрицы $f(A)$ линейны.

5.4. Спектральное разложение для $f(A)$

Следующее (спектральное) разложение для $f(A)$ имеет большое значение и, очевидно, тесно примыкает к уже полученной спектральной теореме для простых матриц (теореме 2.5.1).

Теорема 5.4.1 *Если $A \in \mathscr{C}_{n\times n}$, f определена на спектре A, $f_k^{(j)}$ — значение j-й производной от f в собственном значении λ_k $(k = 1, 2, \ldots, s,\ j = 0, 1 \ldots, m_k - 1)$ и m_k — индекс λ_k, то существуют такие независимые от f матрицы Z_{kj}, что*

$$f(A) = \sum_{k=1}^{s} \sum_{j=1}^{m_k} f_k^{(j-1)} Z_{kj}. \tag{5.4.1}$$

Кроме того, матрицы Z_{kj} линейно независимы как элементы из $\mathscr{C}_{n\times n}$ и коммутируют с A.

Доказательство. Напомним, что f и многочлен g, определяемый равенством (5.3.1), принимают одинаковые значения на спектре матрицы A. Следовательно, $f(A) = g(A)$ и равенство (5.4.1) будет выполняться, если положить

$$Z_{kj} = \varphi_{kj}(A).$$

Так как φ_{kj} определяется свойствами минимального многочлена для A, то Z_{kj} не зависит от f. Немедленно также получаем, что каждое Z_{kj} коммутирует с A.

Остается лишь доказать, что эти матрицы линейно независимы. Мы оставим читателю в качестве упражнения доказать, что многочлены φ_{kj} — линейно независимые элементы пространства $\mathscr{P}_{m-1}$ многочленов степени не выше $m-1$, где $m = m_1 + m_2 + \ldots + m_s$

Предположим, что

$$\sum_{k=1}^{s} \sum_{j=1}^{m_k} c_{kj} Z_{kj} = 0,$$

и определим $h(\lambda) = \sum_{k,j} c_{k,j} \varphi_{kj}(\lambda)$ — многочлен пространства $\mathscr{P}_{m-1}$. Тогда равенство $Z_{kj} = \varphi_{kj}(A)$ означает, что $h(A) = 0$. Но m — степень минимального многочлена матрицы A и, так как h аннулирует A, из теоремы 4.4.1 следует, что h — нулевой многочлен и, следовательно (так как φ_{kj} независимы), что $c_{kj} = 0$ для всех k и j. Таким образом, матрицы Z_{kj} линейно независимы. ◀

Мы будем называть матрицы Z_{kj} *компонентами* матрицы A. Заметим, что, так как они линейно независимы, ни одна из них не может быть нулевой.

Рассмотрим более подробно случай простых матриц. В (5.4.1) мы имеем теперь $m_1 = m_2 = \ldots = m_s = 1$, так что

$$f(A) = \sum_{k=1}^{s} f_k Z_{k1}$$

для простой матрицы A. Сравним это равенство с результатом теоремы 2.5.1. Применяя оба результата к базисным лагранжевым многочленам $l_k(\lambda)$, определенным в § 5.2, получаем, что Z_{k1} равна сумме сопровождающих матриц, соответствующих собственному значению λ_k. Как тогда нетрудно видеть, Z_{k1} — идемпотентная матрица и образ Z_{k1} — это правое собственное подпространство, соответствующее λ_k (т. е. нулевое подпространство матрицы $I\lambda_k - A$). Таким образом, теорема 2.5.1 оказывается частным случаем теоремы 5.4.1. Из следствия теоремы 2.5.1 получаем, что

$$\sum_{k=1}^{s} Z_{k1} = I \quad \text{и} \quad Z_{k1} Z_{j1} = \delta_{kj} Z_{k1}, \tag{5.4.2}$$

хотя первый из этих результатов, как легко видеть, верен для всех матриц $A \in \mathscr{C}_{n \times n}$, следует лишь положить $f(\lambda) = 1$ в равенстве (5.4.1). В следующем параграфе мы покажем, что второй результат тоже верен в общем случае.

Заметим, наконец, что в рассматриваемом случае $\varphi_{k1} = l_k$, $k = 1, 2, \ldots, s$, так что

$$Z_{k1} = \varphi_{k1}(A) = l_k(A) = \prod_{\substack{j=1 \\ j \neq k}}^{s} (A - \lambda_j I) \Big/ \prod_{\substack{j=1 \\ j \neq k}}^{s} (\lambda_k - \lambda_j). \tag{5.4.3}$$

***Упр. 1.** Если $A \in \mathscr{C}_{n\times n}$, то доказать, что

$$A = \sum_{k=1}^{s} (\lambda_k Z_{k1} + Z_{k2}) \quad \text{и}$$

$$(\mu I - A)^{-1} = \sum_{k=1}^{s} \sum_{j=1}^{m_k} \frac{(j-1)!}{(\mu - \lambda_k)^j} Z_{kj}.$$

Упр. 2. Если A — простая матрица, ψ — минимальный многочлен и $C(\lambda)$ — приведенная присоединенная к A матрица, то доказать, что

$$Z_{k1} = \frac{C(\lambda_k)}{\psi^{(1)}(\lambda_k)}.$$

Решение. Заметим сначала, что, так как A простая,

$$\psi(\lambda) = \prod_{j=1}^{s} (\lambda - \lambda_j)$$

и поэтому

$$\psi^{(1)}(\lambda_j) = \prod_{\substack{j=1\\ j\neq k}}^{s} (\lambda_k - \lambda_j) \neq 0.$$

Таким образом, приняв во внимание равенство (5.4.3), следует лишь доказать, что

$$C(\lambda_k) = \prod_{\substack{j=1\\ j\neq k}}^{s} (A - \lambda_j I).$$

Используя результаты § 4.5, записываем $C(\lambda) = \Psi(I\lambda, A)$, где

$$\Psi(\lambda, \mu) = \frac{\psi(\lambda) - \psi(\mu)}{\lambda - \mu}.$$

Так как $\psi(\lambda_k) = 0$, то

$$\Psi(\lambda_k, \mu) = \frac{-\prod_{j=1}^{s} (\mu - \lambda_j)}{\lambda_k - \mu} = \prod_{\substack{j=1\\ j\neq k}}^{s} (\mu - \lambda_j).$$

Следовательно,

$$C(\lambda_k) = \Psi(I\lambda_k, A) = \prod_{\substack{j=1\\ j\neq k}}^{s} (A - \lambda_j I).$$

***Упр. 3.** Доказать результат упр. 2 при предположении, что λ_k имеет лишь линейные элементарные делители (и A не обязательно простая).

5.5. Свойства компонентных матриц

Мы исследуем теперь свойства компонент некоторой общей матрицы $A \in \mathscr{C}_{n \times n}$. При этом впервые находит применение жорданова нормальная форма. Если J — жорданова нормальная форма для A, то существует такая неособая матрица $T \in \mathscr{C}_{n \times n}$, что $A = TJT^{-1}$, и если

$$J = \operatorname{diag}\{J_1, \ldots, J_p\},$$

то из теоремы 5.3.2 получаем

$$Z_{kj} = \varphi_{kj}(A) = T \operatorname{diag}\{\varphi_{kj}(J_1), \ldots, \varphi_{kj}(J_p)\} T^{-1}. \qquad (5.5.1)$$

Рассмотрим две возможности:

Случай 1. Пусть J_μ — жорданова клетка порядка n_μ, соответствующая λ_k. Тогда, используя базисные интерполяционные свойства φ_{kj} в λ_k, т. е. то, что $\varphi_{kj}^{(r-1)}(\lambda_k) = \delta_{rj}$, $r, j = 1, 2, \ldots, m_k$, и упр. 4 из § 5.3, убеждаемся в том, что

$$\varphi_{kj}(J_\mu) = \frac{1}{(j-1)!}(H_{n_\mu})^{j-1}.$$

Заметим также, что $n_\mu \leqslant m_k$.

Случай 2. Другая возможность состоит в том, что жорданова клетка J_ν порядка n_ν соответствует собственному значению $\lambda_l \neq \lambda_k$. Тогда $n_\nu \leqslant m_l$, и мы замечаем, что φ_{kj} вместе со своими первыми $m_l - 1$ производными равна нулю в $\lambda = \lambda_l$. Опять используя упр. 4 из § 5.3, получаем, что

$$\varphi_{kj}(J_\nu) = 0.$$

Для упрощения обозначений рассмотрим теперь случай $k = 1$ и предположим, что $J_1, \ldots, J_q$ — все жордановы клетки матрицы A, соответствующие λ_1. Тогда в равенстве (5.5.1) будем иметь

$$(j-1)!\, Z_{1j} = T \operatorname{diag}\left\{(H_{n_1})^{j-1}, \ldots, (H_{n_q})^{j-1}, 0, \ldots, 0\right\} T^{-1} \qquad (5.5.2)$$

для $j = 1, 2, \ldots, m_1$. В частности, имеем

$$Z_{11} = T \operatorname{diag}\left\{I_{n_1 + \ldots + n_q}, 0\right\} T^{-1}. \qquad (5.5.3)$$

Вид результатов для общих λ_k теперь совершенно ясен, а если это так, то утверждения следующей теоремы очевидны.

Теорема 5.5.1. *Компоненты Z_{kj} $(k = 1, 2, \ldots, s,\ j = 1, 2, \ldots, m_k)$ матрицы $A \in \mathscr{C}_{n \times n}$ удовлетворяют условиям:*

(i) $\sum_{k=1}^{s} Z_{k1} = I$;

(ii) $Z_{kp} Z_{lr} = 0$, *если* $k \neq l$;

(iii) $Z_{kj}^2 = Z_{kj} \Leftrightarrow j = 1$;

(iv) $Z_{k1}Z_{kr} = Z_{kr}$, $r = 1, 2, \ldots, m_k$.

Эти результаты следует рассматривать как обобщения результатов следствия 1 теоремы 2.5.1. Следующая теорема позволяет нам очень легко находить все компоненты, как только известны идемпотентные компоненты Z_{k1}.

Теорема 5.5.2. *Для* $k = 1, 2, \ldots, s$ *и* $j = 1, 2, \ldots, m_k$ *имеем*

$$Z_{kj} = \frac{1}{(j-1)!}(A - \lambda_k I)^{j-1} Z_{k1}.$$

Доказательство. Заметим сначала, что результат тривиален для $j = 1$. Опять без потери общности мы предположим $k = 1$ и воспользуемся соответствующими обозначениями, введенными выше. Равенство (5.5.2) для $j = 2, 3, \ldots, m_1$ дает

$$(j-2)!\, T^{-1} Z_{1,\,j-1} T = \operatorname{diag}\left\{(H_{n_1})^{j-2}, \ldots, (H_{n_q})^{j-2}, 0, \ldots, 0\right\}.$$

Затем, так как $A - \lambda_1 I = T(J - \lambda_1 I)T^{-1}$, имеем также

$$T^{-1}(A - \lambda_1 I)T = \operatorname{diag}\left\{H_{n_1}, \ldots, H_{n_q}, J_{q+1} - \lambda_1 I, \ldots, J_p - \lambda_1 I\right\}.$$

Объединяя эти два равенства, получаем

$$(j-2)!\, T^{-1}(A - \lambda_1 I) Z_{1,\,j-1} T =$$
$$= \operatorname{diag}\left\{(H_{n_1})^{j-1}, \ldots, (H_{n_q})^{j-1}, 0, \ldots, 0\right\} = (j-1)!\, T^{-1} Z_{1j} T,$$

если опять воспользоваться равенством (5.5.2). Таким образом,

$$Z_{1j} = \frac{1}{j-1}(A - \lambda_1 I) Z_{1,\,j-1}, \qquad j = 2, 3, \ldots, m_k.$$

Применяя этот результат повторно, получаем

$$Z_{1j} = \frac{1}{(j-1)!}(A - \lambda_1 I)^{j-1} Z_{11},$$

что и доказывает теорему. ◄

Теорему 5.5.2 следует дополнить некоторым описанием идемпотентных компонент Z_{k1}.

Теорема 5.5.3. *Если* $\mathcal{R}$ *и* $\mathcal{N}$ *обозначают образ и аннулируемое подпространство соответственно, то*

$$\mathcal{R}(Z_{k1}) = \mathcal{N}(A - \lambda_k I)^{m_k}.$$

Доказательство. 1. Мы опять будем рассматривать случай $k = 1$. Если J — жорданова форма для матрицы A, использованная в доказательствах выше, то

$$(J-\lambda_1 I)^{m_1} = \operatorname{diag}\{(H_{n_1})^{m_1}, \ldots, (H_{n_q})^{m_1}, (J_{q+1}-\lambda_1 I)^{m_1}, \ldots$$
$$\ldots, (J_q-\lambda_1 I)^{m_1}\} = \operatorname{diag}\{0, \ldots, 0, (J_{q+1}-\lambda_1 I)^{m_1}, \ldots, (J_q-\lambda_1 I)^{m_1}\},$$

так как $n_1, n_2, \ldots, n_q \leqslant m_1$. Клетки $(J_r-\lambda_1 I)^{m_1}$, как легко видеть, неособые для $r = q+1, \ldots, p$, а поэтому аннулируемое подпространство матрицы $(J-\lambda_1 I)^{m_1}$ натянуто на единичные векторы $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_N$, где $N = n_1+n_2+\ldots+n_q$. Пусть $\mathscr{E}$ обозначает это пространство.

Имеем также

$$(A-\lambda_1 I)^{m_1} = T(J-\lambda_1 I)^{m_1} T^{-1},$$

откуда сразу же следует, что $\boldsymbol{x} \in \mathscr{N}(A-\lambda_1 I)^{m_1} \Leftrightarrow T^{-1}\boldsymbol{x} \in \mathscr{E}$.

2. Равенство (5.5.3) может быть записано в виде

$$Z_{11} = T\,\|\boldsymbol{e}_1\boldsymbol{e}_2 \ldots \boldsymbol{e}_N\boldsymbol{0} \ldots \boldsymbol{0}\|\,T^{-1}.$$

Таким образом, $\xi \in \mathscr{R}(Z_{11})$ в том и только в том случае, когда существует такой вектор $\boldsymbol{\eta}$, что $\xi = Z_{11}\boldsymbol{\eta}$, или

$$T^{-1}\xi = \|\boldsymbol{e}_1 \ldots \boldsymbol{e}_N \;\; \boldsymbol{0} \ldots \boldsymbol{0}\|\,T^{-1}\boldsymbol{\eta},$$

т. е. когда $T^{-1}\xi \in \mathscr{R}(\|\boldsymbol{e}_1 \ldots \boldsymbol{e}_N \;\; \boldsymbol{0} \ldots \boldsymbol{0}\|) = \mathscr{E}$.

Объединяя полученные утверждения, имеем теперь

$$\boldsymbol{x} \in \mathscr{N}(A-\lambda_1 I)^{m_1} \Leftrightarrow T^{-1}\boldsymbol{x} \in \mathscr{E}$$

и

$$\xi \in \mathscr{R}(Z_{11}) \Leftrightarrow T^{-1}\xi \in \mathscr{E},$$

откуда и следует теорема.

К сожалению, результат доказанной теоремы не обобщается до того, чтобы дать столь же простое описание для $\mathscr{R}(Z_{kj})$ при $j > 1$. Однако некоторый результат для $\mathscr{R}(Z_{km_k})$ приводится в упр. 2 ниже.

Упр. 1. Доказать, что $(A-\lambda_k I)^{m_k} Z_{k1} = 0$.

Упр. 2. Доказать, что $\mathscr{R}(Z_{k1}) \supset \mathscr{R}(Z_{k2}) \supset \ldots \supset \mathscr{R}(Z_{km_k})$ и что $\mathscr{R}(Z_{km_k})$ будет подпространством S в $\mathscr{N}(A-\lambda_k I)$. Показать, что размерность S равна числу жордановых клеток, соответствующих λ_k и имеющих порядок m_k.

Упр. 3. Если $\boldsymbol{x}$, $\boldsymbol{y}$ — правый и левый собственные векторы соответственно с собственным значением λ_1, то доказать, что $Z_{11}\boldsymbol{x} = \boldsymbol{x}$ и $\boldsymbol{y}'Z_{11} = \boldsymbol{y}'$.

Упр. 4. Если A — эрмитова матрица, то доказать, что Z_{k1} будет проекцией на правое собственное подпространство для λ_k.

Обращаясь к задаче нахождения матриц Z_{kj}, видим, что очевидный способ — это найти сначала базисные эрмитовы интерполяционные многочлены φ_{kj} и затем положить $Z_{kj}=\varphi_{kj}(A)$. Однако, как мы видели, вычисление многочленов φ_{kj} может быть очень трудоемким, и упражнение 5 указывает, как этого можно избежать.

***Упр. 5.** (ср .упр. 4 из § 4.5). Найти компоненты матрицы

$$A=\begin{Vmatrix} 6 & 2 & 2 \\ -2 & 2 & 0 \\ 0 & 0 & 2 \end{Vmatrix}.$$

Решение. Как мы видели, минимальный многочлен матрицы A равен $(\lambda-4)^2(\lambda-2)$, так что для любой функции f, определенной на спектре A, теорема 5.4.1 дает

$$f(A)=f(4)Z_{11}+f^{(1)}(4)Z_{12}+f(2)Z_{21}.$$

Мы подставим вместо f три целесообразно выбранных линейно независимых многочлена и решим полученную систему уравнений относительно Z_{11}, Z_{12}, Z_{21}. Таким образом, полагая $f(\lambda)=1,\ \lambda-4,\ (\lambda-4)^2$, по очереди получаем

$$Z_{11}+Z_{21}=I,$$

$$Z_{12}-2Z_{21}=A-4I=\begin{Vmatrix} 2 & 2 & 2 \\ -2 & -2 & 0 \\ 0 & 0 & -2 \end{Vmatrix},$$

$$4Z_{21}=(A-4I)^2=\begin{Vmatrix} 0 & 0 & 0 \\ 0 & 0 & -4 \\ 0 & 0 & 4 \end{Vmatrix}.$$

Эти уравнения легко решаются:

$$Z_{11}=\begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{Vmatrix},\quad Z_{12}=\begin{Vmatrix} 2 & 2 & 2 \\ -2 & -2 & -2 \\ 0 & 0 & 0 \end{Vmatrix},$$

$$Z_{21}=\begin{Vmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{Vmatrix}.$$

Заметим, что в общем случае результаты теорем 5.5.1 и 5.5.2 могут помочь в решении уравнений для компонентных матриц.

Упр. 6. Найти компоненты для матрицы

$$\begin{Vmatrix} -1 & 1 & 2 & 1 \\ 0 & -1 & 0 & -3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{Vmatrix}.$$

5.6. Последовательности и ряды матриц

В этом параграфе мы показываем, как можно применить другие понятия анализа к матрицам, в частности к функциям от матриц, как они были определены. Мы рассмотрим сначала множества матриц, определенных неотрицательными целыми числами, т. е. последовательности матриц.

Пусть $A_1, A_2, \ldots$ — некоторая последовательность матриц, принадлежащих $\mathscr{C}_{m\times n}$, и пусть $a_{ij}^{(p)}$ — (i, j)-элемент матрицы A_p, $p = 1, 2, \ldots$ Последовательность $A_1, A_2, \ldots$ называется *сходящейся* к матрице $A \in \mathscr{C}_{m\times n}$, если существуют такие числа a_{ij} (элементы матрицы A), что $a_{ij}^{(p)} \to a_{ij}$ при $p \to \infty$ для каждой пары индексов i, j. Последовательность, которая не сходится, называется *расходящейся*. Таким образом, сходимость последовательностей матриц определяется как поэлементная сходимость. Введенное определение включает также в качестве частных случаев сходимость вектор-столбцов и вектор-строк.

Упр. 1. Пусть $A_1, A_2, \ldots$ и $B_1, B_2, \ldots$ — сходящиеся последовательности из $\mathscr{C}_{m\times n}$ и $A_p \to A$, $B_p \to B$ при $p \to \infty$. Доказать, что при $p \to \infty$ будет $A_p + B_p \to A + B$. Если последовательность $B_1, B_2, \ldots$ из $\mathscr{C}_{n\times l}$, то доказать, что $A_pB_p \to AB$.

[Указание. Эти результаты следуют из основных свойств последовательностей комплексных чисел. Для второго утверждения, если $a_p \to a$ и $b_p \to b$ при $p \to \infty$ (для комплексных чисел), то записать

$$a_nb_n = \{a + (a_n - a)\}\{b + (b_n - b)\},$$

откуда получаем

$$a_nb_n - ab = (a_n - a)\,b + (b_n - b)\,a + (a_n - a)(b_n - b).]$$

Рассмотрим случай, когда члены последовательности матриц определяются некоторой последовательностью функций, действующих на A. Таким образом, если функции $f_1, f_2, \ldots$ определены на спектре A, то $A_p = f_p(A)$, $p = 1, 2, \ldots$ Для каждого члена последовательности по теореме 5.4.1 получаем

$$f_p(A) = \sum_{k=1}^{s} \sum_{j=1}^{m_k} f_p^{(j-1)}(\lambda_k)\, Z_{kj}. \tag{5.6.1}$$

Но компоненты Z_{kj} зависят лишь от A, но не от p, так что каждый член последовательности определен m числами $f_p^{(j-1)}(\lambda_k)$ для $k = 1, 2, \ldots, s$ и $j = 1, 2, \ldots, m_k$. Мы вправе поэтому ожидать, что сходимость последовательности $A_1, A_2, \ldots$ может быть сделана зависящей от сходимости m скалярных последовательностей $f_1^{(j-1)}(\lambda_k)$, $f_2^{(j-1)}(\lambda_k)$, ..., что позволило бы не рассматривать n^2 скалярных последовательностей элементов из матриц $A_1, A_2, \ldots$

Теорема 5.6.1. *Пусть функции $f_1, f_2, \ldots$ определены на спектре матрицы $A \in \mathscr{C}_{n\times n}$ и $A_p = f_p(A)$, $p = 1, 2, \ldots$ Тогда последовательность $A_1, A_2, \ldots$ сходится при $p \to \infty \Longleftrightarrow m$ скалярных последовательностей $f_1^{(j-1)}(\lambda_k)$, $f_2^{(j-1)}(\lambda_k)$, ... (определенных для $k = 1, 2, \ldots, s$ и $j = 1, 2, \ldots, m_k$) сходятся при $p \to \infty$.*

Кроме того, если $\lim\limits_{p\to\infty} f_p^{(j-1)}(\lambda_k) = f^{(j-1)}(\lambda_k)$ для всех j и k и некоторой функции f, то $\lim\limits_{p\to\infty} A_p = f(A)$ и, наоборот, если $\lim\limits_{p\to\infty} A_p$ существует, то существует такая определенная на спектре A функция f, что $\lim\limits_{p\to\infty} A_p = f(A)$.

Доказательство. 1. Если $f_p^{(j-1)}(\lambda_k) \to f^{(j-1)}(\lambda_k)$ при $p \to \infty$ для всех j и k и некоторой f, то равенство (5.6.1) означает, что

$$\lim_{p\to\infty} A_p = \lim_{p\to\infty} f_p(A) = \sum_k \sum_j \left(\lim f_p^{(j-1)}(\lambda_k)\right) Z_{kj} =$$
$$= \sum_k \sum_j f^{(j-1)}(\lambda_k) Z_{kj} = f(A).$$

2. Наоборот, предположим, что $\lim\limits_{p\to\infty} A_p$ существует и что элементами матрицы $A_p = f_p(A)$ будут $a_{\mu\nu}$, $\mu, \nu = 1, 2, \ldots, n$. Напомним, что компоненты некоторой матрицы линейно независимы как элементы в $\mathscr{C}_{n\times n}$ (теорема 5.4.1), и рассмотрим равенства (5.6.1) как множество n^2 уравнений от m неизвестных $f_p^{(j-1)}(\lambda_k)$, причем p фиксировано. Независимость матриц Z_{kj} означает, что $n^2 \times m$-матрица коэффициентов этих уравнений будет ранга m, а поэтому существует единственное решение вида

$$f_p^{(j-1)}(\lambda_k) = \sum_{\mu, \nu=1}^{n} c_{\mu\nu} a_{\mu\nu}^{(p)},$$

где $c_{\mu\nu}$ зависят от j и k, но не от p. Но так как $\lim\limits_{p\to\infty} a_{\mu\nu}^{(p)}$ существует, отсюда следует, что $\lim\limits_{p\to\infty} f_p^{(j-1)}(\lambda_k)$ тоже существует для всех j и k. Это завершает первую часть теоремы.

Из равенства (5.6.1) можно теперь получить, что

$$\lim_{p\to\infty} A_p = \lim_{p\to\infty} f_p(A) = \sum_k \sum_j \left(\lim_{p\to\infty} f_p^{(j-1)}(\lambda_k)\right) Z_{kj},$$

и мы можем выбрать в качестве f интерполяционный многочлен, значениями которого на спектре A будут числа

$$\lim_{p\to\infty} f_p^{(j-1)}(\lambda_k), \qquad k=1,\ 2,\ \ldots,\ s \quad \text{и} \quad j=1,\ 2,\ \ldots,\ m_k.$$

Это завершает доказательство. ◄

Мы хотим теперь получить подобную теорему для рядов матриц. Прежде всего *ряд* матриц $\sum\limits_{p=0}^{\infty} A_p$ называется *сходящимся* к матрице A (*сумме* ряда), если последовательность частичных сумм $B_\nu = \sum\limits_{p=0}^{\nu} A_p$ сходится к A при $\nu\to\infty$. Ряд, который не сходится, называется *расходящимся*.

Следующая теорема немедленно следует из применения теоремы 5.6.1 к последовательности частичных сумм.

Теорема 5.6.2. *Если функции u_1, u_2, ... определены на спектре $A\in\mathscr{C}_{n\times n}$, то $\sum\limits_{p=1}^{\infty} u_p(A)$ сходится $\Leftrightarrow$ ряды $\sum\limits_{p=1}^{\infty} u_p^{(j-1)}(\lambda_k)$ сходятся, где $k=1,\ 2,\ \ldots,\ s$ и $j=1,\ 2,\ \ldots,\ m_k$.*

Кроме того, если

$$\sum_{p=1}^{\infty} u_p^{(j-1)}(\lambda_k) = f^{(j-1)}(\lambda_k)$$

для всех j и k, то

$$\sum_{p=1}^{\infty} u_p(A) = f(A),$$

и наоборот.

Мы можем теперь применить эту теорему к получению одного важного результата, который даст возможность оправдать наше изложение теории функций от матриц. Следующая теорема показывает общность данного в § 5.3 определения, которое при первом знакомстве могло показаться несколько произвольным. Действительно, читатель обнаружит, что описанные в следующей теореме свойства часто используются при определении функции от матрицы. Наш подход допускает определение $f(A)$ для более широкого класса функций f, чем описанный в теореме 5.6.3.

Теорема 5.6.3. *Пусть матрица $A\in\mathscr{C}_{n\times n}$ имеет собственные значения λ_1, λ_2, ..., λ_n. Если функция f имеет ряд Тейлора в точке λ_0:*

$$f(\lambda) = \sum_{p=0}^{\infty} \alpha_p (\lambda-\lambda_0)^p,$$

с кругом сходимости $|\lambda-\lambda_0|=r$ *и если* $|\lambda_j-\lambda_0|<r$, $j=1, 2, \ldots, n$, *то* $f(A)$ *определена и*

$$f(A)=\sum_{p=0}^{\infty}\alpha_p(A-\lambda_0 I)^p.$$

Доказательство. Как хорошо известно, при сделанных предположениях f имеет производные всех порядков в точках λ, для которых $|\lambda-\lambda_0|<r$, и эти производные имеют ряды Тейлора, которые могут быть получены почленным дифференцированием ряда для f. Таким образом, так как $|\lambda_j-\lambda_0|<r$ для каждого j, f определена на спектре матрицы A.

Если записать $u_p(\lambda)=\alpha_p(\lambda-\lambda_0)^p$, то u_p несомненно определена на спектре A, $p=0, 1, 2, \ldots$, и ряды

$$\sum_{p=1}^{\infty}u_p^{(j-1)}(\lambda_k),\qquad k=1, 2, \ldots, s;\quad j=1, 2, \ldots, m_k,$$

все сходятся. Из теоремы 5.6.2 теперь следует, что $\sum_{p=0}^{\infty}u_p(A)$ сходится и имеет сумму $f(A)$. ◀

Если ряд Тейлора для f в начале координат сходится для всех точек на комплексной плоскости (радиус сходимости бесконечен), то f называется *целой* функцией и представление для $f(A)$ в виде ряда будет сходиться для всех $A\in\mathscr{C}_{n\times n}$. Наиболее важные функции такого рода — это тригонометрические и экспоненциальная функции. Таким образом, из соответствующих результатов для скалярных функций мы получаем, что для всех $A\in\mathscr{C}_{n\times n}$

$$\sin A=\sum_{p=0}^{\infty}\frac{(-1)^p}{(2p+1)!}A^{2p+1},\quad \cos A=\sum_{p=0}^{\infty}\frac{(-1)^p}{(2p)!}A^{2p},$$

$$e^A=\sum_{p=0}^{\infty}\frac{A^p}{p!},\quad \operatorname{sh}A=\sum_{p=0}^{\infty}\frac{A^{2p+1}}{(2p+1)!},\quad \operatorname{ch}A=\sum_{p=0}^{\infty}\frac{A^{2p}}{(2p)!}.$$

Если собственные значения матрицы A равны $\lambda_1, \ldots, \lambda_n$, то из биномиальной теоремы и логарифмического ряда получаем, что при $|\lambda_j|<1$ для $j=1, 2, \ldots, n$ имеем

$$(I-A)^{-1}=I+A+A^2+\ldots$$

и

$$\ln(I+A)=\sum_{p=1}^{\infty}\frac{(-1)^{p-1}}{p}A^p.$$

5.7. Свойства некоторых элементарных функций

Получив обычные разложения в степенные ряды для некоторых элементарных функций с матричными аргументами, мы ставим вопрос: будут ли столь же обычные тождества между этими функциями верными в матричном случае? Например, следует ли из тождества $\sin^2\lambda + \cos^2\lambda = 1$, что $\sin^2 A + \cos^2 A = I$, и из $e^{\lambda}e^{-\lambda} = 1$, что $e^{A}e^{-A} = I$? Ответ на эти вопросы (и целый класс таких вопросов) делается в следующей теореме.

Теорема 5.7.1. *Пусть $G(u_1, u_2, \ldots, u_l)$ — многочлен от $u_1, \ldots, u_l$ и $f_1, \ldots, f_l$ — определенные на спектре матрицы $A \in \mathscr{C}_{n\times n}$ функции, для которых функция $g = G(f_1, \ldots, f_l)$ равна нулю на спектре A. Тогда*

$$G(f_1(A), \ldots, f_l(A)) = 0.$$

Доказательство. Пусть $r_1, r_2, \ldots, r_l$ — интерполяционные многочлены для $f_1, \ldots, f_l$ соответственно на спектре A и определим многочлен $h = G(r_1, \ldots, r_l)$. Тогда $r_\nu(A) = f_\nu(A)$, $\nu = 1, 2, \ldots, l$, так что

$$G(f_1(A), \ldots, f_l(A)) = G(r_1(A), \ldots, r_l(A)) = h(A).$$

Но g равна нулю на спектре A и, как мы видим, h и g принимают одинаковые значения на спектре A. Следовательно, $h(A) = 0$. ◀

Для упомянутых выше частных случаев мы выбираем сначала $f_1(\lambda) = \sin\lambda$, $f_2(\lambda) = \cos\lambda$ и $G(u_1, u_2) = u_1^2 + u_2^2 - 1$, чтобы доказать, что для каждой матрицы $A \in \mathscr{C}_{n\times n}$ имеем $\sin^2 A + \cos^2 A = I$. Взяв $f_1(\lambda) = e^{\lambda}$, $f_2(\lambda) = e^{-\lambda}$ и $G(u_1, u_2) = u_1u_2 - 1$, получаем $e^{A}e^{-A} = I$. Таким образом, $e^{-A} = (e^{A})^{-1}$.

Мы можем обсудить также возможность извлечения с помощью теоремы 5.7.1 корней степени n из некоторой матрицы. Предположим, что p — положительное целое число, и пусть $f_1(\lambda) = \lambda^{1/p}$, $f_2(\lambda) = \lambda$ и $G(u_1, u_2) = u_1^p - u_2$, где для f_1 обычно выбирается некоторая однозначная ветвь функции $\lambda^{1/p}$, хотя это в данном случае несущественно. Если, однако, матрица A имеет нулевое собственное значение с индексом, бóльшим 1, то, так как потребуются значения производных от f_1 в начале координат, f_1 не была бы определена на спектре A. Таким образом, если A неособая или особая с нулевым собственным значением индекса 1, то мы определяем $f_1(\lambda) = \lambda^{1/p}$ (с известным соглашением относительно выбора ветвей в собственных значениях) и

$$A^{1/p} \doteq f_1(A) = \sum_{k=1}^{s}\sum_{j=1}^{m_k} f_1^{(j-1)}(\lambda_k) Z_{kj}.$$

Тогда

$$(A^{1/p})^p = A.$$

Это рассуждение следует сравнить с определением квадратного корня из определенной матрицы, которое было дано в доказательстве теоремы 2.14.2.

Следует, пожалуй, предупредить читателя, что наше определение для $A^{1/2}$ не включает, например, всех возможных матриц B, для которых $B^2 = A$. Для иллюстрации этого укажем на то, что все вещественные симметрические ортогональные матрицы B обладают свойством $B^2 = I$, но не могут быть с необходимостью получены применением наших определений для нахождения $I^{1/2}$. Частный случай — это множество матриц, определенных как

$$B(\theta) = \begin{Vmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{Vmatrix}.$$

Упр. 1. Доказать, что для любой $A \in \mathscr{C}_{n\times n}$

$$e^{iA} = \cos A + i \sin A.$$

Упр. 2. Если g, h — взаимно простые многочлены, $\det(h(A)) \neq 0$ и мы определяем рациональную функцию r как $r(\lambda) = g(\lambda)/h(A)$, то доказать, что

$$r(A) = g(A)[h(A)]^{-1} = [h(A)]^{-1} g(A).$$

Упр. 3. Доказать, что если $A \in \mathscr{C}_{n\times n}$, то $e^{A(s+t)} = e^{As}e^{At}$ для любых комплексных чисел s и t.

Упр. 4. Доказать, что если A и B коммутируют, то $e^{A+B} = e^Ae^B$, и если $e^{(A+B)t} = e^{At}e^{Bt}$ для всех t, то A и B коммутируют.

5.8. Использование контурных интегралов*)

Теория функций комплексной переменной дает возможность получить красивое выражение для компонент произвольной матрицы в $\mathscr{C}_{n\times n}$. В частности, нам потребуется теорема Коши о вычетах, которую мы сформулируем. Прежде всего, будем называть функцию f комплексной переменной z *аналитической* на множестве точек D комплексной плоскости, если f непрерывно дифференцируема в каждой точке D (или, что эквивалентно, если f имеет сходящееся разложение в ряд Тейлора в окрестности каждой точки из D).

---

*) Читателю, который не знаком с основными теоремами о функциях комплексной переменной, следует пропустить этот параграф, но тем не менее просмотреть упражнения в конце § 5.8.

Теорема о вычетах утверждает, что если функция f непрерывна на замкнутом контуре $\mathscr{C}$ и аналитична, за исключением конечного числа полюсов, внутри $\mathscr{C}$, то

$$\int_{\mathscr{C}} f(z)\,dz = 2\pi i \{\text{сумма вычетов } f \text{ в ее полюсах внутри } \mathscr{C}\}.$$

Мы будем рассматривать матрицы $A(z)$, элементы которых — функции комплексной переменной, и (как и в § 2.15) определим производную (или интеграл) от матрицы как матрицу, получаемую дифференцированием (или интегрированием) каждого элемента в $A(z)$. Мы пишем

$$\frac{d}{dz} A(z) = A^{(1)}(z) \quad \text{и} \quad \int A(z)\,dz$$

для производной и интеграла матрицы $A(z)$ соответственно.

Пусть A — некоторая фиксированная матрица из $\mathscr{C}_{n\times n}$. Рассмотрим зависящую от A матрицу R_z, определенную как функция комплексной переменной z посредством

$$R_z = (zI - A)^{-1}.$$

Как очевидно, R_z определена $\Longleftrightarrow z$ не является собственным значением матрицы A. Эта матрица известна как *резольвента* для A.

Если, в наших обычных обозначениях, A имеет компоненты Z_{kj}, $k = 1, 2, \ldots, s$ и $j = 1, 2, \ldots, m_k$, то применение теоремы 5.4.1 (ср. упр. 1 из § 5.4) дает

$$R_z = \sum_{k=1}^{s} \sum_{j=1}^{m_k} \frac{(j-1)!}{(z-\lambda_k)} Z_{kj}. \tag{5.8.1}$$

Пусть $\mathscr{C}_\nu$ — окружность на комплексной плоскости с центром λ_ν, не имеющая других собственных значений матрицы A внутри или на $\mathscr{C}_\nu$. Проинтегрируем тогда обе части написанного равенства по $\mathscr{C}_\nu$. Матрица вычетов n^2 интегралов справа равна как раз $Z_{\nu 1}$. Таким образом, из теоремы о вычетах получаем

$$\int_{\mathscr{C}_\nu} R_z\,dz = 2\pi i Z_{\nu 1},$$

и идемпотентные компоненты для A задаются в виде

$$Z_{k1} = \frac{1}{2\pi i} \int_{\mathscr{C}_k} R_z\,dz, \qquad k = 1, 2, \ldots, s.$$

Но проведенное рассуждение может быть легко обобщено до получения формул для всех компонент в виде контурных интегралов. Умножим сначала обе части равенства (5.8.1) на $(z-\lambda_\nu)^t$ $(0 \leqslant t \leqslant m_\nu - 1)$ и затем проинтегрируем их по $\mathscr{C}_\nu$. Таким образом,

$$\int_{\mathscr{C}_\nu} (z-\lambda_\nu)^t R_z\,dz = \sum_k \sum_j (j-1)!\,Z_{kj} \int_{\mathscr{C}_\nu} \frac{(z-\lambda_\nu)^t}{(z-\lambda_k)^j}\,dz.$$

Единственные ненулевые интегралы справа — это те, для которых $k=\nu$ и $j=t+1$, и этот интеграл имеет значение $2\pi i$. Таким образом,

$$\int_{\mathscr{C}_\nu} (z-\lambda_\nu)^t R_z\,dz = (2\pi i)\,t!\,Z_{\nu,\,t+1}$$

и в общем виде

$$Z_{kj} = \frac{1}{(j-1)!\,2\pi i} \int_{\mathscr{C}_k} (z-\lambda_k)^{j-1} R_z\,dz, \tag{5.8.2}$$

$$k=1,\ 2,\ \ldots,\ s, \quad j=1,\ 2,\ \ldots,\ m_k.$$

Контурные интегралы могут быть также использованы для получения третьего выражения для $f(A)$ при подходящих функциях f. Пусть $A \in \mathscr{C}_{n\times n}$ имеет различные собственные значения $\lambda_1, \ldots, \lambda_s$, и пусть $\mathscr{C}$ — замкнутый контур, содержащий $\lambda_1, \ldots, \lambda_s$ внутри себя. Предположим, что функция f непрерывна на $\mathscr{C}$ и аналитична внутри $\mathscr{C}$. Умножая обе части равенства (5.8.1) на $f(z)$ и интегрируя по z вдоль контура $\mathscr{C}$, получим

$$\int_{\mathscr{C}} f(z) R_z\,dz = \sum_k \sum_j (j-1)!\,Z_{kj} \int_{\mathscr{C}} \frac{f(z)}{(z-\lambda_k)^j}\,dz.$$

Как хорошо известно, интегральная формула Коши теперь дает

$$\int_{\mathscr{C}} \frac{f(z)}{(z-\lambda_k)^j}\,dz = \frac{2\pi i}{(j-1)!} f_k^{(j-1)}.$$

Следовательно,

$$\int_{\mathscr{C}} f(z) R_z\,dz = 2\pi i \sum_k \sum_j f_k^{(j-1)} Z_{kj} = 2\pi i f(A)$$

по теореме 5.4.1. Нами доказана

Теорема 5.8.1. *Если $A \in \mathscr{C}_{n\times n}$ имеет различные собственные значения $\lambda_1, \ldots, \lambda_s$, $\mathscr{C}$ — замкнутый контур, содержащий $\lambda_1, \ldots, \lambda_s$ внутри себя и f непрерывна на $\mathscr{C}$ и аналитична внутри $\mathscr{C}$, то*

$$f(A) = \frac{1}{2\pi i} \int_{\mathscr{C}} f(z)(zI-A)^{-1}\,dz.$$

Заметим, что результаты упражнений 1, 2 и 3 ниже применимы равным образом как к вещественной, так и к комплексной переменной t.

***Упр. 1.** Доказать, что

$$\frac{d}{dt}(e^{At}) = Ae^{At} = e^{At}A.$$

Упр. 2. Доказать, что

$$\frac{d}{dt}(A(t))^2 = A^{(1)}A + AA^{(1)},$$

и построить пример, показывающий, что, вообще говоря,

$$\frac{d}{dt}(A(t))^2 \neq 2AA^{(1)}.$$

***Упр. 3.** Доказать, что в случае, когда p — положительное целое число и матрицы существуют:

(i) $$\frac{d}{dt}(A(t))^p = \sum_{j=1}^{p} A^{j-1}A^{(1)}A^{p-j};$$

(ii) $$\frac{d}{dt}(A(t))^{-p} = -A^{-p}\frac{dA^p}{dt}A^{-p}.$$

***Упр. 4.** Доказать, что в обозначениях § 5.8

$$AZ_{k1} + Z_{k2} = \frac{1}{2\pi i}\int_{\mathscr{C}_k} zR_z\,dz.$$

5.9. Приложения к решению дифференциальных уравнений

Мы рассмотрим сначала систему обыкновенных дифференциальных уравнений с постоянными коэффициентами следующего вида:

$$\begin{aligned}
\dot{x}_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n,\\
\dot{x}_2 &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n,\\
&\vdots\\
\dot{x}_n &= a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n,
\end{aligned}$$

где $x_1, \ldots, x_n$ — функции независимой переменной t, $\dot{x}_j$ обозначает производную x_j по t и коэффициенты a_{jk} не зависят от t. Мы можем, очевидно, это множество уравнений заменить одним матричным уравнением

$$\dot{\boldsymbol{x}} = A\boldsymbol{x}.$$

Можно писать $\boldsymbol{x}(t)$ или $\dot{\boldsymbol{x}}(t)$, если есть необходимость отметить зависимость этих функций от t.

Мы можем легко убедиться в существовании решения уравнения $\dot{\boldsymbol{x}} = A\boldsymbol{x}$. Будем искать решение в виде $\boldsymbol{x}(t) = \boldsymbol{\xi}e^{\mu t}$, где $\boldsymbol{\xi}$ и μ не зависят от t. Тогда $\dot{\boldsymbol{x}} = \mu\boldsymbol{x}$ и $\dot{\boldsymbol{x}} = A\boldsymbol{x}$ может быть записано в виде $(A - \mu I)\boldsymbol{\xi} = \mathbf{0}$. Таким образом, если в качестве $\boldsymbol{\xi}$ взять правый собственный вектор для A с собственным значением μ, то $\boldsymbol{x}(t)$ будет решением. Кроме того, такое решение всегда существует, потому что каждая матрица A имеет по крайней мере одно собственное значение с соответствующим правым собственным вектором.

Можно ли характеризовать множество всех решений уравнения $\dot{\boldsymbol{x}} = A\boldsymbol{x}$? Как можно доказать, множество всех решений является линейным пространством размерности n. То, что это множество — линейное пространство, проверяется очень легко, вопрос же о размерности требует более детального рассмотрения, и мы не будем этим заниматься. В данном случае обычное доказательство основано на другой важной теореме: существует *единственное* решение уравнения $\dot{\boldsymbol{x}} = A\boldsymbol{x}$, которое удовлетворяет также начальному условию $\boldsymbol{x}(0) = \boldsymbol{c}$ при произвольном заданном не зависящем от t векторе $\boldsymbol{c}$.

Покажем, как такое решение может быть вычислено. Следующее рассуждение формальное (т. е. не строгое математически), но приводящее к такой вектор-функции, которая, как можно убедиться после дифференцирования, оказывается решением и поэтому единственным решением. Предположим, что вектор-решение может быть разложено в ряд Тейлора в точке $t = 0$. Таким образом, обозначая через $\dot{\boldsymbol{x}}_0$ значение при $t = 0$ первой производной по t от $\boldsymbol{x}$ и т. д., получаем

$$\boldsymbol{x}(t) = \boldsymbol{c} + t\dot{\boldsymbol{x}}_0 + \frac{t^2}{2!}\ddot{\boldsymbol{x}}_0 + \ldots$$

Продифференцируем теперь равенство $\dot{\boldsymbol{x}} = A\boldsymbol{x}$ последовательно для получения

$$\ddot{\boldsymbol{x}} = A\dot{\boldsymbol{x}} = A^2\boldsymbol{x}, \quad \dddot{\boldsymbol{x}} = A\ddot{\boldsymbol{x}} = A^3\boldsymbol{x}$$

и т. д. Подставляя полученные значения в ряд для $\boldsymbol{x}(t)$, получаем

$$\boldsymbol{x}(t) = \boldsymbol{c} + tA\boldsymbol{x}_0 + \frac{t^2}{2!}A^2\boldsymbol{x}_0 + \ldots = \left(I + tA + \frac{t^2}{2!}A^2 + \ldots\right)\boldsymbol{c} = e^{At}\boldsymbol{c}.$$

Таким образом, наше формальное рассуждение подсказывает, что $\boldsymbol{x}(t) = e^{At}\boldsymbol{c}$ может быть решением. Так как производная от

e^{At} равна Ae^{At} (упр. 1 § 5.8), легко проверить, что мы действительно имеем решение уравнения $\dot{x} = Ax$ с $x(0) = c$.

Если в теореме 5.4.1 положить теперь $f(\lambda) = e^{\lambda t}$ и заметить, что $f^{(r)}(\lambda) = t^r e^{\lambda t}$, $r = 0, 1, 2, \ldots$, то будем иметь

$$x(t) = \sum_{k=1}^{s} \left(Z_{k1} + Z_{k2}t + \ldots + Z_{km_k}t^{m_k-1}\right) e^{\lambda_k t} c =$$

$$= \sum_{k=1}^{s} \left(z_{k1} + z_{k2}t + \ldots + z_{km_k}t^{m_k-1}\right) e^{\lambda_k t},$$

где $z_{kj} = Z_{kj}c$ и не зависит от t. Таким образом, как мы видим, решение будет комбинацией многочленов от t с показательными функциями. Если A — простая матрица, то решение сводится к

$$x(t) = \sum_{k=1}^{s} z_{k1} e^{\lambda_k t}$$

и, как мы видим, каждый элемент $x(t)$ будет линейной комбинацией показательных функций без добавлений многочленов.

Упр. 1. Найти решение уравнения $\dot{x} = Ax$ с $x(0) = c$, если

$$A = \begin{Vmatrix} 6 & 2 & 2 \\ -2 & 2 & 0 \\ 0 & 0 & 2 \end{Vmatrix}, \quad c = \begin{Vmatrix} 0 \\ 1 \\ 1 \end{Vmatrix}.$$

Решение. Компоненты для A были найдены в упр. 4 из § 5.5. Векторы $z_{kj} = Z_{kj}c$ из проведенного выше рассуждения оказываются равными

$$z_{11} = \begin{Vmatrix} 0 \\ 2 \\ 0 \end{Vmatrix}, \quad z_{12} = \begin{Vmatrix} 4 \\ -4 \\ 0 \end{Vmatrix}, \quad z_{21} = \begin{Vmatrix} 0 \\ -1 \\ 1 \end{Vmatrix}$$

и, так как $\lambda_1 = 4$, $\lambda_2 = 2$, решением будет

$$x(t) = \left(\begin{Vmatrix} 0 \\ 2 \\ 0 \end{Vmatrix} + \begin{Vmatrix} 4 \\ -4 \\ 0 \end{Vmatrix} t \right) e^{4t} + \begin{Vmatrix} 0 \\ -1 \\ 1 \end{Vmatrix} e^{2t},$$

или

$$x_1(t) = 4te^{4t}, \quad x_2(t) = (2 - 4t)e^{4t} - e^{2t}, \quad x_3(t) = e^{2t}.$$

Рассмотрим теперь неоднородное уравнение

$$\dot{x} = Ax + f,$$

где элементами f будут $f_1(t), \ldots, f_n(t)$ — заданные функции от t, которые не все тождественно нулевые. Пусть $\mathscr{L}$ — пространство всех решений соответствующего однородного уравнения $\dot{x} = Ax$ и $\mathscr{T}$ — множество всех решений для $\dot{x} = Ax + f$. Если нам известно одно решение $y \in \mathscr{T}$, то $z \in \mathscr{T} \Leftrightarrow z = y + w$ для некоторого $w \in \mathscr{L}$. Проверка этого утверждения предоставляется читателю в качестве упражнения. Таким образом, для получения $\mathcal{T}$ требуется знать лишь одно решение неоднородного уравнения, известное как *частный интеграл*, и прибавить к нему все решения однородного уравнения, которые могут быть выражены в виде $e^{At}\boldsymbol{c}$ для некоторого $\boldsymbol{c}$.

Будем искать решение уравнения $\dot{\boldsymbol{x}} = A\boldsymbol{x} + \boldsymbol{f}$ в виде

$$\boldsymbol{x}(t) = e^{At}\boldsymbol{z}(t).$$

Если такое решение существует, то $\dot{\boldsymbol{x}} = A\boldsymbol{x} + e^{At}\dot{\boldsymbol{z}}$, так что $e^{At}\dot{\boldsymbol{z}} = \boldsymbol{f}$, откуда

$$\boldsymbol{z}(t) = e^{-At_0}\boldsymbol{c} + \int e^{-A\tau}\boldsymbol{f}(\tau)\,d\tau,$$

где $\boldsymbol{c} = e^{At_0}\boldsymbol{z}(t_0) = \boldsymbol{x}(t_0)$. Это наводит на мысль, что

$$\boldsymbol{x}(t) = e^{At}\left\{e^{-At_0}\boldsymbol{c} + \int_{t_0}^{t} e^{-A\tau}\boldsymbol{f}(\tau)\,d\tau\right\} = e^{A(t-t_0)}\boldsymbol{c} + \int_{t_0}^{t} e^{A(t-\tau)}\boldsymbol{f}(\tau)\,d\tau$$

будет решением уравнения $\dot{\boldsymbol{x}} = A\boldsymbol{x} + \boldsymbol{f}$, $\boldsymbol{x}(t_0) = \boldsymbol{c}$. Это легко проверяется.

Упр. 2. Найти решение уравнения $\dot{\boldsymbol{x}} = A\boldsymbol{x} + \boldsymbol{f}$, $\boldsymbol{x}(0) = \boldsymbol{c}$, если A, $\boldsymbol{c}$ определены, как в упр. 1, и

$$\boldsymbol{f} = \begin{Vmatrix} ae^{4t} \\ 0 \\ 0 \end{Vmatrix} = ae^{4t}\boldsymbol{e}_1.$$

Решение. Вектор $e^{At}\boldsymbol{c}$ известен из упр. 1. Для частного интеграла имеем

$$\boldsymbol{y}(t) = \int_0^t e^{A(t-\tau)}\boldsymbol{f}(\tau)\,d\tau = \sum_k \sum_l Z_{kl}\int_0^t (t-\tau)^{l-1} e^{\lambda_k(t-\tau)}\boldsymbol{f}(\tau)\,d\tau.$$

Интегралы

$$a\int_0^t e^{4(t-\tau)}e^{4\tau}\,d\tau, \quad a\int_0^t (t-\tau)\,e^{4(t-\tau)}e^{4\tau}\,d\tau, \quad a\int_0^t e^{2(t-\tau)}e^{4\tau}\,d\tau$$

легко находятся и равны ate^{4t}, $\frac{1}{2}at^2e^{4t}$, $\frac{1}{2}ae^{2t}(e^{2t}-1)$ соответственно. Следовательно,

$$\boldsymbol{y}(t) = ate^{4t}Z_{11}\boldsymbol{e}_1 + \frac{1}{2}at^2e^{4t}Z_{12}\boldsymbol{e}_1 + \frac{1}{2}ae^{2t}(e^{2t}-1)\,Z_{21}\boldsymbol{e}_1 =$$
$$= \frac{1}{2}ate^{4t}\left\{2\begin{Vmatrix} 1 \\ 0 \\ 0 \end{Vmatrix} + t\begin{Vmatrix} 2 \\ -2 \\ 0 \end{Vmatrix}\right\} = \begin{Vmatrix} a(1+t)\,te^{4t} \\ -at^2e^{4t} \\ 0 \end{Vmatrix}.$$

Решение тогда задается в виде $\boldsymbol{x}(t) = e^{At}\boldsymbol{c} + \boldsymbol{y}(t)$.

Упр. 3. Пусть $A_1, \ldots, A_n \in \mathscr{C}_{m\times m}$. Доказать, что $\boldsymbol{u}$ будет решением системы уравнений порядка n

$$\boldsymbol{u}^{(n)} + A_1\boldsymbol{u}^{(n-1)} + \ldots + A_{n-1}\boldsymbol{u}^{(1)} + A_n\boldsymbol{u} = \boldsymbol{f}$$

(где индексы обозначают производные) $\Longleftrightarrow$ вектор $\boldsymbol{x}$ будет решением системы первого порядка $d\boldsymbol{x}/dt = A\boldsymbol{x} + \boldsymbol{g}$, где

$$\boldsymbol{x} = \begin{Vmatrix} \boldsymbol{u} \\ \boldsymbol{u}^{(1)} \\ \vdots \\ \boldsymbol{u}^{(n-1)} \end{Vmatrix}, \quad A = \begin{Vmatrix} 0 & I & 0 & \ldots & 0 \\ 0 & 0 & I & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \ldots & & 0 & I \\ -A_n & \ldots & & -A_2 & -A_1 \end{Vmatrix}, \quad \boldsymbol{g} = \begin{Vmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{f} \end{Vmatrix}.$$

Упр. 4. Если определить λ-матрицу $D(\lambda) = I\lambda^n + A_1\lambda^{n-1} + \ldots \ldots + A_n$ и за μ взять скрытый корень для $D(\lambda)$, то найдется такой вектор $\boldsymbol{q} \neq \boldsymbol{0}$, что $D(\mu)\boldsymbol{q} = \boldsymbol{0}$. Доказать, что $\boldsymbol{u}(t) = \boldsymbol{q}e^{\mu t}$ будет решением уравнения $\boldsymbol{u}^{(n)} + A_1\boldsymbol{u}^{(n-1)} + \ldots + A_n\boldsymbol{u} = \boldsymbol{0}$.

Упр. 5. Проверить, что для неособой матрицы A

$$\boldsymbol{x}(t) = \left(\cos A^{1/2}t\right)\boldsymbol{x}_0 + \left(A^{1/2}\right)^{-1}\left(\sin A^{1/2}t\right)\dot{\boldsymbol{x}}_0$$

будет решением уравнения $\ddot{\boldsymbol{x}} = -A\boldsymbol{x}$, $\boldsymbol{x}(0) = \boldsymbol{x}_0$, $\dot{\boldsymbol{x}}(0) = \dot{\boldsymbol{x}}_0$. Как этот результат может быть обобщен до включения особой матрицы A?

Упр. 6. Найти решение уравнения $\ddot{\boldsymbol{x}} = -A\boldsymbol{x} + \boldsymbol{f}$ с $\boldsymbol{x}(0) = \boldsymbol{x}_0$, $\dot{\boldsymbol{x}}(0) = \dot{\boldsymbol{x}}_0$.

***Упр. 7.** Если $A, B \in \mathscr{C}_{n\times n}$ и X — зависящая от t $n \times n$-матрица, то найти решение уравнения

$$\dot{X} = AX + XB, \quad X(0) = C,$$

рассматривая матрицы вида $X = e^{At}Y(t)$.

Смешанные упражнения

1. Если f определена на спектре матрицы A, то доказать, что $f(A') = [f(A)]'$.

2. Как мы видели в упр. 1 из § 5.4,

$$A = \sum_{k=1}^{s} (\lambda_k Z_{k1} + Z_{k2}).$$

Возводя обе части этого равенства в квадрат и используя теоремы 5.5.1 и 5.5.2, доказать, что

$$A^2 = \sum_{k=1}^{s} \left(\lambda_k^2 Z_{k1} + 2\lambda_k Z_{k2} + 2Z_{k3}\right).$$

(Если $m_k = 1$, то $Z_{k2} = Z_{k3} = 0$ и т. д.)

3. Доказать, что если ряд $I + A + A^2 + \ldots$ сходится, то каждое собственное значение матрицы A по модулю меньше 1. (Из биномиального разложения § 5.6 тогда следует, что

$$(I - A)^{-1} = I + A + A^2 + \ldots$$

в том и только в том случае, когда все собственные значения матрицы A по модулю меньше 1.)

4. Показать, что унитарную матрицу U можно представить в виде $U = e^{iH}$, где H эрмитова. (Если пытаться установить аналогию между полем комплексных чисел и множеством (алгеброй) матриц в $\mathscr{C}_{n \times n}$, то эрмитовы матрицы будут соответствовать вещественным числам и унитарные матрицы — числам с единичным модулем.)

[Указание. Показать, что существуют такие вещественные числа $a_1, \ldots, a_n$ и унитарная матрица V, что

$$U = V \operatorname{diag}\left\{e^{ia_1}, \ldots, e^{ia_n}\right\} V^*,$$

и положить

$$H = V \operatorname{diag}\{a_1, \ldots, a_n\} V^*.]$$

5. Для $A \in \mathscr{C}_{n \times n}$ доказать, что существует такая вещественная кососимметрическая матрица S с $A = e^S \Longleftrightarrow A$ — вещественная ортогональная матрица с $\det A = 1$.

6. Определить M^n для всех целых n и всех λ, если

$$M = \begin{Vmatrix} 1 - \lambda + \lambda^2 & 1 - \lambda \\ \lambda - \lambda^2 & \lambda \end{Vmatrix}.$$

7. Если элементы квадратной матрицы $D(\lambda)$ дифференцируемы и $\Delta(\lambda) = \det D(\lambda)$, то показать, что

$$\Delta^{(1)}(\lambda) = \operatorname{tr}\left\{\left(D^{\vee} D^{(1)}\right)\right\},$$

и если $\Delta(\lambda) \neq 0$, то

$$\frac{\Delta^{(1)}(\lambda)}{\Delta(\lambda)} = \operatorname{tr}\left(D^{-1} D^{(1)}\right).$$

8. Предположим, что λ — вещественная переменная и в упр. 7 D — решение матричного дифференциального уравнения $\dot{X}(\lambda) = X(\lambda) A(\lambda)$, где $A(\lambda)$ существует для всех λ. Доказать, что если существует λ_0, для которого $\Delta(\lambda_0) \neq 0$, то

$$\Delta(\lambda) = \Delta(\lambda_0) \exp \int_{\lambda_0}^{\lambda} \operatorname{tr}(A(\mu))\, d\mu$$

и, следовательно, $\Delta(\lambda) \neq 0$ для всех λ. (Это соотношение известно как *тождество Якоби* или *формула Лиувилля*.)

9. Если все собственные значения матрицы A имеют отрицательные вещественные части, то доказать, что

$$A^{-1} = -\int_0^\infty e^{At}\,dt$$

и, следовательно, что если все собственные значения матрицы A лежат в полуплоскости $\mathrm{Re}(z) < \mathrm{Re}(\lambda)$, то

$$(I\lambda - A)^{-1} = \int_0^\infty e^{-(I\lambda - A)}\,dt.$$

[Указание. Использовать равенство (5.8.1).]

Глава 6

НОРМЫ ВЕКТОРОВ И МАТРИЦ

6.1. Матричные нормы

Во многих случаях бывает полезно иметь возможность приписать матрице некоторое однозначно определенное неотрицательное число в качестве меры ее величины. Для комплексного числа мы уже имеем такую меру величины в его модуле, и для вектора $\boldsymbol{x}$ из $\mathscr{C}_n$ величина может быть измерена длиной $\langle \boldsymbol{x}, \boldsymbol{x}\rangle^{1/2}$, введенной в § 1.3. Мы изложим аксиоматический подход к формулировке мер величин, или норм, матриц и начнем с выписывания свойств, выполнения которых мы от такой меры требуем. Как мы увидим, существует много возможных выборов для нормы.

Функция с вещественными значениями, определенная на всех квадратных матрицах A с комплексными элементами, называется *матричной нормой* и обычно записывается в виде $\|A\|$, если она удовлетворяет следующим аксиомам:

(i) $$\|A\| \geqslant 0 \quad \text{и} \quad \|A\| = 0 \Leftrightarrow A = 0;$$

(ii) $\|cA\| = |c|\,\|A\|$ для любого комплексного числа c;

$$\left.\begin{array}{ll} \text{(iii)} & \|A+B\| \leqslant \|A\| + \|B\|, \\ \text{(iv)} & \|AB\| \leqslant \|A\|\,\|B\| \end{array}\right\}$$

для всех квадратных матриц A, B одного и того порядка.

Полезно иметь в виду аналогию с модулем комплексного числа. Аксиомы (i), (ii) и (iii) — непосредственные обобщения, в подтверждение этой аналогии, хотя аксиома (iv) является более либеральной, чем соответствующий результат $|ab| = |a|\,|b|$ для комплексных чисел. Аксиома (iii) известна как неравенство треугольника по очевидной причине.

Заметим, что во всей этой главе все рассматриваемые матрицы будут квадратными.

***Упр. 1.** Доказать, что $\|A - B\| \geqslant |\,\|A\| - \|B\|\,|$.

[Указание. Доказать сначала соответствующий результат для комплексных чисел.]

***Упр. 2.** Доказать, что вещественнозначная функция $\sum_{i,j} |a_{ij}|$ будет матричной нормой.

Упр. 3. Доказать, что вещественнозначная функция $\max_{i,j} |a_{ij}|$ удовлетворяет аксиомам (i), (ii) и (iii), но не аксиоме (iv).

Упр. 4. Доказать, что для любой матричной нормы имеем

$$\|I\| \geqslant 1, \quad \|A^n\| \leqslant \|A\|^n, \quad \|A^{-1}\| \geqslant \frac{1}{\|A\|}.$$

Упр. 5. Если S неособая и $\|\cdot\|$ — матричная норма, то вещественнозначная функция f, определенная как $f(A) = \|SAS^{-1}\|$, будет матричной нормой.

Во многих отношениях аксиомы (i), (ii) и (iii) — это все, что требуется для матричной нормы. Мы поэтому определяем вещественнозначную функцию, удовлетворяющую аксиомам (i), (ii) и (iii) (но не обязательно (iv)), как *обобщенную матричную норму*. Таким образом, матричная норма всегда будет обобщенной матричной нормой, но не наоборот. Упражнение 3 выше дает пример обобщенной матричной нормы, которая не является матричной нормой.

Наша первая теорема утверждает, что обобщенная матричная норма будет обязательно «гладкой» функцией от матричного аргумента в том смысле, что «малые» изменения в матричных элементах приводят к «малому» изменению в матричной норме.

Теорема 6.1.1 *). *Обобщенная матричная норма непрерывно зависит от элементов матрицы, т. е. для заданного $\varepsilon > 0$ существует такое $\delta > 0$, зависящее от ε, что*

$$|\|A\| - \|B\|| < \varepsilon,$$

как только $|a_{ij} - b_{ij}| < \delta$ для всех пар индексов i, j.

Доказательство. Пусть E_{ij} — матрица, имеющая 1 на месте (i, j) и 0 на всех других местах $(E_{ij} = e_i e_j')$. Тогда

$$A - B = \sum_{i,j} (a_{ij} - b_{ij}) E_{ij},$$

и если определить число $k = \max_{i,j} \|E_{ij}\|$, заметив, что $k > 0$ по аксиоме (i), то аксиомы (ii) и (iii) будут означать, что

$$\|A - B\| \leqslant \sum_{i,j} |a_{ij} - b_{ij}| \|E_{ij}\| \leqslant k \sum_{i,j} |a_{ij} - b_{ij}|.$$

*) Строго говоря, в этой теореме следует зафиксировать порядок рассматриваемых матриц. — *Прим. перев.*

Для любого $\varepsilon > 0$ определим $\delta = \varepsilon/kn^2$, где A, B порядка n, и рассмотрим любую пару матриц A, B, для которых

$$|a_{ij} - b_{ij}| < \delta, \quad i, j = 1, 2, \ldots, n.$$

Тогда

$$\| A - B \| < kn^2\delta = \varepsilon.$$

Следовательно, по упр. 1 выше

$$| \| A \| - \| B \| | < \varepsilon,$$

как только $| a_{ij} - b_{ij} | < \delta$. ◀

Наша следующая теорема тоже имеет очень большое значение и связывает величины любых двух обобщенных матричных норм, вычисленных для одного и того же матричного аргумента. Эта теорема сравнения часто сводит изучение свойств общих матричных норм к изучению одной относительно простой нормы (ср. упр. 6 ниже).

Теорема 6.1.2. *Пусть $\| A \|$ и $N(A)$ — любые две обобщенные матричные нормы, вычисленные в A. Тогда существуют такие положительные числа r_1 и r_2, зависящие лишь от выбора норм, что*

$$r_1 \leqslant \frac{\| A \|}{N(A)} \leqslant r_2$$

для всех $A \in \mathscr{C}_{n \times n}$ *).

Доказательство. Мы применим матрицы E_{ij}, использованные в предыдущем доказательстве. Положим $p_2 = \sum_{i, j} \| E_{ij} \|$, $a = \max_{i, j} | a_{ij} |$. Тогда по аксиомам (ii) и (iii) имеем

$$\| A \| \leqslant \sum_{i, j} | a_{ij} | \| E_{ij} \| \leqslant ap_2, \tag{6.1.1}$$

где p_2 не зависит от A.

Рассмотрим теперь множество $\mathscr{M}$ всех матриц, для которых $a = 1$, и пусть

$$p_1 = \min_{B \in \mathscr{M}} \| B \|.$$

Множество $\mathscr{M}$ замкнуто и ограничено; по теореме 6.1.1 $\| B \|$ непрерывно зависит от элементов B, поэтому **) существует такая матрица $B_0 \in \mathscr{M}$, что $p_1 = \| B_0 \|$. Аксиома (i) тогда означает, что $p_0 > 0$ и тоже не зависит от A.

Для данной матрицы A можно записать $A = a_{\mu\nu} B$, где $|a_{\mu\nu}| = a$ и $B \in \mathscr{M}$. Тогда $\| A \| = a \| B \| \geqslant ap_1$. Объединяя это с неравенством (6.1.1), получаем

$$ap_1 \leqslant \| A \| \leqslant ap_2.$$

---

*) Матрицу A следует считать ненулевой. — *Прим. перев.*

**) См. дополнение 1.

Подобно этому существуют такие положительные числа q_1 и q_2, зависящие лишь от нормы N, что

$$aq_1 \leqslant N(A) \leqslant aq_2.$$

Эти неравенства означают, что

$$r_1 \leqslant \frac{\| A \|}{N(A)} \leqslant r_2,$$

где $r_1 = p_1/q_2$, $r_2 = p_2/q_1$.

Упр. 6. Если $A_1, A_2, A_3, \ldots$ — последовательность матриц из $\mathscr{C}_{n\times n}$, то при любой обобщенной матричной норме $A_p \to A$ при $p \to \infty \Longleftrightarrow \| A_p - A \| \to 0$ при $p \to \infty$.

[Указание. Доказать результат для нормы из упр. 2 выше и затем воспользоваться теоремой 6.1.2.]

Упр. 7. Если $A_1, A_2, A_3, \ldots \in \mathscr{C}_{n\times n}$ и $B_N = \sum\limits_{p=1}^{N} A_p$ для каждого положительного целого числа N, то доказать, что $\sum\limits_{p=1}^{\infty} A_p$ сходится к матрице $A \Longleftrightarrow$ при любой матричной норме $\| B_N - A \| \to 0$ при $N \to \infty$.

Наиболее полезные матричные нормы — это часто те, которые просто определяются в терминах элементов матричного аргумента. Возможно, что мера величины для матриц могла бы быть основана на величинах их собственных значений, хотя бы это могло оказаться не очень полезным практически. Если $\mu_1, \ldots, \mu_n$ — собственные значения матрицы A, то $\mu_A = \max\limits_j |\mu_j|$ $(1 \leqslant j \leqslant n)$ будет такой мерой величины и известна как *спектральный радиус* A.

Рассмотрим матрицу H_n из упр. 3 § 5.3. Как легко видеть, все ее собственные значения нулевые и, следовательно, спектральный радиус для H_n равен нулю. Так как H_n — ненулевая матрица, мы убеждаемся в том, что спектральный радиус не удовлетворяет аксиоме (i) для матричной нормы. Однако, хотя спектральный радиус не является матричной нормой, он тесно связан с величиной матричных норм. Действительно, матричные нормы часто используются при получении границ для спектрального радиуса. Следующая теорема — первый результат такого рода.

Теорема 6.1.3. *Если $A \in \mathscr{C}_{n\times n}$ и μ_A — спектральный радиус матрицы A, то при любой матричной норме имеем $\| A \| \geqslant \mu_A$.*

Доказательство. Пусть μ — собственное значение матрицы A с $\mu^A = |\mu|$. Тогда существует такой вектор $\boldsymbol{x} \neq \boldsymbol{0}$, что

$Ax = \mu x$. Определим $n \times n$-матрицу

$$A_x = \| x \ 0 \ 0 \ \ldots \ 0 \|.$$

Тогда, очевидно,

$$AA_x = \mu A_x.$$

Используя аксиомы (ii) и (iv) для матричной нормы, получаем

$$|\mu| \| A_x \| \leqslant \|A\| \| A_x \|,$$

а так как $A_x \neq 0$, аксиома (i) означает, что $\| A_x \| \neq 0$. Следовательно,

$$|\mu| = \mu_A \leqslant \| A \|. \blacktriangleleft$$

В следующей последовательности упражнений мы вводим большое многообразие норм.

Упр. 8 (евклидова норма, или норма Фробениуса). Доказать, что функция

$$\| A \| = \left\{ \sum_{i,j} | a_{ij} |^2 \right\}^{1/2}$$

является нормой.

Решение. Читатель без затруднений проверит аксиомы (i) и (ii) для матричной нормы. Что касается аксиомы (iii), то заметим сначала, что

$$| a_{ij} + b_{ij} |^2 = (a_{ij} + b_{ij}) (\bar{a}_{ij} + \bar{b}_{ij}) =$$
$$= | a_{ij} |^2 + | b_{ij} |^2 + 2 \operatorname{Re} (a_{ij} \bar{b}_{ij}) \leqslant | a_{ij} |^2 + | b_{ij} |^2 + 2 | a_{ij} b_{ij} |.$$

Таким образом,

$$\| A + B \|^2 = \sum_{i,j} | a_{ij} + b_{ij} |^2 \leqslant \| A \|^2 + \|B\|^2 + 2 \left\{ \sum_{i,j} | a_{ij} b_{ij} | \right\}.$$

Используя неравенство Шварца, получаем

$$\| A + B \|^2 \leqslant \| A \|^2 + \| B \|^2 + 2 \left\{ \sum_{i,j} | a_{ij} |^2 \sum_{i,j} | b_{ij} |^2 \right\}^{1/2} = (\| A \| + \| B \|)^2.$$

Следовательно, $\| A + B \| \leqslant \| A \| + \| B \|$ и аксиома (iii) выполняется.

Для аксиомы (iv) получаем

$$\| AB \|^2 = \sum_{i,j} \left| \sum_k a_{ik} b_{kj} \right|^2 \leqslant \sum_{i,j} \left(\sum_k | a_{ik} |^2 \right) \left(\sum_k | b_{kj} |^2 \right),$$

опять воспользовавшись неравенством Шварца, а следовательно,

$$\| AB \|^2 \leqslant \left(\sum_{i\ \ k} | a_{ik} |^2 \right) \left(\sum_{j,\ k} | b_{kj} |^2 \right) = \| A \|^2 \| B \|^2.$$

Аксиома (iv) следует отсюда немедленно.

Упр. 9. Доказать, что $M(A) = n \max\limits_{i,j} |a_{ij}|$ будет матричной нормой.

Упр. 10. Если $\|\cdot\|$ обозначает евклидову норму и M такова, как определена в упр. 9, то доказать, что $\|A\| \leqslant M(A)$.

Упр. 11. Если $\|\cdot\|$ обозначает евклидову норму, то доказать, что $\|A\|^2 = \operatorname{tr}(A^*A)$.

Упр. 12 (нормы Гёльдера, или p-нормы). Функция

$$\|A\|_p = \left\{ \sum_{i,j} |a_{ij}|^p \right\}^{1/p}$$

будет обобщенной матричной нормой для $p \geqslant 1$ и матричной нормой в том и только в том случае, когда $1 \leqslant p \leqslant 2$.

(Читатель должен быть хорошо знаком с обобщением неравенства Шварца, известным как неравенство Гёльдера, перед тем, как приступить к этому упражнению.)

Упр. 13. Если $p \geqslant 2$ и $(1/p) + (1/q) = 1$, то

$$\|AB\|_p \leqslant \min\{\|A\|_p \|B\|_q, \|A\|_q \|B\|_p\}.$$

6.2. Векторные нормы

Мы увидим, что, как и для матриц, существует несколько полезных мер для величины векторов в $\mathscr{C}_n$. Длина такого вектора, как она определена в § 1.3, как раз является одной из полезных мер, удовлетворяющей следующим аксиомам. Эти аксиомы следует сравнить с аксиомами для обобщенной матричной нормы.

Вещественнозначная функция h, определенная на $\mathscr{C}_n$, называется *векторной нормой* на $\mathscr{C}_n$, если для всех $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{C}_n$:

(а) $h(\boldsymbol{x}) \geqslant 0$ и $h(\boldsymbol{x}) = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}$;

(б) $h(c\boldsymbol{x}) = |c| h(\boldsymbol{x})$ для любого комплексного числа c;

(в) $h(\boldsymbol{x} + \boldsymbol{y}) \leqslant h(\boldsymbol{x}) + h(\boldsymbol{y})$.

Упр. 1. Если h — векторная норма, то доказать, что $h(\boldsymbol{x} - \boldsymbol{y}) \geqslant |h(\boldsymbol{x}) - h(\boldsymbol{y})|$.

***Упр. 2.** Если $\boldsymbol{x}' = \| x_1, x_2, \ldots, x_n \|$, то доказать, что следующие функции будут векторными нормами:

(а) $\rho(\boldsymbol{x}) = \max\limits_j |x_j|$;

(б) $\gamma(\boldsymbol{x}) = \sum\limits_j |x_j|$;

(в) $h(\boldsymbol{x}) = \left\{ \sum\limits_j |x_j|^2 \right\}^{1/2} = \langle \boldsymbol{x}, \boldsymbol{x} \rangle$ (евклидова норма);

(г) $h_p(\boldsymbol{x}) = \left\{ \sum\limits_j |x_j|^p \right\}^{1/p}$, $p \geqslant 1$ (нормы Гёльдера).

***Упр. 3.** Доказать, что евклидова векторная норма инвариантна относительно унитарных преобразований. (Как мы видели в упр. 3 из § 2.13, ортогональное преобразование в $\mathscr{R}_3$ соответствует вращению осей координат вокруг оси, проходящей через начало координат. Мы вправе ожидать, что такое преобразование оставляет длину вектора инвариантной.)

Упр. 4. Если $P \in \mathscr{C}_{n\times n}$ — эрмитова положительно определенная матрица, то доказать, что

$$h(\boldsymbol{x}) = \langle \boldsymbol{x}, P\boldsymbol{x}\rangle^{1/2}$$

будет векторной нормой.

***Упр. 5.** Доказать аналоги теорем 6.1.1 и 6.1.2 для векторных норм.

Упр. 6. Сформулировать и доказать аналоги упражнений 6 и 7 из § 6.1, применимые к векторам и векторным нормам.

Упр. 7. Если h — векторная норма на $\mathscr{C}_n$ и $A \in \mathscr{C}_{n\times n}$ неособая, то доказать, что функция g, определенная как $g(\boldsymbol{x}) = h(A\boldsymbol{x})$, тоже будет векторной нормой.

Очень часто встречается задача нахождения нормы вектора, заданного в виде $A\boldsymbol{x}$, где $A \in \mathscr{C}_{n\times n}$ и $\boldsymbol{x} \in \mathscr{C}_n$. Приобретенный в связи с аксиомой (iv) для матричной нормы опыт позволяет ожидать, что будут существовать матричные нормы $\|\cdot\|$ и векторные нормы h, для которых $h(A\boldsymbol{x}) \leqslant \|A\| h(\boldsymbol{x})$. Если это так, то мы говорим, что нормы согласованы. Дадим точное определение: векторная норма h и матричная норма $\|\cdot\|$ *согласованы*, если для всех $\boldsymbol{x} \in \mathscr{C}_n$ и для всех $A \in \mathscr{C}_{n\times n}$ имеем $h(A\boldsymbol{x}) \leqslant \|A\| h(\boldsymbol{x})$.

Упр. 8. Матричная норма $M(A)$ из упр. 9 в § 6.1 согласована с (а) $\rho(\boldsymbol{x})$, (б) $\gamma(\boldsymbol{x})$, (в) $h(\boldsymbol{x})$, которые определены в упр. 2 выше.

Решение.

$$\text{(а)}\quad \rho(A\boldsymbol{x}) = \max_j \left|\sum_{k=1}^n a_{jk}x_k\right| \leqslant \max_j \left(\sum_{k=1}^n |a_{jk}|\,|x_k|\right) \leqslant$$
$$\leqslant \max_j \left(\sum_{k=1}^n |a_{jk}|\right)\rho(\boldsymbol{x}) \leqslant n\,(\max_{j,k}|a_{jk}|)\,\rho(\boldsymbol{x}) = M(A)\,\rho(\boldsymbol{x}).$$

$$\text{(б)}\quad \gamma(A\boldsymbol{x}) = \sum_j \left|\sum_{k=1}^n a_{jk}x_k\right| \leqslant \sum_{j,k} |a_{jk}|\,|x_k| \leqslant$$
$$\leqslant \sum_{j,k} \frac{M(A)}{n}|x_k| = M(A)\,\gamma(\boldsymbol{x}).$$

$$\text{(в)}\quad h^2(A\boldsymbol{x}) = \sum_j \left|\sum_k a_{jk}x_k\right|^2 \leqslant \sum_j \left(\sum_k |a_{jk}|^2\right)\left(\sum_k |x_k|^2\right)$$

по неравенству Шварца. Следовательно, если $\|\cdot\|$ обозначает евклидову матричную норму, то

$$h^2(A\boldsymbol{x}) \leqslant \left(\sum_{j,k}|a_{jk}|^2\right)\left(\sum_k |x_k|^2\right) = \|A\|^2 h^2(\boldsymbol{x}) \leqslant M^2(A)\,h^2(\boldsymbol{x})$$

по упр. 10 из § 6.1. Заметим, что попутно доказана согласованность евклидовых матричной и векторной норм.

Упр. 9. Пусть $\|\cdot\|$ обозначает обобщенную матричную норму, не являющуюся матричной нормой. Доказать, что не существует векторной нормы, согласованной с $\|\cdot\|$ *).

В обычной геометрии пространства $\mathcal{R}_3$ единичная сфера — это множество векторов $\boldsymbol{x} \in \mathcal{R}_3$, для которых $x_1^2 + x_2^2 + x_3^2 = 1$, т. е. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 1$. В более общем случае, когда мы вводим некоторую норму, или длину, в $\mathscr{C}_n$, единичная сфера будет зависеть от выбора нормы. Так, в пространстве $\mathscr{C}_n$ с векторной нормой h *единичная сфера* $\mathcal{O}_h$ — это множество всех векторов $\boldsymbol{x} \in \mathscr{C}_n$, для которых $h(\boldsymbol{x}) = 1$. В частности, использованное в § 3.2 определение дает единичную сферу в $\mathscr{C}_n$ с евклидовой нормой. В более общем случае сфера $\mathscr{S}_h$ в $\mathscr{C}_n$ с центром $\boldsymbol{x}_0$ и радиусом r — это множество всех векторов $\boldsymbol{x}$, для которых $h(\boldsymbol{x} - \boldsymbol{x}_0) = r$.

Поучительно рассмотреть единичные сферы в $\mathcal{R}_2$ для векторных норм, определенных выше в упр. 2. Однако, прежде чем приступить к этому, следует заметить, что все определенные там нормы могут считаться нормами Гёльдера. Ясно, что γ и h — нормы Гёльдера с $p = 1$ и $p = 2$ соответственно. Кроме того, мы утверждаем, что ρ соответствует норме Гёльдера с $p = \infty$. Чтобы убедиться в этом, предположим, что вектор $\boldsymbol{x}$ имеет ν компонент абсолютной величины $\rho(\boldsymbol{x}) = |x_j|$, и запишем затем

$$h_p(\boldsymbol{x}) = |x_j| \{\nu + \varepsilon_1^p + \ldots + \varepsilon_{n-\nu}^p\}^{1/p},$$

где $\varepsilon_i = |x_k|/|x_j| < 1$ для некоторого k и $i = 1, 2, \ldots, n - \nu$. Теперь уже нетрудно видеть, что $h_p(\boldsymbol{x}) \to |x_j| = \rho(\boldsymbol{x})$ при $p \to \infty$.

Возвращаясь теперь к единичным сферам в $\mathcal{R}_2$, читатель легко проверит, что части единичных сфер в первом квадранте для норм γ, h и ρ таковы, как они показаны на рис. 4. Мы предоставим читателю исследовать вид единичных сфер для норм Гёльдера с промежуточными значениями p. Читателю рекомендуется также представить себе вид единичных сфер в $\mathcal{R}_3$ с нормами Гёльдера.

В общем случае симметрии единичной сферы будут зависеть от симметрий функции $h(\boldsymbol{x}) = H(x_1, x_2, \ldots, x_n)$, если ее рассмотреть как функцию H от n скалярных переменных $x_1, \ldots, x_n$. Нормы Гёльдера все являются симметрическими функциями от компонент $x_1, x_2, \ldots, x_n$. Это эквивалентно тому, что $h_p(P\boldsymbol{x}) = h_p(\boldsymbol{x})$ для любого x и любой перестановочной матрицы P. Упражнение 10 дает простой пример векторной нормы, которая не является симметрической функцией от компонент вектора.

*) В присланном письме автор сообщает, что он не имеет решения этого упражнения. — *Прим. перев.*

***Упр. 10.** Для $x \in \mathscr{C}_n$ положим $h(x) = |x_1| + \ldots + |x_{n-1}| + n^{-1}|x_n|$. Доказать, что h — векторная норма на $\mathscr{C}_n$. Каков вид единичных сфер для $\mathscr{R}_2$ и $\mathscr{R}_3$ с нормой h?

Какую бы векторную норму ни ввести в $\mathscr{C}_n$, множество точек на и внутри единичной сферы, как можно доказать, будет обладать свойством выпуклости. В $\mathscr{R}_2$ или $\mathscr{R}_3$ не представляет затруднений убедиться в этом, и утверждение звучит правдоподобно. Для придачи ему смысла в $\mathscr{C}_n$ мы должны привести сначала несколько определений.

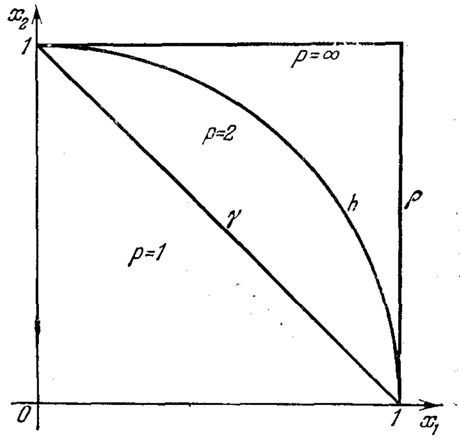

Рис. 4. Единичные сферы в $\mathscr{R}_2$ для p-норм.

Прежде всего, если $x, y \in \mathscr{C}_n$, то *прямой сегмент*, соединяющий x и y, определяется как множество точек в $\mathscr{C}_n$, имеющих вид $tx + (1-t)y$ для некоторого t в замкнутом отрезке $[0, 1]$. Это естественно обобщает понятие, используемое в $\mathscr{R}_2$ и $\mathscr{R}_3$. Далее, множество векторов $\mathscr{L} \subseteq \mathscr{C}_n$ *выпукло*, если для каждой пары x, $y \in \mathscr{L}$ прямой сегмент, соединяющий x и y, тоже содержится в $\mathscr{L}$.

Наконец, в пространстве $\mathscr{C}_n$ с векторной нормой h будем называть множество точек на и внутри единичной сферы *единичным шаром* (в обозначении $\mathscr{B}_h$), т. е. $\mathscr{B}_h$ состоит из всех $x \in \mathscr{C}_n$ с $h(x) \leqslant 1$.

Теорема 6.2.1. *Если h — векторная норма в $\mathscr{C}_n$, то единичный шар $\mathscr{B}_h$ будет выпуклым множеством.*

Доказательство. Пусть x, y — любые два вектора в $\mathscr{B}_h$. Тогда

$$z = tx + (1-t)y, \qquad 0 \leqslant t \leqslant 1,$$

— типичный представитель прямого сегмента, соединяющего x и y. Мы должны лишь доказать, что $z \in \mathscr{B}_h$. Имеем (используя аксиомы (б) и (в) для векторной нормы)

$$h(z) \leqslant th(x) + (1-t)h(y)$$

и $h(x)$, $h(y) \leqslant 1$, так как x, $y \in \mathscr{B}_h$. Следовательно, $h(z) \leqslant t + (1-t) = 1$, так что $z \in \mathscr{B}_h$. ◀

***Упр. 11.** Множество векторов x в $\mathscr{C}_n$, для которых $h(x - x_0) \leqslant r$, есть *шар* с центром x_0 и радиусом r. Доказать, что шар — выпуклое множество.

***Упр. 12.** Пусть Z_k — некоторое множество k различных векторов в $\mathscr{C}_n$. Наименьшее выпуклое множество $H(Z_k)\subseteq\mathscr{C}_n$ с $Z_k\subseteq H(Z_k)$ называется *выпуклой оболочкой* Z_k. (Слово «наименьшее» здесь означает, что если $Z_k\subseteq\mathscr{T}\subseteq\mathscr{C}_n$ и $\mathscr{T}$ выпукло, то $H(Z_k)\subseteq\mathscr{T}$.) Если $z_1, z_2, \ldots, z_k$ — элементы из Z_k, то доказать, что $z\in H(Z_k)\Longleftrightarrow$ существуют числа $\theta_i\geqslant 0$, $i=1, 2, \ldots, k$, для которых

$$\sum_{i=1}^{k}\theta_i=1 \quad \text{и} \quad \boldsymbol{z}=\sum_{i=1}^{k}\theta_i\boldsymbol{z}_i.$$

Решение. Пусть $\mathscr{L}_k$ — множество векторов вида $\sum_{i=1}^{k}\theta_i\boldsymbol{z}_i$ с $\theta_i\geqslant 0$, $\sum\theta_i=1$. Нетрудно доказать, что $\mathscr{L}_k$ выпукло и что $Z_k\subseteq\mathscr{L}_k$.

Пусть теперь $\mathscr{T}$ — *любое* выпуклое множество в $\mathscr{C}_n$ с $Z_k\subseteq\mathscr{T}$. Мы получим результат, доказав, что $\mathscr{L}_k\subseteq\mathscr{T}$ и, следовательно, что $\mathscr{L}_k=H(Z_k)$. Доказательство проводится по индукции относительно k и в случае $k=2$ следует непосредственно из определения выпуклого множества.

Читателю предоставляется доказать, что из $\mathscr{L}_{k-1}\subseteq\mathscr{T}$ и $z_k\in\mathscr{T}$ следует $\mathscr{L}_k\subseteq\mathscr{T}$.

6.3. Индуцированные матричные нормы

Мы исследуем теперь способ получения некоторой матричной нормы из любой заданной векторной нормы. Если h — векторная норма в $\mathscr{C}_n$ и $A\in\mathscr{C}_{n\times n}$, то рассмотрим отношение $h(A\boldsymbol{x})/h(\boldsymbol{x})$, определенное для каждого ненулевого $\boldsymbol{x}\in\mathscr{C}_n$. Это отношение, очевидно, неотрицательно. Рассмотрим множество всех чисел, получаемых из этого отношения, когда $\boldsymbol{x}$ принимает все возможные ненулевые значения в $\mathscr{C}_n$, и определим *)

$$\|A\|=\sup_{\boldsymbol{x}\neq 0}\frac{h(A\boldsymbol{x})}{h(\boldsymbol{x})}. \tag{6.3.1}$$

Мы докажем, что для фиксированной матрицы A это неотрицательное число определено и что получающаяся в результате функция $\|\cdot\|$ будет матричной нормой. Мы будем называть эту матричную норму *индуцированной* h. Следует обратить внимание на содержательную аналогию введенного отношения с отношением Релея (для эрмитовых матриц), определенным в § 3.2.

Следующие результаты непосредственно получаются из дополнения 1.

*) «sup» — сокращение для супремума, или наименьшей верхней грани.

Лемма. *Для функции* $\|\cdot\|$, *определенной на* $\mathscr{C}_{n\times n}$ *по* (6.3.1), *имеем*

(i) $\|A\| = \max\limits_{h(\boldsymbol{x})=1} (A\boldsymbol{x})$

и

(ii) *существует такой вектор* $\boldsymbol{x}_0$, *зависящий от* A, *что* $h(\boldsymbol{x}_0) = 1$ *и* $\|A\| = h(A\boldsymbol{x}_0)$.

Доказательство. Заметим сначала, что замена $\boldsymbol{x}$ на $c\boldsymbol{x}$ не меняет отношения $h(A\boldsymbol{x})/h(\boldsymbol{x})$. Таким образом, полный образ значений функции $\|\cdot\|$ останется тем же, если предположить вектор $\boldsymbol{x}$ нормализованным, так что $h(\boldsymbol{x}) = 1$. Поэтому $\|A\| = \sup h(A\boldsymbol{x})$, где наименьшая верхняя грань берется по единичной сфере $\mathscr{O}_h$.

Как мы видели (упр. 5 из § 6.2), векторная норма непрерывно зависит от компонент векторного аргумента. Кроме того, $\mathscr{O}_h$ — замкнутое и ограниченное множество, а поэтому обе части леммы непосредственно следуют из дополнения 1. В части (ii) просто утверждается, что существует $\boldsymbol{x}_0 \in \mathscr{O}_h$, для которой максимум достигается.

Значение индуцированных норм частично объясняется частью (в) следующей теоремы.

Теорема 6.3.1. *Если* $\mathscr{O}_h$ — *единичная сфера в* $\mathscr{C}_n$ *с векторной нормой* h, *то*:

(а) $\|A\| = \max\limits_{\boldsymbol{x}\in\mathscr{O}_h} h(A\boldsymbol{x})$ — *матричная норма*;

(б) h *и* $\|\cdot\|$ *согласованы*;

(в) *если* N — *любая согласованная с* h *матричная норма, то* $\|A\| \leqslant N(A)$ *для всех* $A \in \mathscr{C}_{n\times n}$.

Доказательство. 1. Мы намереваемся установить выполнимость аксиом для матричной нормы. Относительно аксиомы (i) ясно, что $\|A\| \geqslant 0$ и что $\|A\| = 0$ при $A = 0$. Если $A \neq 0$, то существует такой $\boldsymbol{x} \in \mathscr{C}_n$, что $A\boldsymbol{x} \neq \boldsymbol{0}$ и $h(\boldsymbol{x}) = 1$. Следовательно, $\|A\| > 0$. Аксиома (i) поэтому выполняется.

Используя аксиому (б) для векторной нормы, получаем

$$\|cA\| = \max_{\boldsymbol{x}\in\mathscr{O}_h} h(cA\boldsymbol{x}) = |c| \max_{\boldsymbol{x}\in\mathscr{O}_h} h(A\boldsymbol{x}) = |c|\,\|A\|,$$

так что аксиома (ii) тоже выполняется.

Имеем тогда

$$h((A+B)\boldsymbol{x}) = h(A\boldsymbol{x} + B\boldsymbol{x}) \leqslant h(A\boldsymbol{x}) + h(B\boldsymbol{x}),$$

откуда

$$\max_{\boldsymbol{x}\in\mathscr{O}_h} h((A+B)\boldsymbol{x}) \leqslant \max\{h(A\boldsymbol{x}) + h(B\boldsymbol{x})\} \leqslant \max h(A\boldsymbol{x}) + \max h(B\boldsymbol{x}).$$

Таким образом, $\|A+B\| \leqslant \|A\| + \|B\|$ и получена аксиома (iii).

2. Мы отступим теперь от естественного хода доказательства и докажем часть (б) теоремы. Для любого $\boldsymbol{x} \neq \boldsymbol{0}$ определим $\boldsymbol{\xi} = \boldsymbol{x}/h(\boldsymbol{x})$, так что $h(\boldsymbol{\xi}) = 1$. Тогда

$$h(A\boldsymbol{x}) = h(h(\boldsymbol{x})A\boldsymbol{\xi}) = h(\boldsymbol{x})\,h(A\boldsymbol{\xi}) \leqslant h(\boldsymbol{x}) \max_{y \in \mathcal{O}_h} h(A\boldsymbol{y}) = h(\boldsymbol{x})\,\|A\|,$$

и это есть условие того, что h и $\|\cdot\|$ согласованы (хотя мы и не утверждаем пока, что $\|\cdot\|$ — матричная норма).

3. Возвратимся к аксиоме (iv) для матричной нормы. Из части (ii) леммы выводим, что существует вектор $\boldsymbol{x}_0$ с $h(\boldsymbol{x}_0) = 1$ и $h(AB\boldsymbol{x}_0) = \|AB\|$ при заданных $A, B \in \mathscr{C}_{n\times n}$. Тогда

$$\|AB\| = h(A(B\boldsymbol{x}_0)) = h(AB\boldsymbol{x}_0) \leqslant \|A\|\,h(B\boldsymbol{x}_0),$$

ввиду только что доказанной согласованности h и $\|\cdot\|$. Используя этот результат еще раз, получаем

$$\|AB\| \leqslant \|A\|\,\|B\|\,h(\boldsymbol{x}_0) = \|A\|\,\|B\|.$$

Таким образом, аксиома (iv) выполняется и части (а) и (б) теоремы доказаны.

4. Для доказательства части (в) предположим, что $h(\boldsymbol{y}_0) = 1$ и $h(A\boldsymbol{y}_0) = \|A\|$. Тогда

$$\|A\| = h(A\boldsymbol{y}_0) \leqslant N(A)\,h(\boldsymbol{y}_0) = N(A). \blacktriangleleft$$

Некоторые из наиболее важных и полезных матричных норм — это индуцированные нормы, обсуждаемые в следующем ряде упражнений.

Прежде чем оставить идею индуцирования матричных норм из векторных, мы можем поставить обратный вопрос: можно ли использовать некоторую заданную матричную норму для индуцирования согласованной векторной нормы? Это можно сделать, определяя h как

$$h(\boldsymbol{x}) = \|\boldsymbol{x}\boldsymbol{a}'\| \tag{6.3.2}$$

для некоторого фиксированного вектора $\boldsymbol{a} \neq \boldsymbol{0}$.

Упр. 1. Доказать, что функция h, определяемая равенством (6.3.2), является векторной нормой.

Упр. 2. Пусть $\|\cdot\|$ обозначает матричную норму, индуцированную векторной нормой h, и пусть H_n определено, как в § 4.13. Доказать, что $\|I\| = 1$. Если дополнительно h — симметрическая функция от абсолютных значений компонент ее векторного аргумента, то доказать, что $\|H_n\| = 1$.

Упр. 3. Пусть h — векторная норма, определенная в упр. 10 § 6.2. Доказать, что если $\|\cdot\|$ — индуцированная h матричная норма, то $\|H_n\| = n$. Найти векторную норму, для которой $\|H_n\| < 1$ в индуцированной норме.

Упр. 4. Пусть $\|\cdot\|$ обозначает матричную норму, индуцированную евклидовой нормой (эта норма известна как *спектральная*), и для $A \in \mathscr{C}_{n\times n}$ пусть λ_A будет спектральным радиусом матрицы A^*A. Доказать, что $\|A\| = \lambda_A^{1/2}$.

Решение. Матрица A^*A эрмитова и определенная, так как

$$(A^*A)^* = A^*A$$

и

$$x^*(A^*A)x = (Ax)^* Ax = \langle Ax, Ax\rangle \geqslant 0$$

для всех $x \in \mathscr{C}_n$. Из теоремы 2.14.1 тогда следует, что собственные значения для A^*A вещественны и неотрицательны. Следовательно, на самом деле λ_A будет собственным значением матрицы A^*A.

Пусть $x_1, x_2, \ldots, x_n$ — множество ортонормированных правых собственных векторов для A^*A с соответствующими собственными значениями $\lambda_1, \lambda_2, \ldots, \lambda_n$. Пусть h обозначает евклидову векторную норму и для любого x с $h(x) = 1$ запишем $x = \sum_{j=1}^{n} \xi_j x_j$. Тогда

$$A^*Ax = \sum_{j=1}^{n} \xi_j \lambda_j x_j$$

и

$$h^2(Ax) = (Ax)^* Ax = x^*(A^*Ax) = \Big(\sum_j \xi_j x_j\Big)^* \Big(\sum_k \xi_k \lambda_k x_k\Big) = \sum_j |\xi_j|^2 \lambda_j,$$

если воспользоваться тем, что $x_j^* x_k = \delta_{jk}$. Таким образом,

$$h(Ax) = \Big\{\sum_j |\xi_j|^2 \lambda_j\Big\}^{1/2}, \quad \sum_j |\xi_j|^2 = 1,$$

и

$$\lambda_j \geqslant 0, \quad j = 1, 2, \ldots, n.$$

Отсюда непосредственно следует, что

$$\|A\| = \max_{x \in \mathcal{O}_n} h(Ax) = \lambda_A^{1/2}.$$

Упр. 5. Если A — унитарная матрица, то доказать, что спектральная норма A равна 1.

Упр. 6. Если $A = \operatorname{diag}\{d_1, d_2, \ldots, d_n\}$ и $\|\cdot\|$ — матричная норма, индуцированная любой из норм (а), (б), (в) в упр. 2 § 6.2, то доказать, что $\|A\| = \max_j |d_j|$.

Упр. 7. Если A, $U \in \mathscr{C}_{n\times n}$ и U унитарна, то доказать, что для евклидовой и унитарной норм $\|A\| = \|UA\| = \|AU\|$ и, следовательно, что для этих двух норм $\|A\|$ инвариантна при унитарно подобных преобразованиях.

Упр. 8. Если A — нормальная матрица и $\|\cdot\|$ обозначает спектральную норму, то доказать, что $\|A\| = \mu_A$ — спектральному радиусу A. (Сравнить этот результат с теоремой 6.1.3.) Если f определена на спектре матрицы A, то доказать, что $\|f(A)\|$ равно спектральному радиусу $f(A)$.

***Упр. 9.** Показать, что матричная норма, индуцированная векторной нормой $\rho(\boldsymbol{x}) = \max\limits_j |x_j|$, задается в виде

$$\|A\|_\rho = \max_j \sum_k |a_{jk}|.$$

(Заметим, что берутся суммы по *строкам абсолютных значений* матричных элементов, а затем выбирается наибольшая из этих сумм.)

Решение. Пусть $\boldsymbol{x}$ — вектор с $\rho(\boldsymbol{x}) = 1$. Тогда

$$\rho(A\boldsymbol{x}) = \max_j \left| \sum_k a_{jk} x_k \right| \leqslant \max_j \sum_k |a_{jk}| |x_k| \leqslant$$

$$\leqslant \rho(\boldsymbol{x}) \max_j \sum_k |a_{jk}| = \|A\|_\rho.$$

Таким образом, $\max\limits_{\rho(\boldsymbol{x})=1} \rho(A\boldsymbol{x}) \leqslant \|A\|_\rho$, так что индуцированная ρ норма не может превзойти $\|\cdot\|_\rho$. Чтобы доказать, что они равны, мы должны лишь показать, что существует ξ с $\rho(\xi) = 1$ и $\rho(A\xi) \geqslant \|A\|_\rho$.

Предположим, что $\|A\|_\rho = \sum\limits_k |a_{mk}|$, и рассмотрим вектор ξ с элементами

$$\xi_k = \begin{cases} |a_{mk}|/a_{mk}, & \text{если} \quad a_{mk} \neq 0, \\ 0, & \text{если} \quad a_{mk} = 0, \end{cases}$$

для $k = 1, 2, \ldots, n$. Тогда $\rho(\xi) = 1$, и если $A\xi = \eta$, то

$$\eta_m = \sum_k a_{mk} \xi_k = \sum_k |a_{mk}| = \|A\|_\rho$$

по определению ξ. Таким образом,

$$\rho(A\xi) = \max_j |\eta_j| \geqslant \eta_m = \|A\|_\rho,$$

откуда следует, что

$$\|A\|_\rho = \max_{\rho(\boldsymbol{x})=1} \rho(A\boldsymbol{x}).$$

Упр. 10. Показать, что матричная норма, индуцированная векторной нормой $\gamma(\boldsymbol{x}) = \sum\limits_j |x_j|$, задается в виде

$$\|A\|_\gamma = \max_k \sum_j |a_{jk}|.$$

(Заметим, что берутся суммы по *столбцам абсолютных значений* матричных элементов, а затем выбирается наибольшая из этих сумм.)

Упр. 11. Если $\|\cdot\|_s$ обозначает спектральную норму и $\|\cdot\|_\rho$, $\|\cdot\|_\gamma$ определены в упр. 9 и 10, то доказать, что

$$\|A\|_s^2 \leqslant \|A\|_\rho \|A\|_\gamma.$$

Решение. Используя результат упр. 4, имеем $\|A\|_s = \lambda_A^{1/2}$, где λ_A — максимальное собственное значение матрицы A^*A, и по теореме 6.1.3 $\lambda_A \leqslant \|A^*A\|_\gamma$. Таким образом,

$$\|A\|_s^2 \leqslant \|A^*A\|_\gamma \leqslant \|A^*\|_\gamma \|A\|_\gamma = \|A\|_\rho \|A\|_\gamma.$$

6.4. Абсолютные векторные нормы

Читатель, по-видимому, обратил внимание на то, что почти все рассмотренные в упражнениях векторные нормы зависят лишь от абсолютных значений элементов векторного аргумента. Такие нормы называются *абсолютными* векторными нормами. В общей теории норм, которую мы до сих пор излагали, не предполагается, что обязательно имеет место этот случай. Этот параграф будет посвящен доказательству двух свойств абсолютных векторных норм. Докажем сначала одну лемму, которая часто бывает полезна и сама по себе.

Лемма. *Если h — абсолютная векторная норма и мы записали*

$$h(\boldsymbol{x}) \equiv H(x_1, x_2, \ldots, x_n),$$

то H — неубывающая функция от абсолютных значений величин $x_1, x_2, \ldots, x_n$.

Доказательство. Сосредоточим внимание на H как на функции первой компоненты x_1. То же рассуждение будет применимо к каждой из n переменных. Таким образом, мы предполагаем $x_2, \ldots, x_n$ фиксированными, полагаем

$$H(x, x_2, \ldots, x_n) = f(x)$$

и рассматриваем поведение f при изменяющемся на вещественной оси x.

Если утверждение леммы не выполняется, то существуют неотрицательные числа p, q, для которых $p < q$ и $f(p) > f(q)$. Но так как $f(x) = f(|x|)$, то $f(-q) = f(q)$, так что векторы

$$\boldsymbol{y}_1 = \begin{Vmatrix} -q \\ x_2 \\ \cdot \\ \vdots \\ x_n \end{Vmatrix} \quad \text{и} \quad \boldsymbol{y}_2 = \begin{Vmatrix} q \\ x_2 \\ \cdot \\ \vdots \\ x_n \end{Vmatrix}$$

принадлежат шару, состоящему из векторов $\boldsymbol{x}$, для которых $h(\boldsymbol{x}) \leqslant f(q)$. Но шар — обязательно выпуклое множество (упр. 12 § 6.2), а так как вектор

$$\boldsymbol{z} = \begin{Vmatrix} p \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{Vmatrix},$$

очевидно, принадлежит прямому отрезку, соединяющему $\boldsymbol{y}_1$ и $\boldsymbol{y}_2$, то $h(\boldsymbol{z}) \leqslant f(q)$. Но $h(\boldsymbol{z}) = f(p) > f(q)$. Полученное противоречие доказывает лемму.

Одно из свойств абсолютной векторной нормы, которое мы установим, называется свойством *монотонности*. Будем обозначать через $|\boldsymbol{x}|$ вектор, элементы которого — абсолютные значения элементов вектора $\boldsymbol{x}$, и при $\boldsymbol{y}, \boldsymbol{z} \in \mathscr{R}_n$ будем писать $\boldsymbol{y} \leqslant \boldsymbol{z}$, если это неравенство выполняется для соответствующих элементов $\boldsymbol{y}$ и $\boldsymbol{z}$. Подобное соглашение относится и к другим знакам неравенств. Векторная норма h называется монотонной тогда и только тогда, когда из $|\boldsymbol{x}| \leqslant |\boldsymbol{y}|$ следует $h(\boldsymbol{x}) \leqslant h(\boldsymbol{y})$. Лемма означает, что это условие непременно выполняется для абсолютных векторных норм. В следующей теореме мы покажем, что обратное утверждение тоже верно: монотонная норма будет обязательно абсолютной.

Второе из получаемых свойств относится к индуцированной матричной норме. Когда матричная норма вычисляется для диагональной матрицы $D = \operatorname{diag}\{d_1, \ldots, d_n\}$, мы считаем норму хорошей, если для нее выполняется свойство $\|D\| = \max\limits_j |d_j|$. Мы покажем, что $\|\cdot\|$ обладает этим свойством *тогда и только тогда, когда* она является нормой, индуцированной абсолютной векторной нормой.

Теорема 6.4.1. *Если h — векторная норма и $\|\cdot\|$ обозначает матричную норму, индуцированную h, то следующие условия эквивалентны*:

(i) *h абсолютная*;

(ii) *h монотонная*;

(iii) *$\|D\| = \max\limits_j |d_j|$ для любой диагональной матрицы $D = \operatorname{diag}\{d_1, \ldots, d_n\}$.*

Доказательство. Мы получим эквивалентность всех свойств, показав, что из (i) следует (ii), из (ii) следует (iii) и из (iii) следует (i).

1. Из леммы получаем, что из (i) следует (ii).

2. Для доказательства того, что из (ii) следует (iii), можем предположить $D \neq 0$. Определим $d = \max\limits_j |d_j|$. Ясно, что $|D\boldsymbol{x}| \leqslant$

$\leqslant |d\boldsymbol{x}|$, так что условие (ii) означает, что $h(D\boldsymbol{x}) \leqslant h(d\boldsymbol{x}) = dh(|\boldsymbol{x}|)$. Таким образом,

$$\|D\| = \sup_{\boldsymbol{x} \neq 0} \frac{h(D\boldsymbol{x})}{h(\boldsymbol{x})} \leqslant d.$$

Если мы сможем показать, что существует вектор $\boldsymbol{a}$, для которого $h(D\boldsymbol{a})/h(\boldsymbol{a}) = d$, то доказательство будет закончено. Предположим, что m — целое число, для которого $d = |d_m|$. Тогда

$$\frac{h(D\boldsymbol{e}_m)}{h(\boldsymbol{e}_m)} = \frac{h(d_m \boldsymbol{e}_m)}{h(\boldsymbol{e}_m)} = |d_m| = d.$$

Следовательно, $\sup\limits_{\boldsymbol{x} \neq 0} h(D\boldsymbol{x})/h(\boldsymbol{x}) = d$, или $\|D\| = d$. Таким образом, из (ii) следует (iii).

3. Предположим теперь, что выполняется условие (iii). Для любого вектора $\boldsymbol{x}$ существует такая диагональная матрица*) D с $|d_j| = 1$, $j = 1, 2, \ldots, n$, что $|\boldsymbol{x}| = D\boldsymbol{x}$. Свойство (iii) означает, что $\|D\| = \|D^{-1}\| = 1$. Имеем тогда

$$h(|\boldsymbol{x}|) \leqslant \|D\| h(\boldsymbol{x}) = h(\boldsymbol{x})$$

и, так как $\boldsymbol{x} = D^{-1}|\boldsymbol{x}|$,

$$h(\boldsymbol{x}) \leqslant \|D^{-1}\| h(|\boldsymbol{x}|) = h(|\boldsymbol{x}|).$$

Следовательно, $h(\boldsymbol{x}) = h(|\boldsymbol{x}|)$ для любого $\boldsymbol{x}$, так что из (iii) следует (i). Это завершает доказательство теоремы.

Упр. 1. Если $\|\cdot\|$ — матричная норма, индуцированная абсолютной векторной нормой, и $|A|$ — матрица с элементами $|a_{jk}|$, то показать, что $\|A\| \leqslant \||A|\|$. Будет ли эта матричная норма монотонна в том смысле, что из $|A| \leqslant |B|$ следует $\|A\| \leqslant \|B\|$?

6.5. Нижние грани

Мы начали обсуждение матричных норм с рассмотрения меры величины матрицы или меры удаления ее от нулевой матрицы. Мера удаления матрицы от вырожденности также была бы полезной. Таким образом, мы попытаемся найти функцию, определенную на всех квадратных матрицах A, которая неотрицательна и равна нулю в том и только в том случае, когда $\det A = 0$. С первого взгляда удивляет то, что такая мера может быть получена из определения, дополнительного к определению индуцированной нормы в (6.3.1). Однако это удивит меньше, если вспомнить, что $\boldsymbol{x} \neq \boldsymbol{0}$ с $A\boldsymbol{x} = \boldsymbol{0}$ существует в том и только в том случае, когда $\det A = 0$. Мы определяем

---

*) Полагаем $d_j = 1$ при $x_j = 0$ и $d_j = |x_j|/x_j$ при $x_j \neq 0$.

нижнюю грань матрицы A относительно векторной нормы h (в обозначении $\mathrm{glb}_h(A)$) по формуле

$$\mathrm{glb}_h(A) = \inf_{x \neq 0} \frac{h(Ax)}{h(x)}.$$

Если ясно, какая векторная норма подразумевается, мы будем писать $\mathrm{glb}(A)$. Следующий результат составляет пару с леммой § 6.3 и доказывается тем же самым способом.

Лемма. (i) $\mathrm{glb}_h(A) = \min\limits_{h(x)=1} h(Ax)$ *и*

(ii) *существует такой вектор* y_0, *зависящий от* A, *что* $h(y_0) = 1$ *и* $\mathrm{glb}_h(A) = h(Ay_0)$.

Если принять во внимание, что $\|A^{-1}\|$ неограниченно возрастает по величине, когда A стремится к особой матрице, то следующая теорема будет оправдывать нашу претензию на то, что нижняя грань является мерой удаления от вырожденности.

Теорема 6.5.1. *Если* $\|\cdot\|$ — *матричная норма, индуцированная векторной нормой* h, *то*

$$\mathrm{glb}_h(A) = \begin{cases} 1/\|A^{-1}\|, & \text{если} \quad \det A \neq 0, \\ 0, & \text{если} \quad \det A = 0. \end{cases}$$

Доказательство. 1. Если $\det A \neq 0$, то определим вектор y через x по формуле $y = Ax$. Тогда для любого $x \neq 0$ будет $y \neq 0$ и

$$\mathrm{glb}_h(A) = \inf \frac{h(Ax)}{h(x)} = \inf \frac{h(y)}{h(A^{-1}y)} = \left[\sup \frac{h(A^{-1}y)}{h(y)}\right]^{-1} = \frac{1}{\|A^{-1}\|}.$$

(При переходе от инфимума к супремуму использован результат о том, что если $\mathscr{A}$ — некоторое множество положительных чисел и $\mathscr{B}$ — множество чисел $1/a$, где $a \in \mathscr{A}$, то $\inf \mathscr{A} = 1/\sup \mathscr{B}$.)

2. Если $\det A = 0$, то существует такой вектор ξ, что $A\xi = 0$ и $h(\xi) = 1$. Тогда, согласно лемме, $\mathrm{glb}_h(A) = \min\limits_{h(x)=1} h(Ax) = 0$; минимум достигается при $x = \xi$.

Следствие 1. *Если* $A, B \in \mathscr{C}_{n \times n}$, *то*

$$\mathrm{glb}_h(AB) \geqslant \mathrm{glb}_h(A)\, \mathrm{glb}_h(B).$$

Доказательство этого следствия предоставляется читателю в качестве простого упражнения.

Следствие 2. *Если* h — *абсолютная норма и* $D = \mathrm{diag}\{d_1, \ldots, d_n\}$, *то*

$$\mathrm{glb}_h(D) = \min_j |d_j|.$$

Доказательство. Если D неособая, то $d_j \neq 0$ для каждого j и

$$\mathrm{glb}_h(D) = \frac{1}{\|D^{-1}\|} = \frac{1}{\max\limits_j (1/|d_j|)} = \min_j |d_j|$$

по теореме 6.4.1.

Если D особая, то $\mathrm{glb}_h(D) = 0$ по теореме и, очевидно, $\min\limits_j |d_j| = 0$.

Упр. 1. Пусть $\mu_1, \mu_2, \ldots, \mu_n$ — собственные значения матрицы A. Доказать, что для любой векторной нормы h имеем $\mathrm{glb}_h(A) \leqslant \min\limits_j |\mu_j|$. Доказать также, что неравенство переходит в равенство, если A нормальна и h — евклидова векторная норма. (Заметим, что, объединяя этот результат с теоремой 6.1.3, мы можем определить кольцеобразную область комплексной плоскости, внутри которой должны лежать собственные значения матрицы A.)

6.6. Поле значений

Если $A \in \mathscr{C}_{n\times n}$, то собственные значения A образуют множество из n точек (не обязательно различных) на комплексной плоскости. Некоторые полезные представления о распределении этих точек могут быть получены, если исходить из понятия «поля значений» матрицы A. Условимся сначала, что всюду в этом параграфе h будет обозначать евклидову векторную норму. Определим *поле значений* $F(A)$ матрицы $A \in \mathscr{C}_{n\times n}$ как множество всех чисел

$$\boldsymbol{x}^* A\boldsymbol{x} = \langle \boldsymbol{x}, A\boldsymbol{x}\rangle = \sum_{j,\,k=1}^{n} a_{jk} x_j \bar{x}_k,$$

где $\boldsymbol{x} \in \mathscr{C}_n$ и $h(\boldsymbol{x}) = 1$. Ссылаясь на § 3.2, мы можем также описать $F(A)$ как множество всех возможных значений отношения Релея для A.

Так как $F(A)$ — образ сферы $\mathcal{O}_h$ при непрерывной функции, то (теорема 2 дополнения 1) $F(A)$ будет замкнутым и ограниченным множеством на комплексной плоскости. Можно также доказать, что $F(A)$ — выпуклое множество, но мы этого доказательства приводить не будем. Ясно, что каждое собственное значение μ_j матрицы A принадлежит $F(A)$, $j = 1, 2, \ldots, n$, ибо если $A\boldsymbol{x}_j = \mu_j \boldsymbol{x}_j$ и $h(\boldsymbol{x}_j) = 1$, то

$$\mu_j = \boldsymbol{x}_j^* A \boldsymbol{x}_j \in F(A).$$

Теорема 6.6.1. *Поле значений матрицы A инвариантно при унитарно подобных преобразованиях. Таким образом, если U унитарна, то*

$$F(A) = F(UAU^*).$$

Доказательство. Если $z \in F(UAU^*)$, то существует $\boldsymbol{x} \in \mathscr{C}_n$ с $h(\boldsymbol{x}) = 1$ и $z = \boldsymbol{x}^* UAU^* \boldsymbol{x} = \boldsymbol{y}^* A \boldsymbol{y}$, где $\boldsymbol{y} = U^* \boldsymbol{x}$. Но $h(\boldsymbol{y}) = 1$ (упр. 3 § 6.2), так что $z \in F(A)$. Следовательно, $F(UAU^*) \subseteq F(A)$. Подобным проведенному рассуждением доказываем обратное включение, так что $F(A) = F(UAU^*)$.

Для нормальных матриц (§ 2.10) характеризовать геометрию поля значений теперь очень легко. Следует вспомнить сначала определение выпуклой оболочки некоторого множества точек на комплексной плоскости (частный случай упр. 12 § 6.2).

Теорема 6.6.2. *Поле значений нормальной матрицы совпадает с выпуклой оболочкой ее собственных значений.*

Доказательство. Если A — нормальная матрица с собственными значениями $\mu_1, \ldots, \mu_n$, то по теореме 2.10.2 существует унитарная матрица U, для которой $A = UDU^*$ и $D = \operatorname{diag}\{\mu_1, \ldots, \mu_n\}$. По предшествующей теореме имеем $F(A) = F(D)$. Таким образом, $z \in F(A)$ в том и только в том случае, когда существует $\boldsymbol{x}$ с $h(\boldsymbol{x}) = 1$ и

$$z = \boldsymbol{x}^* D \boldsymbol{x} = \sum_{j=1}^{n} |x_j|^2 \mu_j.$$

Так как $h(\boldsymbol{x}) = \sum_{j=1}^{n} |x_j|^2 = 1$, из упр. 12 § 6.2 следует, что z принадлежит выпуклой оболочке собственных значений, так что теорема доказана.

Заметим, что если в общем случае Z_n — множество собственных значений матрицы $A \in \mathscr{C}_{n \times n}$, то, так как каждое собственное значение принадлежит $F(A)$ и $F(A)$ выпукло, из определения выпуклой оболочки получаем, что $F(A) \subseteq H(Z_n)$.

Теорема 6.6.3. *Поле значений матрицы A является отрезком вещественной прямой тогда и только тогда, когда A эрмитова.*

Этот результат следует сравнить с теоремой 3.2.1 и упр. 3 из § 3.2.

Доказательство. Если A эрмитова, то A нормальна и, согласно теореме 6.6.2, $F(A)$ — выпуклая оболочка собственных значений матрицы A. Но собственные значения A все вещественны, так что их выпуклая оболочка, очевидно, совпадает с отрезком вещественной прямой.

Наоборот, если $A \in \mathscr{C}_{n \times n}$, то мы полагаем

$$B = \frac{1}{2}(A + A^*), \quad C = -\frac{1}{2} i (A - A^*).$$

Как нетрудно видеть, $B = B^*$, $C = C^*$ и $A = B + iC$. Тогда $\boldsymbol{x}^* A \boldsymbol{x} = \boldsymbol{x}^* B \boldsymbol{x} + i \boldsymbol{x}^* C \boldsymbol{x}$ и $\boldsymbol{x}^* B \boldsymbol{x}$, $\boldsymbol{x}^* C \boldsymbol{x}$ вещественны. Если $F(A)$ состоит лишь из вещественных чисел, то $\boldsymbol{x}^* C \boldsymbol{x} = 0$ для каждого $\boldsymbol{x} \in \mathscr{C}_n$, и, следовательно, $C = 0$ (упр. 9 из § 2.9). Но $C = 0$ означает, что $A = A^*$, что и требовалось доказать.

Глава 7

ТЕОРИЯ ВОЗМУЩЕНИЙ И ОЦЕНКИ ДЛЯ СОБСТВЕННЫХ ЗНАЧЕНИЙ

7.1. Возмущения в решении линейных уравнений

Бо́льшая часть этой главы посвящена исследованию поведения матричных собственных значений при возмущениях элементов матрицы. Прежде чем приступить к нему, будет поучительно (и, в частности, будет иметь важное значение для аналитика-числовика) установить воздействие возмущения A и $\boldsymbol{b}$ на решение $\boldsymbol{x}$ уравнения $A\boldsymbol{x} = \boldsymbol{b}$. Мы предположим, что $A \in \mathscr{C}_{n \times n}$ и $\det A \neq 0$, так что (теорема 1.16.5) $\boldsymbol{x}$ будет *единственным* решением уравнения $A\boldsymbol{x} = \boldsymbol{b}$.

В общем случае некоторая задача считается устойчивой, если «малые» причины (возмущения в A или $\boldsymbol{b}$) приводят к «малым» воздействиям (возмущениям в $\boldsymbol{x}$). Степень малости участвующих величин обычно измеряется в зависимости от их значений в невозмущенном состоянии. Таким образом, если $\boldsymbol{x} + \boldsymbol{y}$ — решение задачи после возмущения, то мы измеряем величину возмущения в векторе-решении отношением $h(\boldsymbol{y})/h(\boldsymbol{x})$ для некоторой векторной нормы h. Мы будем называть это отношение *относительным возмущением* $\boldsymbol{x}$.

Прежде чем исследовать задачу с общими коэффициентами матрицы A, очень полезно рассмотреть задачу-прототип вариации в коэффициентах уравнения $\boldsymbol{x} = \boldsymbol{b}$. Таким образом, мы пришли к необходимости исследовать решения уравнения $(I + M)\xi = \boldsymbol{b}$, где M «мало» по сравнению с I. Отклонение ξ от $\boldsymbol{x}$ будет, очевидно, зависеть от вида матрицы $(I + M)^{-1}$, если это обращение существует. Имеем:

Теорема 7.1.1. *Если $\|\cdot\|$ обозначает любую матричную норму, для которой $\|I\| = 1$, и если $\nu = \|M\| < 1$, то $(I + M)^{-1}$ существует:*

$$(I + M)^{-1} = I - M + M^2 - \ldots$$

и

$$\|(I + M)^{-1}\| \leqslant \frac{1}{1-\nu}.$$

Доказательство. Теорема 6.1.3 означает, что для любого собственного значения μ матрицы M имеем $|\mu| \leqslant \nu < 1$. Более того, функция f, определенная как $f(\lambda) = (1+\lambda)^{-1}$, имеет разложение в степенной ряд в точке $\lambda = 0$ с радиусом сходимости, равным 1. Следовательно (теорема 5.6.3), $f(M) = (I+M)^{-1}$ существует и

$$(I+M)^{-1} = I - M + M^2 - \ldots$$

Если положить $S_p = I - M + M^2 - \ldots \pm M^{p-1}$ (p-я частичная сумма) и заметить, что $\|M^r\| \leqslant \|M\|^r$, то получим

$$\|S_p\| \leqslant \|I\| + \|M\| + \ldots + \|M^{p-1}\| \leqslant$$
$$\leqslant 1 + \nu + \ldots + \nu^{p-1} = \frac{1-\nu^p}{1-\nu} \leqslant \frac{1}{1-\nu}.$$

Так как эта граница не зависит от p, то

$$\|S_\infty\| = \|(I+M)^{-1}\| \leqslant \frac{1}{1-\nu}. \quad \blacktriangleleft$$

Возвратимся к нашей первоначальной задаче. Имеем

$$A\boldsymbol{x} = \boldsymbol{b}$$

с $\det A \neq 0$. Предположим, что $\boldsymbol{b}$ переходит в $\boldsymbol{b}+\boldsymbol{k}$, A переходит в $A+F$ и $\boldsymbol{x}+\boldsymbol{y}$ удовлетворяет уравнению

$$(A+F)(\boldsymbol{x}+\boldsymbol{y}) = \boldsymbol{b}+\boldsymbol{k}.$$

Мы вычислим верхнюю оценку для относительного возмущения $h(\boldsymbol{y})/h(\boldsymbol{x})$.

Вычитая первое уравнение из второго, получаем

$$(A+F)\boldsymbol{y} + F\boldsymbol{x} = \boldsymbol{k}$$

или

$$A\left(I + A^{-1}F\right)\boldsymbol{y} = \boldsymbol{k} - F\boldsymbol{x}.$$

Мы предположим, что $\delta = \|A^{-1}\|\|F\| < 1$ и что $\|I\| = 1$. Так как тогда $\|A^{-1}F\| \leqslant \|A^{-1}\|\|F\| < 1$, теорема означает, что $(I + A^{-1}F)^{-1}$ существует и что $\|(I+A^{-1}F)^{-1}\| < (1-\delta)^{-1}$. Таким образом,

$$\boldsymbol{y} = \left(I + A^{-1}F\right)^{-1} A^{-1}\boldsymbol{k} - \left(I + A^{-1}F\right)^{-1} A^{-1}F\boldsymbol{x},$$

и если h — любая векторная норма, согласованная с $\|\cdot\|$, то

$$h(\boldsymbol{y}) \leqslant \frac{\|A^{-1}\|}{1-\delta} h(\boldsymbol{k}) + \frac{\delta}{1-\delta} h(\boldsymbol{x}).$$

Чтобы получить удобное выражение для искомой меры $h(\boldsymbol{y})/h(\boldsymbol{x})$ — величины возмущения, мы используем то, что из

$Ax = b$ следует $h(b) \leqslant \|A\| h(x)$, откуда $1/h(x) \leqslant \|A\|/h(b)$. Таким образом,

$$\frac{h(y)}{h(x)} \leqslant \frac{\|A\| \|A^{-1}\|}{1-\delta} \frac{h(k)}{h(b)} + \frac{\delta}{1-\delta}.$$

Первый член справа содержит относительное возмущение b и последний член не зависит от правой части первоначального уравнения. Определим теперь $k(A) = \|A\| \|A^{-1}\|$ как *условное число* для A (относительно нормы $\|\cdot\|$) и заметим, что

$$\delta = \|A^{-1}\| \|F\| = \frac{k(A) \|F\|}{\|A\|}.$$

Таким образом, окончательно получаем

$$\frac{h(y)}{h(x)} \leqslant \frac{k(A)}{1 - k(A)\|F\|/\|A\|} \left(\frac{h(k)}{h(b)} + \frac{\|F\|}{\|A\|} \right).$$

Это дает нам верхнюю оценку для относительного возмущения x через относительные возмущения b и A и условное число $k(A)$. В частности, следует отметить решающую роль при этом условного числа. Оно участвует в оценке во всех случаях, будут ли возмущения происходить только в b, только в A и в b и A одновременно. Условное число будет иметь большое значение также в исследовании возмущений собственных значений (§ 7.4).

Упр. 1. Доказать, что если $\|A^{-1}\| \|F\| < 1$, то $\|F\| < \|A\|$.

Упр. 2. Если $A = \operatorname{diag}\{1, 2, 3, \ldots, 10\}$, $\rho(b) = 1$ (норма из упр. 2 (а) § 6.2) и известно лишь, что элементы F и k ограничены по абсолютной величине ε с $\varepsilon < 0,1$, то

$$\frac{\rho(y)}{\rho(x)} \leqslant \frac{20\varepsilon}{1 - 10\varepsilon}.$$

Если дополнительно

$$b = \begin{Vmatrix} 0,1 \\ 0,2 \\ \cdot \\ \vdots \\ \cdot \\ 1,0 \end{Vmatrix}, \quad k = \begin{Vmatrix} -\varepsilon \\ 0 \\ \cdot \\ \vdots \\ \cdot \\ 0 \end{Vmatrix}, \quad F = \begin{Vmatrix} \varepsilon & \varepsilon & \ldots & \varepsilon \\ 0 & 0 & \ldots & 0 \\ \cdot & & & \cdot \\ \vdots & & & \vdots \\ \cdot & & & \cdot \\ 0 & 0 & \ldots & 0 \end{Vmatrix},$$

то показать, что

$$\frac{\rho(y)}{\rho(x)} = \frac{20\varepsilon}{1 + \varepsilon}.$$

Упр. 3. Если $k(A)$ — *спектральное условное число* для A, т. е. условное число, полученное при использовании спектральной нормы, и λ_1, λ_n — наименьшее и наибольшее собственные значения матрицы A^*A, то доказать, что $k(A) = (\lambda_n/\lambda_1)^{1/2}$.

Упр. 4. Если линейное уравнение $Ax = b$ с $\det A \neq 0$ заменяется на эквивалентное уравнение $Bx = A^*b$, где $B = A^*A$, то доказать, что это может быть сделано за счет спектрального условного числа, т. е. что $k(B) \geqslant k(A)$.

Упр. 5. Доказать, что при предположениях теоремы 7.1.1 $M^p \to 0$ при $p \to \infty$. Доказать в общем случае, что $M^p \to 0$ при $p \to \infty \Longleftrightarrow$ спектральный радиус матрицы M меньше 1.

7.2. Теорема Гершгорина

Мы приступаем теперь к рассмотрению вопроса о зависимости собственных значений и собственных векторов матрицы A от элементов A. Для $A \in \mathscr{C}_{n \times n}$ и ее собственных значений $\mu_1, \ldots, \mu_n$ наши первоначальные результаты будут описывать области комплексной плоскости, в которых должны лежать собственные значения. Прежде чем переходить к деталям, напомним, что мы уже имеем один такой результат. Он упомянут в замечании к упр. 1 из § 6.5.

Первым существенным результатом является то, что собственные значения непрерывно зависят от элементов матрицы A. Так как собственные значения — это в точности корни характеристического многочлена, то последний результат получаем сразу же, если известно, что корни многочлена непрерывно зависят от его коэффициентов. Это результат элементарной теории алгебраических функций и не будет здесь излагаться. В § 7.5 будут приведены еще некоторые подробности, когда мы будем рассматривать аналитические возмущения; при этом будут доказаны более сильные результаты, описывающие поведение собственных значений в некоторых специальных случаях.

Следует заметить, что некоторая информация утрачивается, когда мы имеем дело лишь с характеристическим многочленом, ибо существует много различных матриц с заданным характеристическим многочленом. Поэтому не удивительно, что более сильные результаты учитывают строение матриц.

Следующее упражнение иллюстрирует то, что, хотя собственные значения непрерывно зависят от матричных элементов, они могут тем не менее изменяться очень быстро при патологических обстоятельствах.

Упр. 1. Рассмотрим 10×10-матрицу $A(\varepsilon)$, определяемую как функция вещественной переменной ε посредством

$$A(\varepsilon) = \begin{Vmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & & \cdot & \\ \cdot & \cdot & & & \cdot \\ 0 & 0 & \ldots & 0 & 1 \\ \varepsilon & 0 & \ldots & 0 & 0 \end{Vmatrix}.$$

Доказать, что собственные значения при $\varepsilon = 0$ и при $\varepsilon = 10^{-10}$ отличаются по абсолютной величине на 0,1. (Заметим, что разница между $A(0)$ и $A(10^{-10})$ настолько мала, что она не может быть различима для вычислительной машины со слишком короткой длиной слов.)

Мы можем доказать теперь одну из наиболее полезных и легко применяемых теорем, дающую оценки для собственных значений. Она известна как теорема Гершгорина и была впервые опубликована всего лишь в 1937 г.

Теорема 7.2.1. *Если* $A \in \mathscr{C}_{n\times n}$ *и*

$$\rho_j = \sum_k{}' |a_{jk}|,$$

где $\sum\limits_k{}'$ *обозначает суммирование от* $k=1$ *до* n, $k \neq j$, *то каждое из собственных значений матрицы* A *лежит по крайней мере в одном из кругов*

$$|z - a_{jj}| \leqslant \rho_j, \qquad j = 1, 2, \ldots, n,$$

комплексной z*-плоскости. Кроме того, множество из* m *кругов, не пересекающихся с остальными* $n-m$ *кругами, содержит* m *и только* m *собственных значений матрицы* A.

Доказательство. 1. Пусть μ — некоторое собственное значение матрицы A с правым собственным вектором $\boldsymbol{x}$. Тогда $A\boldsymbol{x} = \mu\boldsymbol{x}$, или, записывая это соотношение в виде n скалярных равенств,

$$\sum_{k=1}^{n} a_{jk}x_k = \mu x_j, \qquad j = 1, 2, \ldots, n.$$

Пусть $|x_p| = \max\limits_j |x_j|$; тогда p-е равенство дает

$$|\mu - a_{pp}|\,|x_p| = \Big|\sum_k{}' a_{pk}x_k\Big| \leqslant \sum_k{}' |a_{pk}|\,|x_k| \leqslant |x_p| \sum_k{}' |a_{pk}|.$$

Но так как $\boldsymbol{x} \neq \boldsymbol{0}$, то $|x_p| \neq 0$, и мы имеем $|\mu - a_{pp}| \leqslant \rho_p$. Первое утверждение доказано.

2. Предположим теперь, что $A = D + C$, где $D = \mathrm{diag}\{a_{11}, \ldots, a_{nn}\}$, и пусть $B(t) = D + tC$. Тогда $B(0) = D$ и $B(1) = A$. Мы рассмотрим поведение собственных значений матрицы $B(t)$ при t, меняющемся в отрезке $[0, 1]$, и воспользуемся их непрерывностью как функций от t. Итак, для любого t в этом отрезке собственные значения матрицы $B(t)$ лежат в кругах с центрами a_{jj} и радиусами $t\rho_j$, $j = 1, 2, \ldots, n$.

Предположим, что j-й круг для $A = B(1)$ не пересекается с остальными $n-1$ кругами. Очевидно тогда, что j-й круг для $B(t)$ не пересекается со всеми остальными кругами при всех t из $[0, 1]$. Но при $t = 0$ собственными значениями матрицы $B(0)$ будут $a_{11}, \ldots, a_{nn}$ и из них a_{jj} — единственное собственное значение в j-м (вырожденном) круге. Так как собственные значения матрицы $B(t)$ являются непрерывными функциями от t и j-й круг всегда не пересекается со всеми остальными кругами, то для всех $t \in [0, 1]$ существует одно и только одно собственное значение матрицы $B(t)$ в j-м круге. В частности, это будет так в случае $t = 1$ и $B(1) = A$.

Это доказывает вторую часть теоремы для $m = 1$. Читателю предоставляется закончить доказательство для любого $m \leqslant n$.

Заметим, что, применяя теорему к A', получаем подобный результат, но с использованием сумм по столбцам матрицы A, а не сумм по строкам, для определения радиусов новых кругов.

Отметим также, что если ρ — векторная норма $\rho(\boldsymbol{x}) = \max\limits_j |x_j|$, то индуцированная матричная норма $\|\cdot\|_\rho$ задается в виде

$$\|A\|_\rho = \max_j \sum_k |a_{jk}|$$

(упр. 9 из § 6.3). Но, очевидно, никакая точка из круга Гершгорина не может быть удаленной от начала координат более чем на расстояние $\|A\|_\rho$, а так как все собственные значения лежат в этих кругах, то $\|A\|_\rho \geqslant \mu_A$ — спектрального радиуса матрицы A. Это — частный случай теоремы 6.1.3. Подобно этому, применяя теорему Гершгорина к A', находим, что $\|A\|_\gamma \geqslant \mu_A$ (упр. 10 из § 6.3).

На практике значение теоремы Гершгорина часто может быть увеличено тем, что собственные значения матрицы инвариантны при подобных преобразованиях. Таким образом, если S — любая неособая матрица, то применение теоремы к SAS^{-1} может дать дальнейшую информацию о локализации собственных значений матрицы A. Чтобы избежать очень длинных вычислений матрицы S^{-1} (которые могут быть лучше применены к самой A), обычно ограничиваются рассмотрением диагональных преобразующих матриц S.

***Упр. 2.** Найти круги Гершгорина для следующей матрицы и для ее транспонированной:

$$A = \begin{Vmatrix} 0 & 1 & 0 & i \\ 1 & 6 & 1 & 1 \\ i/2 & i & 5 & 0 \\ 0 & 1/2 & 1/2 & -2 \end{Vmatrix}.$$

Если $S = \operatorname{diag}\{1, 1, 1, 4\}$, то вычислить SAS^{-1} и получить, что A имеет собственное значение в круге $|z + 2| \leqslant \frac{1}{2}$.

Упр. 3. Показать, что матрица A будет неособой, если

$$|a_{jj}| > \sum_k{}' |a_{jk}| = \rho_j, \qquad j = 1, 2, \ldots, n.$$

(Такая матрица называется матрицей с *доминирующей диагональю.*)

Упр. 4. Пусть матрица A имеет доминирующую диагональ и $d_j = |a_{jj}| - \rho_j$, $j = 1, 2, \ldots, n$. Доказать, что $|\det A| \geqslant d_1 d_2 \ldots \ldots d_n$. (Верхняя оценка для $|\det A|$ будет получена в упр. 1 из § 7.3.)

Решение. Определим матрицу B элементами $b_{jk} = a_{jk}/d_j$. Тогда

$$\det B = \frac{1}{d_1 d_2 \ldots d_n} \det A,$$

или

$$\det A = d_1 d_2 \ldots d_n \det B.$$

Поставленная задача будет решена, если мы сможем показать, что $|\det B| \geqslant 1$. Пусть μ — некоторое собственное значение матрицы B. Тогда по теореме 7.2.1 имеем

$$|\mu - b_{jj}| \leqslant \sum_k{}' |b_{jk}|$$

для некоторого j. Но $|\mu - b_{jj}| \geqslant ||\mu| - |b_{jj}||$ и, используя определения B и d_j, получаем, что

$$|\mu| \geqslant |b_{jj}| - \sum_k{}' |b_{jk}| = \frac{|a_{jj}| - \rho_j}{d_j} = 1.$$

Таким образом, собственные значения матрицы B все имеют абсолютные значения не меньше 1. Но $\det B$ равен произведению собственных значений для B (§ 2.1), так что $|\det B| \geqslant 1$.

Упр. 5. Доказать, что эрмитова матрица с доминирующей диагональю, у которой все диагональные элементы положительны, будет положительно определенной.

Упр. 6. Если $A = B + C$, где $B = \operatorname{diag}\{1, 2, 2\}$ и $|c_{jk}| \leqslant \varepsilon < \frac{1}{6}$ для $j, k = 1, 2, 3$, то доказать, что существует собственное значение матрицы A в круге $|z - 1 - c_{11}| \leqslant 12\varepsilon^2$.

Упр. 7 (обобщение упр. 6). Если $A = B + C$, где $B = \operatorname{diag}\{b_1, \ldots, b_n\}$, $\beta_j = \min\limits_{k \neq j} |b_j - b_k| > 0$ и $|c_{kl}| \leqslant \varepsilon < \beta_j/2n$ для $k, l = 1, 2, \ldots, n$, то доказать, что существует собственное значение матрицы A в круге

$$|z - b_j - c_{jj}| \leqslant 2n(n-1)\varepsilon^2/\beta_j.$$

(Заметим, что если записать $A = B + \zeta C$ и снять ограничение на C, то будет существовать собственное значение $\mu(\zeta)$ матрицы A, для которого при $\zeta \to 0$

$$\mu(\zeta) = b_j + \zeta c_{jj} + O(|\zeta|^2).$$

Этот результат будет улучшен в § 7.7.)

Упр. 8. Если $f(z) = a_0 + a_1 z + \ldots + a_n z^n$, $a_n \neq 0$, и z_0 — корень f, то доказать, что

$$|z_0| \leqslant 1 + \max_k |a_k/a_n|.$$

7.3. Теорема Шура

Мы приведем теперь результат в несколько другом направлении, принадлежащий Шуру и датируемый 1909 г. Напомним (упр. 8 из § 2.9) о разложении некоторой матрицы $A \in \mathscr{C}_{n\times n}$ в сумму эрмитовой матрицы B и косоэрмитовой матрицы C. Итак, мы полагаем $B = \frac{1}{2}(A + A^*)$, $C = \frac{1}{2}(A - A^*)$ и отмечаем, что $A = B + C$, $B^* = B$ и $C^* = -C$. Если $A^* = A$, то $C = 0$ и все собственные значения матрицы A вещественны. Это наводит на мысль, что норма матрицы C может задавать верхнюю оценку для величин мнимых частей собственных значений матрицы A. Подобно этому, если $A^* = -A$, то $B = 0$ и все собственные значения матрицы A имеют нулевые вещественные части, так что можно ожидать, что норма матрицы B задает верхнюю оценку для величин вещественных частей собственных значений матрицы A. Теорема Шура содержит результаты этого рода. В следующем утверждении $\operatorname{Re}(\mu)$, $\operatorname{Im}(\mu)$ будут обозначать вещественную и мнимую части комплексного числа μ.

Теорема 7.3.1. *Если $A \in \mathscr{C}_{n\times n}$, $\|\cdot\|$ означает евклидову матричную норму и $\mu_1, \ldots, \mu_n$ — собственные значения матрицы A, то*

$$\sum_{r=1}^{n} |\mu_r|^2 \leqslant \|A\|^2, \quad \sum_{r=1}^{n} |\operatorname{Re}(\mu_r)|^2 \leqslant \|B\|^2, \quad \sum_{r=1}^{n} |\operatorname{Im}(\mu_r)|^2 \leqslant \|C\|^2,$$

где $B = \frac{1}{2}(A + A^*)$ и $C = \frac{1}{2}(A - A^*)$. Равенство в любом одном из этих неравенств влечет равенство во всех трех и реализуется в том и только в том случае, когда A нормальна.

Доказательство. Теорема 2.10.1 информирует нас о существовании таких унитарной матрицы U и верхней треугольной матрицы Δ, что $A = U^*\Delta U$, и собственные значения матрицы A будут элементами на главной диагонали в Δ. Но евклидова норма инвариантна при унитарных подобных преобразованиях (упр. 7 из § 6.3), так что $\|A\|^2 = \|\Delta\|^2$. Если элементы в Δ обозначить d_{rs} ($r, s = 1, 2, \ldots, n$), то

$$\|A\|^2 = \sum_{r=1}^{n} |\mu_r|^2 + \sum_{r<s} |d_{rs}|^2 \geqslant \sum_{r=1}^{n} |\mu_r|^2$$

и первое неравенство доказано.

Имеем также $B = \frac{1}{2} U^*(\Delta + \Delta^*) U$, так что, как и прежде,

$$\|B\|^2 = \sum_{r=1}^{n} \left| \frac{\mu_r + \bar{\mu}_r}{2} \right|^2 + \sum_{r \neq s} \left| \frac{d_{rs} + \bar{d}_{rs}}{2} \right|^2 =$$

$$= \sum_{r=1}^{n} |\operatorname{Re}(\mu_r)|^2 + 2 \sum_{r<s} \left| \frac{d_{rs}}{2} \right|^2 \geqslant \sum_{r=1}^{n} |\operatorname{Re}(\mu_r)|^2.$$

Подобное доказательство дает также и третье неравенство.

Ясно, что равенство реализуется в каждом из трех неравенств $\Leftrightarrow$ $d_{rs} = 0$ для всех $r \neq s$ и $r, s = 1, 2, \ldots, n$, т. е. в том и только в том случае, когда A унитарно подобна диагональной матрице из ее собственных значений. Утверждение следует теперь из теоремы 2.10.2.

Следствие 1 (Хирш, 1905 г.). *Если положить*

$$\rho = \max_{r,s} |a_{rs}|, \quad \sigma = \max_{r,s} |b_{rs}|, \quad \tau = \max_{r,s} |c_{rs}|$$

и если μ — *любое собственное значение матрицы* A, *то*

$$|\mu| \leqslant n\rho, \quad |\operatorname{Re}(\mu)| \leqslant n\sigma, \quad |\operatorname{Im}(\mu)| \leqslant n\tau.$$

Доказательство. Очевидно, $\|A\|^2 \leqslant \sum_{r,s} \rho^2 = n^2\rho^2$. Первое неравенство теоремы тогда дает $|\mu| \leqslant \left(\sum_r |\mu_r|^2\right)^{1/2} \leqslant \|A\| \leqslant n\rho$, что является первым результатом следствия. Другие два результата получаются подобно первому.

Отметим, что частный случай $\tau = 0$ приводит к другому (с превышением по мощности) доказательству того, что собственные значения эрмитовой матрицы вещественны. Подобно этому случай $\sigma = 0$ приводит к результату, что собственные значения косоэрмитовой матрицы имеют нулевые вещественные части.

Наконец, следует заметить, что $n\rho$ совпадает с нормой $M(A)$, определенной в упр. 9 из § 6.1. Результат $|\mu| \leqslant n\rho$ является поэтому частным случаем теоремы 6.1.3.

Следствие 2 (Бендиксон, 1902 г.). *Если* $A \in \mathscr{R}_{n \times n}$ *и* $\tau = \frac{1}{2} \max_{r,s} |a_{rs} - a_{sr}|$, *то для любого собственного значения* μ *матрицы* A *имеем*

$$|\operatorname{Im}(\mu)| \leqslant \tau \sqrt{n(n-1)/2}.$$

Доказательство. Для вещественной матрицы A третье неравенство теоремы означает, что

$$\sum_{r=1}^{n} |\operatorname{Im}(\mu_r)|^2 \leqslant \sum_{\substack{r,\, s=1 \\ r \neq s}} \left| \frac{a_{rs} - a_{sr}}{2} \right|^2 \leqslant \tau^2 n(n-1).$$

Так как A вещественна, все ее комплексные собственные значения распадаются на комплексно сопряженные пары, так что каждый ненулевой член в сумме слева входит в нее по крайней мере дважды. Следовательно, $2|\operatorname{Im}(\mu_r)|^2 \leqslant \tau^2 n(n-1)$ и следствие доказано.

В заключение следует отметить, что оценки, получаемые в следствиях, вообще говоря, менее строгие, чем соответствующие оценки, получаемые из теоремы Гершгорина. Это хорошо иллюстрируется применением следствия 1 к упр. 2 из § 7.2.

Упр. 1. Доказать, что $|\det A| \leqslant n^{n/2}\rho^n$, где $\rho = \max\limits_{r,s} |a_{rs}|$.

Решение. Прежде всего напомним неравенство между арифметическим и геометрическим средними: для любых неотрицательных чисел $a_1, a_2, \ldots, a_n$ имеем

$$(a_1 a_2 \ldots a_n)^{1/n} \leqslant \frac{a_1 + a_2 + \ldots + a_n}{n}.$$

Если $\mu_1, \mu_2, \ldots, \mu_n$ — собственные значения матрицы A, то

$$|\det A|^2 = |\mu_1|^2 |\mu_2|^2 \ldots |\mu_n|^2 \leqslant \left(\frac{|\mu_1|^2 + |\mu_2|^2 + \ldots + |\mu_n|^2}{n} \right) \leqslant$$
$$\leqslant \left(\frac{1}{n} \|A\|^2 \right)^n \leqslant \left(\frac{1}{n} n^2 \rho^2 \right)^n$$

и извлечение квадратного корня приводит к результату.

Упр. 2. Если $\mu_1, \mu_2, \ldots, \mu_n$ — собственные значения матрицы A и $\lambda_1, \lambda_2, \ldots, \lambda_n$ — собственные значения для A^*A, то показать, что

$$\sum_{r=1}^{n} |\mu_r|^2 \leqslant \sum_{r=1}^{n} \lambda_r,$$

причем равенство имеет место в том и только в том случае, когда A нормальна.

Упр. 3. В обозначениях упр. 2 показать, что A будет нормальной в том и только в том случае, когда при некоторой нумерации собственных значений $\lambda_r = |\mu_r|^2$, $r = 1, 2, \ldots, n$.

7.4. Возмущение собственных значений простой матрицы

Пусть A — простая матрица с собственными значениями $\mu_1, \mu_2, \ldots, \mu_n$. Если тогда $U = \operatorname{diag}\{\mu_1, \ldots, \mu_n\}$, то существует такая неособая матрица P, что $A = PUP^{-1}$. В теореме 6.4.1 мы имели возможность характеризовать некоторый класс матричных норм с удобными свойствами, если их применять к диагональным матрицам, так что мы приходим к мысли оценить $\|A\|$ для такой нормы через $\|U\|$. Таким образом, если $\|\cdot\|$ — матричная норма, индуцированная абсолютной векторной нормой, то равенство $A = PUP^{-1}$ и теорема 6.4.1 означают, что

$$\| A \| \leqslant k(P) \max_j |\mu_j|,$$

где $k(P)$ — условное число для P относительно $\|\cdot\|$, как определено в § 7.1. Однако матрица P определена равенством $A = PUP^{-1}$ не однозначно. В частности, $A = QUQ^{-1}$ для любой матрицы $Q = PD$, где D — некоторая неособая диагональная матрица. Мы получим поэтому лучшую возможную оценку

$$\| A \| \leqslant \nu(A) \max_j |\mu_j|,$$

если положить $\nu(A) = \inf\limits_P k(P)$, где инфимум берется относительно всех матриц P, для которых $P^{-1}AP$ диагональна. Этот результат может рассматриваться как нижняя оценка для спектрального радиуса матрицы A и вместе с теоремой 6.1.3 дает первую часть следующей теоремы.

Теорема 7.4.1. *Если A — простая матрица, $\|\cdot\|$ — матричная норма, индуцированная некоторой абсолютной векторной нормой h, и $\nu(A) = \inf\limits_P k(P)$, где $P^{-1}AP$ диагональна, то*

$$\frac{\| A \|}{\nu(A)} \leqslant \max_j |\mu_j| \leqslant \| A \|,$$

и

$$\operatorname{glb}_h(A) \leqslant \min_j |\mu_j| \leqslant \nu(A) \operatorname{glb}_h(A).$$

Доказательство. Первая часть последнего результата в точности совпадает с упр. 1 из § 6.5. Что касается второй части, то равенство $A = PUP^{-1}$ и следствия 1 и 2 теоремы 6.5.1 означают, что

$$\operatorname{glb}_h(A) \geqslant \operatorname{glb}_h(P) \operatorname{glb}_h(P^{-1}) \min_j |\mu_j|.$$

Теорема 6.5.1 тогда дает

$$\operatorname{glb}_h(A) \geqslant (k(P))^{-1} \min_j |\mu_j|.$$

Отсюда следует, что $\min\limits_j |\mu_j| \leqslant \nu(A) \operatorname{glb}_h(A)$.

Для матрицы A общего вида доказанная теорема имеет сомнительное вычислительное значение, так как величину $\nu(A)$ оценить трудно. Однако следует заметить, что $PP^{-1} = I$ означает, что $k(P) = \|P\| \|P^{-1}\| \geqslant 1$, а следовательно, что $\nu(A) \geqslant 1$.

Если A — нормальная матрица, то мы знаем (теорема 2.10.2), что P может быть выбрана унитарной матрицей. Если, кроме того, мы используем евклидову векторную норму h, а следовательно, спектральную матричную норму $\|\cdot\|$, то $\|P\| = \|P^{-1}\| = 1$. Таким образом, в этом частном случае имеем $\nu(A) = 1$, и теорема превращается в известные случаи:

$$\|A\| = \max_j |\mu_j|, \quad \mathrm{glb}_h(A) = \min_j |\mu_j|.$$

Следующая наша теорема наглядно показывает значение условного числа $\nu(A)$ для A как отражающего устойчивость собственных значений при возмущении матричных элементов. Нам понадобится следующий простой результат.

Лемма. *Если $R, S \in \mathscr{C}_{n\times n}$ и $Rx = Sx$ для некоторого ненулевого $x \in \mathscr{C}_n$, то для любой векторной нормы h и индуцированной ею матричной нормы $\|\cdot\|$ имеем*

$$\mathrm{glb}_h(R) \leqslant \|S\|, \quad \|R\| \geqslant \mathrm{glb}_h(S).$$

Доказательство. Используя определения матричной нормы и нижней грани, получаем, что

$$\mathrm{glb}_h(R) \leqslant \frac{h(Rx)}{h(x)} = \frac{h(Sx)}{h(x)} \leqslant \|S\|,$$

т. е. получено первое неравенство. Воспользовавшись симметрией предположения в лемме, можем поменять местами R и S в первом неравенстве для получения второго. ◀

Теорема 7.4.2. *Пусть $A, B \in \mathscr{C}_{n\times n}$ и A простая. Если A имеет собственные значения $\mu_1, \ldots, \mu_n$, μ — собственное значение матрицы $A + B$ и для матричной нормы, индуцированной некоторой абсолютной векторной нормой,*

$$r = \|B\| \nu(A),$$

то μ лежит по крайней мере в одном из кругов

$$|z - \mu_j| \leqslant r, \quad j = 1, 2, \ldots, n,$$

комплексной z-плоскости.

Доказательство. Так как μ — собственное значение для $A + B$, то существует $y \neq 0$, для которого

$$(A + B) y = \mu y.$$

Если $A = PUP^{-1}$, где $U = \operatorname{diag}\{\mu_1, \ldots, \mu_n\}$, то

$$\left(U + P^{-1}BP\right)x = \mu x,$$

где $x = P^{-1}y \neq 0$. Таким образом,

$$(\mu I - U)\,x = P^{-1}BPx. \tag{7.4.1}$$

Используя лемму, получаем $\operatorname{glb}_h(\mu I - U) \leqslant \|P^{-1}BP\| \leqslant k(P)\,\|B\|$, откуда

$$\min_j |\mu - \mu_j| \leqslant k(P)\,\|B\|.$$

Так как это верно для каждой P, приводящей A к диагональному виду, то

$$\min_j |\mu - \mu_j| \leqslant \nu(A)\,\|B\| = r.$$

Следовательно, μ лежит по крайней мере в одном из кругов $|z - \mu_j| \leqslant r$. ◄

Упр. 1. Показать, что в упр. 2 из § 7.2 собственные значения лежат в кругах с радиусом 2 и центрами в $3 \pm \sqrt{10}$, $5i$ и -2.

Упр. 2. Доказать и добавить к теореме 7.4.2 свойство отделимости для собственных значений матрицы $A + B$, аналогичное свойству отделимости в теореме 7.2.1.

Если про матрицу возмущения B из теоремы 7.4.2 известно также, что она простая, то радиус получаемых в этой теореме кругов может быть выбран в другом виде, который указан в следующей теореме.

Теорема 7.4.3. *Если, в дополнение к предположениям теоремы 7.4.2, B — простая матрица с собственными значениями $\mu_1', \ldots, \mu_n'$, то μ лежит по крайней мере в одном из кругов*

$$|z - \mu_j| \leqslant \nu(A, B)\max_j |\mu_j'|, \qquad j = 1, 2, \ldots, n,$$

где

$$\nu(A, B) = \inf_{P,\,Q} k\left(P^{-1}Q\right)$$

и P, Q приводят к диагональному виду A и B соответственно.

Доказательство. Мы можем переписать теперь равенство (7.4.1) в виде

$$(\mu I - U)\,x = \left(P^{-1}Q\right)U_1\left(P^{-1}Q\right)^{-1}x,$$

где $U_1 = \operatorname{diag}\{\mu_1', \ldots, \mu_n'\}$. Следовательно,

$$\operatorname{glb}_h(\mu I - U) \leqslant k\left(P^{-1}Q\right)\|U_1\|,$$

откуда и следует результат.

Очень легко можно получить более сильную форму этой теоремы. Заметим, что P будет приводить к диагональному виду A в том и только в том случае, когда P будет приводить к диагональному виду $\alpha I + A$ для любого комплексного числа α, и подобно этому для Q, B и $-\alpha I + B$. Так как

$$(\alpha I + A) + (-\alpha I + B) = A + B,$$

то, применяя теорему к $\alpha I + A$ и $-\alpha I + B$, получаем

Следствие 1. *При предположениях теоремы* 7.4.3 μ *лежит по крайней мере в одном из кругов*

$$|z - (\alpha + \mu_j)| \leqslant \nu(A, B) \max_i |\mu'_i - \alpha|$$

для любого фиксированного комплексного числа α.

Преимущество этого результата в том, что α может быть выбрано теперь минимизирующим величину $\max\limits_i |\mu'_i - \alpha|$ и, следовательно, минимизирующим площадь кругов, содержащих собственные значения.

Случай нормальных A и B достаточно важен, чтобы оправдать отдельное утверждение.

Следствие 2. *Если при предположениях теоремы* 7.4.3 *матрицы A и B нормальны, то собственные значения для $A + B$ содержатся в кругах*

$$|z - (\alpha + \mu_j)| \leqslant \max_i |\mu'_i - \alpha|$$

для любого фиксированного комплексного числа α.

Для доказательства следствия заметим, что P и Q могут быть предположены унитарными. Это означает, что $P^{-1}Q$ унитарна и, следовательно, если мы используем спектральную матричную норму, будем иметь $\nu(A, B) = k(P^{-1}Q) = 1$. Результат следует теперь из первого следствия.

Упр. 3. Если A — нормальная матрица с собственными значениями $\mu_1, \ldots, \mu_n$ и B имеет все элементы равными ε, то доказать, что собственные значения для $A + B$ лежат в кругах

$$\left|z - \left(\frac{n\varepsilon}{2} + \mu_j\right)\right| \leqslant \frac{n}{2}|\varepsilon|, \quad j = 1, 2, \ldots, n.$$

Упр. 4. Если A — нормальная матрица и $B = \operatorname{diag}\{b_1, \ldots, b_n\}$ вещественна, то доказать, что собственные значения для $A + B$ лежат в кругах

$$|z - (\alpha + \mu_j)| \leqslant \beta,$$

где

$$2\alpha = \max_i b_i + \min_i b_i \quad \text{и} \quad 2\beta = \max_i b_i - \min_i b_i.$$

Упр. 5. Пусть $A \in \mathscr{R}_{n\times n}$ и $\mu_1, \ldots, \mu_n$ — собственные значения матрицы $\frac{1}{2}(A+A')$. Если k — спектральный радиус матрицы $\frac{1}{2}(A+A')$ и

$$\min_{j\neq k}|\mu_j-\mu_k|>2k,$$

то доказать, что собственные значения матрицы A вещественны.

Упр. 6. Предположим, что $A, B \in \mathscr{R}_{n\times n}$, $A'=A$ и $|b_{jk}|\leqslant b$ для $j, k=1, 2, \ldots, n$. Если $\mu_1, \ldots, \mu_n$ — собственные значения для A и

$$\min_{j\neq k}|\mu_j-\mu_k|>2nb,$$

то доказать, что собственные значения матрицы $A+B$ вещественны.

7.5. Аналитические возмущения *)

Мы приведем теперь совершенно другой подход к задаче возмущения. Предположим, что $A \in \mathscr{C}_{n\times n}$ — невозмущенная матрица, имеющая собственные значения $\mu_1, \ldots, \mu_n$. Мы предположим, что $A(\zeta)$ — $n\times n$-матрица, элементы которой являются аналитическими функциями ζ в окрестности **) ζ_0, и что $A(\zeta_0)=A$. Для удобства и без потери общности можно считать $\zeta_0=0$, так что $A(0)=A$. Обозначим через $\mu_1(\zeta), \ldots, \mu_n(\zeta)$ собственные значения матрицы $A(\zeta)$. Тогда (§ 7.2), так как собственные значения для $A(\zeta)$ непрерывно зависят от элементов $A(\zeta)$, можно предположить, что $\mu_j(\zeta)\to\mu_j$ при $\zeta\to 0$ для $j=1, 2, \ldots, n$.

Что можно сказать о поведении функций $\mu_j(\zeta)$ для достаточно малых $|\zeta|$? Как мы увидим, в некоторых важных частных случаях функции $\mu_j(\zeta)$ также будут аналитическими в окрестности $\zeta=0$. То, что это не так в общем случае, показывает матрица

$$A(\zeta)=\begin{Vmatrix} 0 & \zeta \\ \zeta^2 & 0 \end{Vmatrix}.$$

Собственные значения для $A(\zeta)$ — это корни многочлена

$$\det(\mu I-A(\zeta))=\mu^2-\zeta^3,$$

т. е. $\mu_1(\zeta)=\zeta^{3/2}$ (определяемое выбором ветви функции $\zeta^{1/2}$) и $\mu_2(\zeta)=-\mu_1(\zeta)$. Таким образом, мы видим, что собственные значения несомненно не аналитичны в окрестности $\zeta=0$.

---

*) Читатель, не знакомый с основными теоремами о функциях комплексной переменной, может пропустить остающуюся часть этой главы.

**) Окрестность точки ζ_0 — открытый круг, имеющий ζ_0 центром.

Следует заметить, что $\det(\mu I - A(\zeta))$ — многочлен от μ, коэффициенты которого являются аналитическими функциями ζ. Из теории алгебраических функций следует тогда, что:

(i) если μ_j — некратное собственное значение матрицы A, то $\mu_j(\zeta)$ аналитична в окрестности $\zeta = 0$;

(ii) если μ_j имеет кратность m и $\mu_k(\zeta) \to \mu_j$ для $k = \alpha_1, \ldots, \alpha_m$, то $\mu_k(\zeta)$ будет аналитической функцией от $\zeta^{1/l}$ в окрестности точки $\zeta = 0$, где $l \leqslant m$ и $\zeta^{1/l}$ — одна из l ветвей функции $\zeta^{1/l}$.

Результат (i) мы докажем вместе с некоторым результатом о соответствующих собственных векторах. Заметим, что в результате (ii) мы *можем* иметь $l = 1$ и $m > 1$, когда $\mu_k(\zeta)$ может быть аналитической. Как и следует ожидать, именно случай (ii) с $m > 1$ представляет реальные трудности для анализа. Разложения по дробным степеням ζ, описываемые в (ii), известны как *ряды Пюизье*. Выполнимость этих разложений будет предполагаться при обсуждениях в §§ 7.9 и 7.10.

Положение становится более сложным в случае, когда мы пытаемся следить за поведением собственных векторов. Матрица

$$A(\zeta) = \begin{Vmatrix} 0 & \zeta \\ 0 & 0 \end{Vmatrix}$$

показывает, что даже в случае, когда собственные значения аналитичны, правые собственные векторы не обязательно непрерывны в $\zeta = 0$.

Тем не менее мы должны начать рассмотрение задачи о возмущении с анализа пространств, натянутых на собственные векторы, поскольку эти пространства определяются компонентными матрицами для $A(\zeta)$ (§ 5.5). Основание для этого в том, что, во-первых, резольвента $R_z(\zeta) = (zI - A(\zeta))^{-1}$ непосредственно выражается через $A(\zeta)$ и, во-вторых, при помощи равенства (5.8.2) мы можем затем исследовать компоненты матрицы для $A(\zeta)$. В частности, мы сможем вывести свойства векторов в образе некоторой компонентной матрицы.

7.6. Возмущение компонентных матриц

Предположим, что некоторое невозмущенное собственное значение μ «расщепляется» при возмущении, определяемом матрицей $A(\zeta)$, на различные собственные значения $\mu_1(\zeta), \ldots, \mu_\beta(\zeta)$. Можно доказать, что сумма первых компонентных матриц для каждого из этих собственных значений аналитична около $\zeta = 0$, и мы показываем в упр. 1, что отдельные первые компонентные матрицы не обязательно обладают этим свойством. Эта теорема является краеугольным камнем, на который опирается весь наш последующий анализ задачи о возмущении.

Теорема 7.6.1. *Пусть* $A(\zeta)$ *аналитична в окрестности* $\zeta = 0$ *и* $A(0) = A$. *Предположим, что* $\mu_1(\zeta), \ldots, \mu_\beta(\zeta)$ — *различные собственные значения матрицы* $A(\zeta)$, *для которых* $\mu_\nu(\zeta) \to \mu$ *при* $\zeta \to 0$, $\nu = 1, 2, \ldots, \beta$, *и пусть* $Z_{\nu 1}(\zeta)$ — *идемпотентная компонентная матрица для* $A(\zeta)$, *соответствующая* $\mu_\nu(\zeta)$. *Если* Z — *идемпотентная компонентная матрица для* A, *соответствующая* μ, *и мы положили*

$$Y(\zeta) = \sum_{\nu=1}^{\beta} Z_{\nu 1}(\zeta),$$

то существует окрестность точки $\zeta = 0$, *в которой имеют место следующие разложения*:

$$Y(\zeta) = Z + \zeta C^{(1)} + \zeta^2 C^{(2)} + \ldots,$$
$$A(\zeta) Y(\zeta) = AZ + \zeta B^{(1)} + \zeta^2 B^{(2)} + \ldots$$

для не зависящих от ζ *матриц* $C^{(r)}$, $B^{(r)}$, $r = 1, 2, \ldots$

Доказательство. Мы предполагаем $|\zeta_0|$ настолько малым, чтобы можно было провести простую замкнутую кривую J, содержащую внутри себя μ, $\mu_1(\zeta), \ldots, \mu_\beta(\zeta)$ и не содержащую других собственных значений A или $A(\zeta)$ для всех ζ с $|\zeta| \leqslant |\zeta_0|$. Записывая тогда $R_z = (Iz - A)^{-1}$ и $R_z(\zeta) = (Iz - A(\zeta))^{-1}$, получаем (по (5.8.2))

$$Z = \frac{1}{2\pi i} \int_J R_z \, dz, \qquad Y(\zeta) = \frac{1}{2\pi i} \int_J R_z(\zeta) \, dz. \tag{7.6.1}$$

Заметим, что при $z \in J$ элементы матрицы $R_z(\zeta)$ являются рациональными выражениями от функций ζ, которые аналитичны около $\zeta = 0$. Таким образом, существуют такие не зависящие от ζ матрицы $R_z^{(r)}$, $r = 1, 2, \ldots$, что

$$R_z(\zeta) = R_z + \zeta R_z^{(1)} + \zeta^2 R_z^{(2)} + \ldots \tag{7.6.2}$$

Используя (7.6.1), для получения первого результата теоремы нам следует лишь проинтегрировать этот ряд почленно по кривой J. Это допустимо, если только каждый коэффициент $R_z^{(r)}$ $(r = 0, 1, 2, \ldots)$ интегрируем по J. Используя упр. 3 из § 5.8, убеждаемся в том, что

$$R_z^{(1)}(\zeta) = -R_z(\zeta) A^{(1)}(\zeta) R_z(\zeta) \tag{7.6.3}$$

и, следовательно, что для каждого r $R_z^{(r)}$ будет линейной комбинацией матриц $R_z(A^{(p)} R_z)^s$ для некоторых целых чисел p, s с $p \leqslant r$. Но $A^{(p)}(0)$ не зависит от z и R_z несомненно непрерывна по z на J. Следовательно, $R_z^{(r)}$ непрерывна на J. Интегрируя ряд для $R_z(\zeta)$ по J, получаем теперь $Y(\zeta) = Z + \zeta C^{(1)} + \zeta^2 C^{(2)} + \ldots$, где $C^{(r)} = \frac{1}{2\pi i}\int\limits_J R_z^{(r)}\,dz, \quad r = 1, 2, \ldots$

Воспользовавшись упр. 4 из § 5.8, получаем

$$A(\zeta)\,Y(\zeta) = \frac{1}{2\pi i}\int\limits_J zR_z(\zeta)\,dz = AZ + \zeta B^{(1)} + \zeta^2 B^{(2)} + \ldots, \tag{7.6.4}$$

где

$$B^{(r)} = \frac{1}{2\pi i}\int\limits_J zR_z^{(r)}\,dz,$$

что завершает доказательство. Может быть доказана подобная этой теорема, по которой из *непрерывности* $A(\zeta)$ в $\zeta = 0$ следует непрерывность $Y(\zeta)$ в $\zeta = 0$.

Упр. 1. Найти матрицы Z, $Z_{11}(\zeta)$, $Z_{21}(\zeta)$ и $Y(\zeta)$ для матрицы

$$A(\zeta) = \begin{Vmatrix} 0 & \zeta \\ \zeta^2 & 0 \end{Vmatrix}.$$

Решение. При $\zeta = 0$ матрица A нулевая и, как легко видеть, $Z = I$. Собственными значениями $A(\zeta)$ будут $\mu_1(\zeta) = \zeta^{3/2}$ и $\mu_2(\zeta) = -\mu_1(\zeta)$. Имеем тогда

$$R_z(\zeta) = \begin{Vmatrix} z & -\zeta \\ -\zeta^2 & z \end{Vmatrix}^{-1} = \frac{1}{z^2 - \zeta^3}\begin{Vmatrix} z & \zeta \\ \zeta^2 & z \end{Vmatrix}.$$

Разложение на рациональные дроби дает

$$R_z(\zeta) = \frac{1}{2}\begin{Vmatrix} \dfrac{1}{z - \zeta^{3/2}} + \dfrac{1}{z + \zeta^{3/2}} & \dfrac{\zeta^{-1/2}}{z - \zeta^{3/2}} - \dfrac{\zeta^{-1/2}}{z + \zeta^{3/2}} \\ \dfrac{\zeta^{1/2}}{z - \zeta^{3/2}} - \dfrac{\zeta^{1/2}}{z + \zeta^{3/2}} & \dfrac{1}{z - \zeta^{3/2}} + \dfrac{1}{z + \zeta^{3/2}} \end{Vmatrix}.$$

Используя (5.8.2), получаем

$$Z_{11}(\zeta) = \frac{1}{2}\begin{Vmatrix} 1 & \zeta^{-1/2} \\ \zeta^{1/2} & 1 \end{Vmatrix} \quad \text{и} \quad Z_{21}(\zeta) = \frac{1}{2}\begin{Vmatrix} 1 & -\zeta^{-1/2} \\ -\zeta^{1/2} & 1 \end{Vmatrix}.$$

Следовательно, $Y(\zeta) = Z_{11}(\zeta) + Z_{21}(\zeta) = I$. Этот пример подчеркивает тот факт, что хотя $Y(\zeta)$ может быть аналитична по ζ, компонентные матрицы (суммой которых $Y(\zeta)$ является) не обязательно аналитичны. В самом деле, в этом примере компонентные матрицы имеют особенность в $\zeta = 0$. Следует также заметить, что хотя собственные значения не аналитичны в окрестности, содержащей $\zeta = 0$, они все же «ведут себя лучше», чем компонентные матрицы.

Упр. 2. Если матрица

$$A(\zeta) = \begin{Vmatrix} 0 & \zeta & 0 \\ 0 & 0 & \zeta \\ \zeta & 0 & 1 \end{Vmatrix}$$

имеет собственные значения $\mu_1(\zeta)$, $\mu_2(\zeta)$ и $\mu_3(\zeta)$ и $\mu_{1,2}(\zeta) \to 0$ при $\zeta \to 0$, то показать, что можно записать

$$\mu_{1,2}(\zeta) = \pm i\zeta^{3/2} - \frac{1}{2}\zeta^3 + \ldots$$

Если $\mu = 0$ — исследуемое невозмущенное собственное значение, то найти Z, $Z_{11}(\zeta)$, $Z_{12}(\zeta)$ и $Y(\zeta)$ и проверить, что $Y(\zeta)$ аналитична в окрестности $\zeta = 0$.

7.7. Возмущение некратного собственного значения

Мы имеем теперь возможность доказать одну теорему о возмущении, имеющую очень большое практическое значение. Обозначения были введены в двух предшествующих параграфах.

Теорема 7.7.1. *Если μ — собственное значение матрицы A кратности* 1, *то для достаточно малого $|\zeta|$ существует такое собственное значение $\mu(\zeta)$ матрицы $A(\zeta)$, что*

$$\mu(\zeta) = \mu + \zeta\mu^{(1)} + \zeta^2\mu^{(2)} + \ldots,$$

существуют правый и левый собственные векторы $\boldsymbol{x}(\zeta)$, $\boldsymbol{y}(\zeta)$ соответственно с собственным значением $\mu(\zeta)$, для которых

$$\boldsymbol{y}'(\zeta)\,\boldsymbol{x}(\zeta) = 1$$

и

$$\boldsymbol{x}(\zeta) = \boldsymbol{x} + \zeta\boldsymbol{x}^{(1)} + \zeta^2\boldsymbol{x}^{(2)} + \ldots,$$
$$\boldsymbol{y}(\zeta) = \boldsymbol{y} + \zeta\boldsymbol{y}^{(1)} + \zeta^2\boldsymbol{y}^{(2)} + \ldots$$

Доказательство. 1. Мы предполагаем всюду далее, что $|\zeta|$ настолько мало, что $\mu(\zeta)$ имеет кратность 1. Тогда будем иметь $Y(\zeta) = Z_{11}(\zeta)$ — компонентной матрице. Кроме того, по терминологии § 2.5 $Y(\zeta)$ — сопутствующая матрица, и мы можем написать

$$Z = \boldsymbol{x}\boldsymbol{y}', \qquad Y(\zeta) = \boldsymbol{x}(\zeta)\,\boldsymbol{y}'(\zeta),$$

где

$$\boldsymbol{y}'\boldsymbol{x} = \boldsymbol{y}'(\zeta)\,\boldsymbol{x}(\zeta) = 1.$$

В таком случае, очевидно, имеем

$$A(\zeta)\,Y(\zeta) = \mu(\zeta)\,Y(\zeta),$$

так что, умножая это равенство слева на y' и справа на x, получаем

$$\mu(\zeta) = \frac{y'A(\zeta)Y(\zeta)x}{y'Y(\zeta)x} = \frac{\mu + \zeta b_1 + \zeta^2 b_2 + \dots}{1 + \zeta c_1 + \zeta^2 c_2 + \dots}$$

после применения теоремы 7.6.1. Коэффициенты справа задаются формулами

$$b_r = y'B^{(r)}x, \quad c_r = y'C^{(r)}x, \quad r = 1, 2, \dots$$

Отсюда следует, что $\mu(\zeta)$ аналитична в окрестности $\zeta = 0$. Первая часть теоремы доказана.

2. Ясно, что в утверждении теоремы $x(\zeta)$ определен лишь с точностью до ненулевой мультипликативной константы (которая может меняться вместе с ζ). Мы определим $x(\zeta)$ более точно, замечая прежде всего, что в определении матрицы $Y(\zeta)$ нет неопределенности. Затем мы выбираем собственные векторы x, y с $y'x = 1$ и, следовательно, $xy' = Z$.

Тогда $Y(\zeta) = x(\zeta)y'(\zeta)$ означает, что

$$y'Y(\zeta)x = (y'x(\zeta))(y'(\zeta)x),$$

а так как $Y(\zeta) \to xy'$ при $\zeta \to 0$, то $(y'x(\zeta))(y'(\zeta)x) \to 1$ при $\zeta \to 0$. Таким образом, можно предположить, что $y'x(\zeta)$ и $y'(\zeta)x$ ненулевые для достаточно малых $|\zeta|$. Теперь мы можем определить $x(\zeta)$, так его нормализуя, чтобы было

$$y'x(\zeta) = 1. \tag{7.7.1}$$

Так как $Y(\zeta)x = x(\zeta)(y'(\zeta)x)$, то $y'Y(\zeta)x = y'(\zeta)x$, так что

$$x(\zeta) = \frac{1}{y'Y(\zeta)x} Y(\zeta)x.$$

Так как матрица $Y(\zeta)$ аналитична в окрестности $\zeta = 0$ и $y'Y(0)x = 1$, то $x(\zeta)$ тоже аналитичен в окрестности $\zeta = 0$. Тот же самый результат получаем для $y(\zeta)$, так как он является правым собственным вектором для $A'(\zeta)$. ◀

Мы доказали существование разложений в степенные ряды по ζ для некратного собственного значения и его собственных векторов. Следующий шаг в анализе задачи о возмущении заключается в обсуждении методов вычисления коэффициентов этих разложений.

7.8. Оценка коэффициентов возмущения

Мы обсудим сначала одну частную задачу, которая нам далее понадобится в нескольких местах. Если μ — собственное значение матрицы $A \in \mathscr{C}_{n\times n}$ и $b \neq 0$, то мы должны решить уравнение

$$(\mu I - A)x = b. \tag{7.8.1}$$

Как мы знаем (теорема 1.16.3), вектор-решение существует $\Leftrightarrow$ $\boldsymbol{b} \in \mathscr{R}(A - \mu I)$. Необходимое условие для существования решения может быть сформулировано в терминах левого собственного пространства $\mathscr{M}$ для μ. Мы пишем $\boldsymbol{y} \in \mathscr{M}$ в том и только в том случае, когда $\boldsymbol{y}'(\mu I - A) = \boldsymbol{0}'$. Ясно, что $\boldsymbol{b}$ должен удовлетворять условию $\boldsymbol{y}'\boldsymbol{b} = 0$ для каждого $\boldsymbol{y} \in \mathscr{M}$, т. е. $\boldsymbol{b}$ и $\boldsymbol{y}$ квазиортогональны для каждого $\boldsymbol{y} \in \mathscr{M}$.

Мы стремимся выразить решение $\boldsymbol{x}$ для (7.8.1) через спектральные свойства матрицы A. Если $\mu = \mu_1, \mu_2, \ldots, \mu_s$ — различные собственные значения, то мы сначала определяем матрицу

$$(I - Z_{11})(I\mu - A)^{-1}(I - Z_{11}), \tag{7.8.2}$$

ограничение резольвенты на образ $I - Z_{11}$
(§§ 5.4 и 5.5 и равенство (5.8.1)). Принимая во внимание спектральное разложение для A (упр. 1 из § 5.4)

$$A = \mu Z_{11} + Z_{12} + \sum_{k=2}^{s} (\mu_k Z_{k1} + Z_{k2}),$$

после некоторых преобразований можно показать, что

$$\mu E - EA = I - Z_{11}. \tag{7.8.3}$$

Таким образом, умножая (7.8.1) слева на E, получаем

$$(I - Z_{11})\boldsymbol{x} = E\boldsymbol{b}. \tag{7.8.4}$$

Особенно мы будем заинтересованы в векторах-решениях $\boldsymbol{x}$, которые удовлетворяют также условию $\boldsymbol{y}'\boldsymbol{x} = 0$ для каждого $\boldsymbol{y} \in \mathscr{M}$. Такое решение нетрудно получить, если индекс μ равен 1, ибо в этом случае компонентная матрица Z_{11} будет суммой сопутствующих матриц (§ 2.5). Итак, для квазибиортогональных систем базисных векторов $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_\alpha$ (для $\mathscr{M}$) и $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\alpha$ (для $\mathscr{N}(A - \mu I)$) имеем

$$Z_{11} = \sum_{j=1}^{\alpha} \boldsymbol{x}_j \boldsymbol{y}_j'.$$

Ясно теперь, что для такой Z_{11} имеем $\boldsymbol{y}'\boldsymbol{x} = 0$ при каждом $\boldsymbol{y} \in \mathscr{M} \Leftrightarrow Z_{11}\boldsymbol{x} = \boldsymbol{0}$. Подстановкой в (7.8.4) получаем следующий результат.

Лемма. *Если μ — собственное значение индекса* 1 *матрицы A и $\mathscr{M}$ — левое собственное подпространство для μ, то вектор $\boldsymbol{\xi} = E\boldsymbol{b}$ (где E задается равенством* (7.8.2)) *будет решением уравнения $(\mu I - A)\boldsymbol{x} = \boldsymbol{b}$ (где $\boldsymbol{b} \in \mathscr{R}(\mu I - A)$), для которого $\boldsymbol{y}'\boldsymbol{\xi} = 0$ при каждом $\boldsymbol{y} \in \mathscr{M}$.*

Упр. 1. Если $\mathscr{N}^* = \mathscr{N}(\mu I - A)^*$ и $\mathscr{M}$ — левое собственное подпространство для $\mu I - A$, то доказать, что два следующих условия эквивалентны:

(i) $\boldsymbol{y}'\boldsymbol{b} = 0$ для каждого $\boldsymbol{y} \in \mathscr{M}$.

(ii) $\boldsymbol{b} \in \mathscr{C}_n \ominus \mathscr{N}^*$.

Затем, используя упр. 5 на стр. 120, доказать, что уравнение (7.8.1) имеет решение $\Longleftrightarrow$ выполняется условие (i).

Упр. 2. Получить равенство (7.8.3).

***Упр. 3.** Если $\mu = \mu_1$ — собственное значение A и $\mu_1, \ldots, \mu_s$ — различные собственные значения A, то показать, что равенство (5.8.1.) можно записать в виде

$$(zI - A)^{-1} = \sum_{j=1}^{m_1} \frac{(j-1)!\, Z_{1j}}{(z-\mu)^j} + E(z),$$

где $E(z)$ — аналитическая в окрестности μ матрица и $Z_{11}E(z) = E(z)Z_{11} = 0$. Полагая $E = E(\mu)$, использовать это равенство и теорему 5.5.1 для получения (7.8.3).

Предположим теперь, что выполняются предположения теоремы 7.7.1. Тогда μ имеет кратность 1, и для достаточно малых $|\zeta|$ имеем

$$(A + \zeta A^{(1)} + \zeta^2 A^{(2)} + \ldots)(\boldsymbol{x} + \zeta \boldsymbol{x}^{(1)} + \zeta^2 \boldsymbol{x}^{(2)} + \ldots) = (\mu + \zeta\mu^{(1)} + \ldots)(\boldsymbol{x} + \zeta\boldsymbol{x}^{(1)} + \ldots).$$

Можно приравнять коэффициенты при ζ^n, получаемые при перемножении этих рядов, для $n = 0, 1, 2, \ldots$ Коэффициент при ζ^0 приводит к равенству для невозмущенного собственного значения: $A\boldsymbol{x} = \mu\boldsymbol{x}$. Приравнивание коэффициентов при ζ дает

$$(\mu I - A)\,\boldsymbol{x}^{(1)} = (A^{(1)} - \mu^{(1)} I)\,\boldsymbol{x}. \tag{7.8.5}$$

Если $\boldsymbol{y}$ — левый собственный вектор для μ, то мы знаем, что, так как $\boldsymbol{x}^{(1)}$ существует, $\boldsymbol{y}'(A^{(1)} - \mu^{(1)} I)\boldsymbol{x} = 0$. Так как $\boldsymbol{y}'\boldsymbol{x} = 1$, из этого условия получаем

$$\mu^{(1)} = \boldsymbol{y}' A^{(1)} \boldsymbol{x}. \tag{7.8.6}$$

Нам известно также, что по определению вектора $\boldsymbol{x}(\zeta)$ соотношение (7.7.1) приводит к равенству

$$\boldsymbol{y}'(\boldsymbol{x} + \zeta\boldsymbol{x}^{(1)} + \zeta^2\boldsymbol{x}^{(2)} + \ldots) = 1,$$

откуда следует, что

$$\boldsymbol{y}'\boldsymbol{x}^{(r)} = 0, \qquad r = 1, 2, \ldots \tag{7.8.7}$$

Воспользуемся теперь результатом леммы. В данном случае пространство $\mathscr{M}$ порождается $\boldsymbol{y}$, так что решением уравнения (7.8.5), согласованным с условием $\boldsymbol{y}'\boldsymbol{x}^{(1)} = 0$, будет

$$\boldsymbol{x}^{(1)} = E\,(A^{(1)} - \mu^{(1)} I)\,\boldsymbol{x}.$$

Кроме того, так как $Z_{11} = xy'$, можно записать $Ex = EZ_{11}x$ и часть (ii) теоремы 5.5.1 означает, что $Ex = 0$. Таким образом, мы завершаем поиск решения для коэффициентов возмущения первого порядка следующим образом:

$$\mu^{(1)} = y'A^{(1)}x, \qquad x^{(1)} = EA^{(1)}x. \tag{7.8.8}$$

Приведенное рассуждение легко обобщается до получения рекурсивных соотношений для вычисления $\mu^{(r)}$ и $x^{(r)}$ при любом r. Приравнивание коэффициентов при ζ^r в $A(\zeta)x(\zeta) = \mu(\zeta)x(\zeta)$ приводит к равенству

$$Ax^{(r)} + A^{(1)}x^{(r-1)} + \ldots + A^{(r)}x = \mu x^{(r)} + \mu^{(1)}x^{(r-1)} + \ldots + \mu^{(r)}x,$$

или

$$(\mu I - A)\,x^{(r)} = (A^{(1)} - \mu^{(1)}I)\,x^{(r-1)} + \ldots + (A^{(r)} - \mu^{(r)}I)\,x.$$

Если предположить теперь, что $\mu^{(1)}, \ldots, \mu^{(r-1)}$ и $x^{(1)}, \ldots, x^{(r-1)}$ известны, то условие квазиортогональности для правой части приводит к

$$\mu^{(r)} = y'A^{(r)}x + y'A^{(r-1)}x^{(1)} + \ldots + y'A^{(1)}x^{(r-1)} \tag{7.8.9}$$

для $r = 1, 2, \ldots$ и $x^{(0)} = x$. По лемме тогда получаем

$$x^{(r)} = E\,[(A^{(1)} - \mu^{(1)}I)\,x^{(r-1)} + \ldots$$
$$\ldots + (A^{(r-1)} - \mu^{(r-1)}I)\,x^{(1)} + A^{(r)}x]. \tag{7.8.10}$$

Во многих приложениях возмущения матрицы A *линейны* по ζ. Таким образом, мы имеем $A(\zeta) = A + \zeta B$ для некоторой матрицы B, не зависящей от ζ. В этом случае $A^{(1)} = B$ и $A^{(2)} = A^{(3)} = \ldots = 0$. Равенства (7.8.9) и (7.8.10) превращаются тогда в

$$\mu^{(r)} = y'Bx^{(r-1)}, \qquad x^{(r)} = E\,(B - x^{(r-1)} - \sum_{j=1}^{r-1} \mu^{(j)}x^{(r-j)})$$

для $r = 1, 2, \ldots$ Напомним, что в этих формулах верхние индексы не обозначают производные. Например, имеем

$$\mu^{(r)} = \frac{1}{r!}\frac{d^r\mu}{d\zeta^r}\bigg|_{\zeta=0}.$$

Упр. 4. Показать, что равенство (7.8.9) в случае $r = 2$ можно записать в виде

$$\mu^{(2)} = \operatorname{tr}(A^{(2)}Z_{11}) + \operatorname{tr}(A^{(1)}EA^{(1)}Z_{11}).$$

7.9. Возмущение кратного собственного значения

Мы можем теперь рассматривать решенной задачу об аналитических возмущениях некратного собственного значения. Результат (i) из § 7.5 получен. Если μ — собственное значение матрицы A кратности $m > 1$, то мы должны обратиться к представлению возмущенного собственного значения $\mu(\zeta)$, которое стремится к μ при $\zeta \to 0$ (результат (ii) из § 7.5), в виде ряда Пюизье. В § 7.6 были получены некоторые дополнительные сведения о поведении компонентных матриц для возмущенных собственных значений. Как мы убедились, хотя их сумма $Y(\zeta)$ аналитична для достаточно малых $|\zeta|$ (теорема 7.6.1), отдельные компонентные матрицы могут иметь особенности в $\zeta = 0$ (упр. 1 из § 7.6).

Предположим теперь, что на матрицу $A(\zeta)$ наложено дополнительное условие о том, что она *эрмитова* и аналитична в окрестности $\zeta = 0$. Непосредственное рассмотрение уравнений Коши — Римана показывает тогда, что во избежание тривиальности ζ следует считать вещественным. Тогда возмущенные собственные значения должны быть вещественны и сопровождающие матрицы должны быть эрмитовы. Эти дополнительные ограничения в задаче, как можно показать, означают, что возмущенные собственные значения аналитичны по ζ независимо от того, какова кратность невозмущенного собственного значения. Возмущенные сопровождающие матрицы также будут аналитичны, и, следовательно, собственные векторы матрицы $A(\zeta)$ можно считать аналитичными и ортонормированными всюду в окрестности $\zeta = 0$. Эти важные результаты были впервые доказаны Ф. Реллихом в начале 1930-х годов. Первое доказательство*) Реллиха зависит в основном от того, что возмущенные собственные значения вещественны. Мы докажем результат (теорема 7.10.1), содержащий перечисленные выше результаты, при помощи более современной и простой техники, принадлежащей Като.

Мы рассмотрим сначала задачу о возмущении, не делая предположения относительно симметричности первоначальной матрицы. Мы докажем, что если μ — собственное значение матрицы A индекса 1 и $\mu_j(\zeta)$ — собственное значение матрицы $A(\zeta)$, для которого $\mu_j(\zeta) \to \mu$ при $\zeta \to 0$, то при $\zeta \to 0$

$$\mu_j(\zeta) = \mu + a_j\zeta + O\left(|\zeta|^{1+(1/l)}\right).$$

Таким образом, в разложении Пюизье для $\mu_j(\zeta)$, описанном в результате (ii) из § 7.5, коэффициенты при $\zeta^{j/l}$ равны нулю для $j = 1, 2, \ldots, l-1$.

Прежде чем приступить к доказательству этого результата, еще раз подчеркнем, что всюду до конца этой главы $A(\zeta)$ аналитична от комплексной переменной ζ в окрестности $\zeta = 0$ и $A(0) = A$. Кроме того, $\mu_j(\zeta)$ — собственное значение матрицы $A(\zeta)$, для которого $\mu_j(\zeta) \to \mu$ при $\zeta \to 0$, и $\mu_j(\zeta)$ имеет разложе-

*) F. Rellich, Math. Ann. **113** (1937), 602.

ние в степенной ряд около $\zeta = 0$ по степеням $\zeta^{1/l}$ для некоторого положительного целого числа l.

Лемма 1. *Существует правый собственный вектор $\boldsymbol{x}(\zeta)$, соответствующий $\mu_j(\zeta)$ для всех ζ из некоторой окрестности $\mathcal{N}$ точки $\zeta = 0$ и обладающий такими свойствами, что при евклидовой векторной норме $h(\boldsymbol{x}(\zeta)) = 1$ для $\zeta \in \mathcal{N}$ и $\boldsymbol{x}(\zeta)$ аналитичен по $\zeta^{1/l}$ для $\zeta \in \mathcal{N}$.*

Доказательство. Положив $B(\zeta) = A(\zeta) - \mu_j(\zeta) I$, можем считать элементы матрицы B аналитичными по $\eta = \zeta^{1/l}$. Мы построим некоторый вектор $\boldsymbol{x}$, который аналитичен по η и обладает свойством: $B(\eta)\boldsymbol{x}(\eta) = \boldsymbol{0}$ для всех η в некоторой окрестности точки $\eta = 0$.

Предположим, что $B(\eta)$ имеет ранг r в некоторой окрестности точки $\eta = 0$. Для простоты и без потери общности будем считать, что минор порядка r

$$B\begin{pmatrix} 1, & 2, & \ldots, & r \\ 1, & 2, & \ldots, & r \end{pmatrix}$$

не равен нулю в некоторой окрестности $\eta = 0$. Рассмотрим тогда разложение минора

$$B\begin{pmatrix} 1, & 2, & \ldots, & r+1 \\ 1, & 2, & \ldots, & r+1 \end{pmatrix}$$

по его последней строке. Обозначим через $l_1(\eta), \ldots, l_{r+1}(\eta)$ алгебраические дополнения для $b_{r+1,1}(\eta), \ldots, b_{r+1,r+1}(\eta)$ в этом определителе. Таким образом,

$$B\begin{pmatrix} 1, & 2, & \ldots, & r+1 \\ 1, & 2, & \ldots, & r+1 \end{pmatrix} = \sum_{j=1}^{r+1} b_{r+1,j}(\eta)\, l_j(\eta) = 0.$$

Определим теперь $l_{r+2}(\eta) = \ldots = l_n(\eta) \equiv 0$ и затем вектор $\boldsymbol{l}(\eta)$ с элементами $l_1(\eta), \ldots, l_n(\eta)$. Заметим, что

$$B_{j*}(\eta)\,\boldsymbol{l}(\eta) = \sum_{k=1}^{n} b_{jk}(\eta)\, l_k(\eta) = B\begin{pmatrix} 1, & 2, & \ldots, & r, & j \\ 1, & 2, & \ldots, & r, & r+1 \end{pmatrix} = 0.$$

Мы получаем нуль справа для $j = 1, 2, \ldots, r$ потому, что определитель имеет две одинаковые строки, и нуль для $j = r+1, \ldots \ldots, n$ потому, что в этом случае имеем определитель, совпадающий с минором в B порядка $r+1$. Таким образом, $B(\eta)\boldsymbol{l}(\eta) = 0$.

Для завершения доказательства мы должны лишь показать, что $\boldsymbol{l}$ может быть соответствующим образом нормализован. Это предоставляется сделать читателю в качестве упражнения.

Заметим, что этой леммой можно было бы воспользоваться в теореме 7.7.1 для доказательства того, что правый собственный вектор $\boldsymbol{x}(\zeta)$ можно считать аналитическим, если известно, что $\boldsymbol{\mu}(\zeta)$ аналитично.

Лемма 2. *Для собственного вектора* $\boldsymbol{x}(\zeta)$, *определенного в лемме* 1,

$$h((I - Z_{11})\boldsymbol{x}(\zeta)) = O(|\zeta|) \quad \textit{при} \quad \zeta \to 0.$$

Заметим, что, полагая $\boldsymbol{x} = Z_{11}\boldsymbol{x} + (I - Z_{11})\boldsymbol{x}$, получаем первую составляющую $\boldsymbol{x}$ в образе Z_{11} (которая в точности совпадает с правым собственным подпространством для μ, если μ имеет индекс 1: теорема 5.5.3), а вторая составляющая появляется в формулировке леммы. Мы утверждаем, что величина этой составляющей $\boldsymbol{x}(\zeta)$ стремится к нулю, как $|\zeta|$, при $\zeta \to 0$, даже если величина $\boldsymbol{x}(\zeta)$ стремится к нулю, как $|\zeta|^{1/l}$.

Доказательство. Можно записать $A(\zeta) = A + \zeta K(\zeta)$, где $K(\zeta)$ аналитична около $\zeta = 0$. Тогда уравнение $(A(\zeta) - \mu_j(\zeta)I)\boldsymbol{x}(\zeta) = \boldsymbol{0}$ можно представить в таком виде:

$$(\mu_j(\zeta)I - A)\boldsymbol{x}(\zeta) = \zeta K(\zeta)\boldsymbol{x}(\zeta). \tag{7.9.1}$$

Случай (i). Предположим, что $\mu_j(\zeta) \equiv \mu$ в окрестности $\zeta = 0$. Тогда, умножая это равенство слева на матрицу E из (7.8.2) и используя равенство (7.8.3), получаем

$$(I - Z_{11})\boldsymbol{x}(\zeta) = \zeta E K(\zeta)\boldsymbol{x}(\zeta).$$

Так как $h(\boldsymbol{x}(\zeta)) = 1$, лемма следует отсюда сразу же. Единственная другая возможность есть

Случай (ii). В некоторой окрестности точки $\zeta = 0$ имеем $\mu_j(\zeta) \neq \mu$, исключая $\zeta = 0$. Умножая слева равенство (7.9.1) на $(\mu_j(\zeta)I - A)^{-1}$ и используя выражение для резольвенты, данное в упр. 3 из § 7.8, получаем

$$\boldsymbol{x}(\zeta) = \left\{ \sum_{j=1}^{m_1} \frac{(j-1)!\, Z_{1j}}{(\mu_j(\zeta) - \mu)^j} + E(\mu_j(\zeta)) \right\} \zeta K(\zeta)\boldsymbol{x}(\zeta).$$

Умножив теперь это равенство на $I - Z_{11}$ слева. Так как $Z_{11}Z_{1j} = Z_{1j}$ и $Z_{11}E(z) = 0$, то находим, что

$$(I - Z_{11})\boldsymbol{x}(\zeta) = \zeta E(\mu_j(\zeta)) K(\zeta)\boldsymbol{x}(\zeta).$$

Так как E аналитична по $\zeta^{1/l}$, K аналитична по ζ и $h(\boldsymbol{x}(\zeta)) = 1$, то результат доказан также и в случае (ii). ◀

Умножив равенство (7.9.1) на Z_{11} слева, получим

$$\mu_j(\zeta)Z_{11}\boldsymbol{x}(\zeta) - Z_{11}A\boldsymbol{x}(\zeta) = \zeta Z_{11}K(\zeta)\boldsymbol{x}(\zeta).$$

Если предположить, что μ имеет индекс 1, то $Z_{11}\boldsymbol{x}(\zeta)$ обязательно будет правым собственным вектором для A, соответствующим μ (теорема 5.5.3), а так как $AZ_{11} = Z_{11}A$, то

$$\mu Z_{11}\boldsymbol{x}(\zeta) - Z_{11}A\boldsymbol{x}(\zeta) = \boldsymbol{0}.$$

Эти два равенства означают, что для $\zeta \neq 0$

$$\frac{\mu_j(\zeta)-\mu}{\zeta} Z_{11}\boldsymbol{x}(\zeta) = Z_{11}K(\zeta)\{Z_{11}\boldsymbol{x}(\zeta) + (I - Z_{11})\boldsymbol{x}(\zeta)\}.$$

Следовательно, воспользовавшись евклидовой матричной нормой, имеем

$$h\left[\frac{\mu_j(\zeta)-\mu}{\zeta} Z_{11}\boldsymbol{x}(\zeta) - Z_{11}K(\zeta)Z_{11}\boldsymbol{x}(\zeta)\right] \leqslant \| Z_{11}K(\zeta)\| h[(I - Z_{11})\boldsymbol{x}(\zeta)].$$

Согласно лемме 2 правая часть стремится к нулю при $\zeta \to 0$, так что то же самое должно быть верно для левой части. Так как $\lim\limits_{\zeta\to 0} Z_{11}K(\zeta)Z_{11}\boldsymbol{x}(\zeta)$ тоже существует, то отсюда следует, что найдется такое число a_j, что

$$\lim_{\zeta\to 0}\frac{\mu_j(\zeta)-\mu}{\zeta} = a_j.$$

Кроме того, полагая $\boldsymbol{x} = \boldsymbol{x}(0)$ и замечая, что $Z_{11}\boldsymbol{x} = \boldsymbol{x}$ и $\lim\limits_{\zeta\to 0} K(\zeta) = A^{(1)}$ (первая производная матрица $A(\zeta)$ в $\zeta = 0$), получаем

$$a_j\boldsymbol{x} = Z_{11}A^{(1)}Z_{11}\boldsymbol{x}.$$

Таким образом, a_j является собственным значением матрицы $Z_{11}A^{(1)}Z_{11}$ с правым собственным вектором $\boldsymbol{x}$. Нами доказана

Теорема 7.9.1. *Пусть $A(\zeta)$ — аналитическая по ζ в окрестности $\zeta = 0$ матрица и $A(0) = A$. Пусть μ — собственное значение для A индекса 1 и $\mu_j(\zeta)$ — собственное значение для $A(\zeta)$ с $\mu_j(0) = \mu$. Тогда существует такое число a_j, что*

$$\mu_j(\zeta) = \mu + a_j\zeta + O\left(|\zeta|^{1+(1/l)}\right)$$

при $\zeta \to 0$ и a_j — собственное значение для $Z_{11}A^{(1)}Z_{11}$.

Кроме того, каждому $\mu_j(\zeta)$ соответствует по крайней мере один собственный вектор $\boldsymbol{x}_j(\zeta)$, для которого $h(\boldsymbol{x}_j(\zeta)) = 1$ при достаточно малом $|\zeta|$:

$$\boldsymbol{x}_j(\zeta) = \boldsymbol{x} + \zeta^{1/l}\boldsymbol{x}_1 + \zeta^{2/l}\boldsymbol{x}_2 + \ldots,$$

$$\boldsymbol{x},\ \boldsymbol{x}_1,\ \ldots,\ \boldsymbol{x}_{l-1} \in \mathcal{N}(A - \mu I) \quad \text{и} \quad a_j\boldsymbol{x} = Z_{11}A^{(1)}Z_{11}\boldsymbol{x}.$$

Упр. 1. Найти $Z_{11}A^{(1)}Z_{11}$ для матрицы $A(\zeta)$ в упр. 2 из § 7.6 и показать, что результат теоремы 7.9.1 согласуется с разложениями для $\mu_{1,2}(\zeta)$, данными в этом примере.

Упр. 2. Проверить результаты теоремы 7.9.1 для матрицы $A(\zeta)$ в упр. 1 из § 7.6 и для матрицы

$$\left\|\begin{matrix} 0 & \zeta - 1 \\ \zeta(\zeta-1) & 0 \end{matrix}\right\|.$$

Упр. 3. Доказать, что если a_j — собственное значение матрицы $Z_{11}A^{(1)}Z_{11}$ индекса 1, то существует такое число b_j, что

$$\mu_j(\zeta) = \mu + a_j\zeta + b_j\zeta^2 + O\left(|\zeta|^{2+1/l}\right)$$

и b_j — собственное значение для $Z_{11}(A^{(1)}EA^{(1)} + A^{(2)})Z_{11}$ с правым собственным вектором $\boldsymbol{x}$.

7.10. Редукционный процесс

Теорему 7.6.1 можно использовать в исследовании возмущений собственного значения индекса 1 другим способом. «Редукционный процесс», который мы собираемся описать, недавнего происхождения и принадлежит Като *).

Используя (7.6.1) и (7.6.2), можем написать:

$$(I\mu - A(\zeta))Y(\zeta) = \frac{-1}{2\pi i}\int_J (z-\mu)R_z(\zeta)\,dz = \frac{-1}{2\pi i}\int_J (z-\mu)\sum_{r=0}^{\infty}\zeta^r R_z^{(r)}\,dz.$$

Но $Y(0) = Z$, и если μ имеет индекс 1, то $(I\mu - A)Z = 0$ и член справа, который не зависит от ζ, должен тоже быть равным нулю. Таким образом, мы записываем

$$(I\mu - A(\zeta))Y(\zeta) = \sum_{n=1}^{\infty}\zeta^n\mathscr{A}^{(n)} \qquad (7.10.1)$$

и отмечаем, что

$$\mathscr{A}^{(1)} = \frac{-1}{2\pi i}\int_J (z-\mu)R_z^{(1)}\,dz.$$

Для оценки этого интеграла воспользуемся (7.6.3):

$$R_z^{(1)} = -(Iz - A)^{-1}A^{(1)}(Iz-A)^{-1} =$$
$$= -\left\{\frac{Z}{z-\mu} + E(z)\right\}A^{(1)}\left\{\frac{Z}{z-\mu} + E(z)\right\},$$

согласно упр. 3 из § 7.8, где $E(z)$ аналитична в окрестности μ. Подставляя это выражение в интеграл для $\mathscr{A}^{(1)}$, из теоремы о вычетах получаем

$$\mathscr{A}^{(1)} = ZA^{(1)}Z. \qquad (7.10.2)$$

Воспользовавшись (7.10.1), определим теперь

$$\mathscr{A}(\zeta) = \zeta^{-1}(\mu I - A(\zeta))Y(\zeta) = \sum_{r=0}^{\infty}\zeta^r\mathscr{A}^{(r+1)} \qquad (7.10.3)$$

*) См. дополнение 3.

и заметим, что $\mathscr{A}(0) = ZA^{(1)}Z$ и $\mathscr{A}(\zeta)$ аналитична в окрестности $\zeta = 0$.

С помощью разложения Пюизье для собственных значений матрицы $A(\zeta)$ можно теперь легко доказать первую часть теоремы 7.9.1. Пусть $\boldsymbol{x}_j(\zeta)$ — правый собственный вектор для $A(\zeta)$ с собственным значением $\mu_j(\zeta)$. Тогда $\boldsymbol{x}_j(\zeta)$ обязательно принадлежит образу проекции $Y(\zeta)$. Таким образом, $\boldsymbol{x}_j(\zeta) = Y(\zeta)\boldsymbol{x}_j(\zeta)$ и, умножая (7.10.3) справа на $\boldsymbol{x}_j(\zeta)$, получаем

$$\mathscr{A}(\zeta)\,\boldsymbol{x}_j(\zeta) = \zeta^{-1}(\mu - \mu_j(\zeta))\,\boldsymbol{x}_j(\zeta).$$

Таким образом, $\zeta^{-1}(\mu - \mu_j(\zeta))$ является собственным значением для $\mathscr{A}(\zeta)$, а так как таковое имеет разложение Пюизье около $\zeta = 0$ по степеням $\zeta^{1/l}$, например, то мы имеем

$$\mu_j(\zeta) = \mu + \zeta\mu^{(1)} + O\left(|\zeta|^{1+(1/l)}\right).$$

Кроме того, постоянный член в разложении Пюизье $\mu^{(1)}$ является собственным значением невозмущенной матрицы $ZA^{(1)}Z$.

Ясно, что если $\mu^{(1)}$ — собственное значение для $ZA^{(1)}Z$ индекса 1, то мы можем продолжить редукционный процесс еще на один шаг и получить результат упр. 3 из § 7.9:

$$\mu_j(\zeta) = \mu + \zeta\mu^{(1)} + \zeta^2\mu^{(2)} + O\left(|\zeta|^{2+(1/l)}\right).$$

В противоположность методам из § 7.9 изложенный редукционный процесс имеет то преимущество, что он *сохраняет симметричность*. Так, предположим, что $A(\varepsilon)$ нормальна для вещественного ε. Тогда $Y(\varepsilon) = Y(\varepsilon)^*$ (ср. упр. 4 из § 5.5) и, так как $A(\varepsilon)$ и $Y(\zeta)$ коммутируют, мы можем написать

$$\mathscr{A}(\varepsilon) = \zeta^{-1}Y(\varepsilon)\,\{I\mu - A(\varepsilon)\}\,Y(\varepsilon).$$

Проверка того, что $\mathscr{A}(\zeta)$ нормальна, является нетрудным упражнением. Таким образом, $\mathscr{A}(0)$ нормальна и *все* собственные значения матрицы $ZA^{(1)}Z$ должны иметь индекс 1. Так как симметричность сохраняется при редукционном процессе и невозмущенные матрицы нормальны на каждом шаге, то собственные значения $\mu_j(\zeta)$ имеют разложения по *целым* степеням ζ вблизи $\zeta = 0$. Поэтому нами доказана

Теорема 7.10.1. *Пусть $A(\zeta)$ аналитична и нормальна при вещественном ζ. Тогда собственные значения матрицы $\mathscr{A}(\zeta)$ аналитичны в некоторой окрестности $\zeta = 0$.*

Глава 8

ПРЯМЫЕ ПРОИЗВЕДЕНИЯ, РЕШЕНИЕ МАТРИЧНЫХ УРАВНЕНИЙ И ЗАДАЧИ УСТОЙЧИВОСТИ

8.1. Введение

Эта глава состоит из трех разделов, находящихся в последовательной зависимости. Мы исследуем сначала некоторую новую операцию, называемую прямым произведением, которая в полной общности является бинарной операцией из $\mathcal{F}_{k\times l} \times \mathcal{F}_{m\times n}$ в $\mathcal{F}_{km\times ln}$. Однако мы ограничимся рассмотрением случая $k = l$, $m = n$. Это наиболее полезный случай в приложениях, в котором имеются красивые результаты относительно зависимости собственных значений матрицы-произведения от сомножителей.

Знание прямого произведения облегчает анализ матричных уравнений. Мы сосредоточимся здесь на линейных уравнениях от матричной переменной, коэффициенты которых будут тоже матрицы. Полезные результаты известны также для полиномиальных уравнений более высокой степени, которые не слишком недоступны, но мы не будем ими заниматься, так как для них нет непосредственного приложения.

Третий раздел имеет наибольшее практическое значение. Матрица $A \in \mathscr{C}_{n\times n}$, все собственные значения которой имеют отрицательные вещественные части, называется *устойчивой*. Это понятие возникает при изучении решений дифференциального уравнения $\dot{x} = Ax$ (§ 5.9), где точка обозначает дифференцирование по времени. Такое уравнение с устойчивой матрицей коэффициентов A, как известно, имеет лишь устойчивые решения в интуитивно понятном смысле. Наше прежнее изучение решения матричных уравнений будет иметь большое значение при изложении подхода Ляпунова к задаче об устойчивости.

8.2. Прямое произведение

Если $A \in \mathcal{F}_{m\times m}$ и $B \in \mathcal{F}_{n\times n}$, то *прямое произведение* A и B (в записи $A \otimes B$) определяется как блочная матрица

$$A \otimes B = \begin{Vmatrix} a_{11}B & a_{12}B & \dots & a_{1m}B \\ a_{21}B & a_{22}B & \dots & a_{2m}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mm}B \end{Vmatrix} \in \mathcal{F}_{mn\times mn}.$$

В определении несомненна некоторая степень произвольности. Так, можно спросить, почему $A \otimes B$ не определяется как

$$\begin{Vmatrix} Ab_{11} & \dots & Ab_{1n} \\ \vdots & & \vdots \\ Ab_{n1} & \dots & Ab_{nn} \end{Vmatrix}?$$

Ответ состоит в том, что для наших целей, используя два возможных определения, можно было бы получить параллельные свойства и, что касается приложений, оба определения были бы одинаково полезны. Понятие прямого произведения естественно возникает в теории групп и в результате имеет важные приложения в физике элементарных частиц.

Нам понадобится явный вид элементов матрицы $A \otimes B$. Как нетрудно видеть, для $r, i = 1, 2, \dots, m$ и $s, j = 1, 2, \dots, n$ *элементом в строке* $(r-1)n+s$ *и столбце* $(i-1)n+j$ *будет* $a_{ri}b_{sj}$. Таким образом, если положить $C = A \otimes B$ и

$$k = (r-1)n+s, \qquad l = (i-1)n+j,$$

где $1 \leqslant r, i \leqslant m$, $1 \leqslant s, j \leqslant n$, то $c_{kl} = a_{ri}b_{sj}$.

Следующие свойства прямого произведения непосредственно следуют из определения. Читатель может провести любую необходимую проверку. Предполагается, что размерности входящих в формулы матриц таковы, что все операции определены.

(а) Если $\mu \in \mathcal{F}$, то $(\mu A) \otimes B = A \otimes (\mu B) = \mu (A \otimes B)$.

(б) $(A + B) \otimes C = A \otimes C + B \otimes C$.

(в) $A \otimes (B + C) = A \otimes B + A \otimes C$.

(г) $A \otimes (B \otimes C) = (A \otimes B) \otimes C$.

(д) $(A \otimes B)' = A' \otimes B'$.

Отметим, что, вообще говоря, прямое произведение не коммутативно. Нам понадобится еще одно свойство, которое менее очевидно и окажется очень полезным. Оно содержится в следующей теореме.

Теорема 8.2.1. *Если* $A, C \in \mathcal{F}_{m\times m}$ *и* $B, D \in \mathcal{F}_{n\times n}$, *то* $(A \otimes B)(C \otimes D) = AC \otimes BD$.

Доказательство. Как мы видели, элемент (k, ν) матрицы $E = A \otimes B$ равен $e_{k\nu} = a_{ru}b_{sv}$, где $k = (r-1)n+s$, $\nu = (u-1)n+v$ и $1 \leqslant r$, $u \leqslant m$, $1 \leqslant s$, $v \leqslant n$. Подобно этому элемент (ν, l) матрицы $F = C \otimes D$ задается в виде

$$f_{\nu l} = c_{ui}d_{vj}, \qquad l = (i-1)n+j,$$

где $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$.

Таким образом, элемент (k, l) матрицы $G = (A \otimes B)(C \otimes D)$ равен

$$g_{kl} = \sum_{\nu=1}^{mn} e_{k\nu}f_{\nu l} = \sum_{\nu=1}^{mn} a_{ru}b_{sv}c_{ui}d_{vj}.$$

Используя выражение для ν, мы пишем теперь $\sum\limits_{u=1}^{m}\sum\limits_{v=1}^{n}$ вместо $\sum\limits_{\nu=1}^{mn}$ и получаем

$$g_{kl} = \sum_{u=1}^{m} a_{ru}c_{ui}\sum_{v=1}^{n} b_{sv}d_{vj} = (AC)_{ri}(BD)_{sj}.$$

Так как $AC \in \mathscr{F}_{m\times m}$, $BD \in \mathscr{F}_{n\times n}$ и индексы имеют соответствующие пределы, убеждаемся в том, что $G = AC \otimes BD$.

Эта теорема приводит к некоторым важным выводам, которые мы перечисляем в следующем следствии.

Следствие. *Если $A \in \mathscr{F}_{m\times m}$ и $B \in \mathscr{F}_{n\times n}$, то*

(i) $A \otimes B = (A \otimes I_n)(I_m \otimes B)$;

(ii) $\det(A \otimes B) = (\det A)^n(\det B)^m$;

(iii) *если дополнительно A и B неособые, то*

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1};$$

(iv) *если $A_1, A_2, \ldots, A_k \in \mathscr{F}_{m\times m}$ и $B_1, B_2, \ldots, B_k \in \mathscr{F}_{n\times n}$, то*

$$(A_1 \otimes B_1)(A_2 \otimes B_2)\ldots(A_k \otimes B_k) = (A_1A_2\ldots A_k) \otimes (B_1B_2\ldots B_k).$$

Доказательство. Часть (i) — непосредственное применение теоремы. Для части (ii) заметим, что

$$I_m \otimes B = \operatorname{diag}\{B, B, \ldots, B\},$$

так что $\det(I_m \otimes B) = (\det B)^m$.

Затем нетрудно убедиться в существовании перестановочной матрицы P, для которой $P'(A \otimes I_n)P = I_n \otimes A$, а так как $\det P = 1$, то $\det(A \otimes I_n) = \det(I_n \otimes A) = (\det A)^n$. Результат получается теперь взятием определителей в части (i).

Для части (iii) из теоремы получаем

$$(A \otimes B)(A^{-1} \otimes B^{-1}) = I_m \otimes I_n = I_{mn},$$

откуда и следует результат.

Часть (iv) получается применением обычного доказательства по индукции.

Упр. 1. Доказать часть (i) следствия непосредственно, используя лишь определение прямого произведения.

Упр. 2. Доказать, что $\operatorname{tr}(A \otimes B) = (\operatorname{tr} A)(\operatorname{tr} B)$.

Упр. 3. Положим $A^{[1]} = A$ и определим $A^{[k+1]} = A \otimes A^{[k]}$, $k = 1, 2, \ldots$ Доказать, что

(i) $A^{[k+l]} = A^{[k]} \otimes A^{[l]}$;

(ii) если $A, B \in \mathscr{F}_{n \times n}$, то $(AB)^{[k]} = A^{[k]} B^{[k]}$. (Отметить противоположность с $(AB)^k$.)

Упр. 4. Если $r(A)$ обозначает ранг матрицы A, то показать, что

$$r(A \otimes B) = r(A)\, r(B).$$

[Указание. Воспользоваться теоремой 4.8.1.]

8.3. Собственные значения составных матриц

Рассмотрим некоторый многочлен φ от двух переменных x и y с комплексными коэффициентами. Таким образом,

$$\varphi(x, y) = \sum_{i, j=0}^{p} c_{ij} x^i y^j$$

для комплексных чисел c_{ij} и целого числа p. Если $A \in \mathscr{C}_{m \times m}$ и $B \in \mathscr{C}_{n \times n}$, то мы рассматриваем матрицы в $\mathscr{C}_{mn \times mn}$ вида

$$\varphi(A; B) = \sum_{i, j=0}^{p} c_{ij} A^i \otimes B^j.$$

Например, если $\varphi(x, y) = 2x + xy^3$, то $2x + xy^3 = 2x^1 y^0 + x^1 y^3$ и

$$\varphi(A; B) = 2A \otimes I_n + A \otimes B^3.$$

Одной из главных причин интереса к прямому произведению является следующая изящно простая связь между собственными значениями матриц A, B и $\varphi(A; B)$.

Теорема 8.3.1. *Если $\lambda_1, \ldots, \lambda_m$ — собственные значения для $A \in \mathscr{C}_{m \times m}$ и $\mu_1, \ldots, \mu_n$ — собственные значения для $B \in \mathscr{C}_{n \times n}$, то собственными значениями матрицы $\varphi(A; B)$ будут mn чисел $\varphi(\lambda_r, \mu_s)$, где $r = 1, 2, \ldots, m$ и $s = 1, 2, \ldots, n$.*

Доказательство. 1. Определим матрицы $P \in \mathscr{C}_{m \times m}$ и $Q \in \mathscr{C}_{n \times n}$ так, что

$$PAP^{-1} = J_1, \quad QBQ^{-1} = J_2$$

и J_1, J_2 — матрицы в жордановой нормальной форме (§ 4.13). Очевидно, что J_1^i — верхняя треугольная матрица с элементами

$\lambda_1^i, \ldots, \lambda_m^i$ на главной диагонали и подобно этому $\mu_1^j, \ldots, \mu_m^j$ — элементы на главной диагонали верхней треугольной матрицы J_2^j.

Кроме того, из определения прямого произведения следует, что $J_1^i \otimes J_2^j$ — тоже верхняя треугольная матрица, имеющая на главной диагонали элементы $\lambda_r^i \mu_s^j$ для $r = 1, 2, \ldots, m$; $s = 1, 2, \ldots, n$. Отсюда получаем, что матрица $\varphi(J_1; J_2)$ верхняя треугольная и имеет на главной диагонали элементы $\varphi(\lambda_r, \mu_s)$. Но элементы на главной диагонали верхней треугольной матрицы являются ее собственными значениями. Следовательно, $\varphi(J_1; J_2)$ имеет собственные значения $\varphi(\lambda_r, \mu_s)$. Теперь нам остается лишь показать, что $\varphi(J_1; J_2)$ и $\varphi(A; B)$ имеют одинаковые собственные значения.

2. Применяя часть (iv) следствия теоремы 8.2.1, находим, что

$$J_1^i \otimes J_2^j = PA^iP^{-1} \otimes QB^jQ^{-1} = (P \otimes Q)(A^i \otimes B^j)(P^{-1} \otimes Q^{-1}),$$

а по части (iii) того же следствия $P^{-1} \otimes Q^{-1} = (P \otimes Q)^{-1}$. Таким образом,

$$J_1^i \otimes J_2^j = (P \otimes Q)(A^i \otimes B^j)(P \otimes Q)^{-1}$$

и, следовательно,

$$\varphi(J_1;\ J_2) = (P \otimes Q)\, \varphi(A;\ B)\, (P \otimes Q)^{-1}.$$

Это показывает, что $\varphi(J_1; J_2)$ и $\varphi(A; B)$ подобны, так что они должны иметь одинаковые собственные значения (теорема 2.4.1). ◀

Следующие два частных случая, вероятно, наиболее часто используются. Если $\varphi(x, y) = xy$, то теорема означает, что *собственные значения матрицы* $A \otimes B$ *совпадают с* mn *числами* $\lambda_r\mu_s$, *где* $r = 1, 2, \ldots, m$; $s = 1, 2, \ldots, n$. Если же $\varphi(x, y) = x + y$, то получаем, что *собственные значения матрицы* $A \otimes I_n + I_m \otimes B$ *совпадают с* mn *числами* $\lambda_r + \mu_s$. Матрица $A \otimes I_n + I_m \otimes B$ часто называется *кронекеровской суммой* матриц A и B.

Упр. 1. Проверить теорему 8.3.1 в случаях $\varphi(x, y) \equiv 1$; $\varphi(x, y) \equiv x$.

Упр. 2. Воспользовавшись примененным при доказательстве теоремы 6.1.3 приемом, показать, что если $A\boldsymbol{x} = \lambda\boldsymbol{x}$ и $B\boldsymbol{y} = \mu\boldsymbol{y}$, то $(A \otimes B)\boldsymbol{z} = \lambda\mu\boldsymbol{z}$, где

$$\boldsymbol{z}' = \| x_1\boldsymbol{y}' x_2\boldsymbol{y}' \ \ldots \ x_m\boldsymbol{y}' \|.$$

Упр. 3. Исследовать собственные векторы матрицы $\varphi(A; B)$, если дано, что A — простая матрица.

Упр. 4. Доказать, что $A \otimes I_n$ и $I_m \otimes B$ коммутируют.

Упр. 5. Если $C = A \otimes I_n + I_m \otimes B$, то доказать, что $e^C = e^A \otimes e^B$.

Упр. 6. Пусть $A \in \mathscr{C}_{n\times n}$ и $B \in \mathscr{C}_{mn\times mn}$ — блочная матрица, каждый из m^2 блоков которой равен A. Если A имеет собственные значения $\mu_1, \ldots, \mu_n$ (не обязательно различные), то доказать, что собственные значения матрицы B равны $m\mu_1, \ldots, m\mu_n$ и 0 в количестве $n(m-1)$.

8.4. Решение линейных матричных уравнений

Рассмотрим общее линейное матричное уравнение для неизвестной матрицы $X \in \mathscr{C}_{n\times n}$:

$$A_1XB_1 + A_2XB_2 + \ldots + A_kXB_k = C, \tag{8.4.1}$$

где $A_1, \ldots, A_k, B_1, \ldots, B_k$ и C — элементы из $\mathscr{C}_{n\times n}$. Временно рассмотрим уравнение лишь с одним членом слева: $AXB = C$. Это уравнение можно рассматривать как сокращенную запись для n^2 скалярных уравнений с n^2 элементами из X. Таким образом, если определить элементы из $\mathscr{C}_{n^2}$ с помощью *)

$$\boldsymbol{x} = \begin{Vmatrix} X'_{1*} \\ X'_{2*} \\ \vdots \\ X'_{n*} \end{Vmatrix}, \quad \boldsymbol{c} = \begin{Vmatrix} C'_{1*} \\ C'_{2*} \\ \vdots \\ C'_{n*} \end{Vmatrix}, \tag{8.4.2}$$

то уравнение $AXB = C$ будет эквивалентно уравнению $G\boldsymbol{x} = \boldsymbol{c}$ для некоторой $G \in \mathscr{C}_{n^2\times n^2}$. Выделяя элемент r, s из AXB, получаем

$$A_{r*}XB_{*s} = \sum_{i,\,j} a_{ri}x_{ij}b_{js} = c_{rs}.$$

Таким образом, коэффициент при x_{ij} в $[(r-1)n+s]$-м уравнении равен $a_{ri}b_{js}$, т. е. матрица G задается элементами $a_{ri}b_{js}$ в строке $(r-1)n+s$ и столбце $(i-1)n+j$, где $1 \leqslant r, s, i, j \leqslant n$.

Сравнивая эти элементы с элементами произведения $A \otimes B$, вид которых выяснен в § 8.2, убеждаемся в том, что $G = A \otimes B'$. Возвращаясь к уравнению (8.4.1), видим, что оно эквивалентно уравнению $G\boldsymbol{x} = \boldsymbol{c}$, где

$$G = A_1 \otimes B'_1 + A_2 \otimes B'_2 + \ldots + A_k \otimes B'_k. \tag{8.4.3}$$

Теорема 8.4.1. *Считая, что* $\boldsymbol{x}$ *и* $\boldsymbol{c}$ *определяются формулой* (8.4.2), *будем иметь, что множество решений уравнения* (8.4.1)

*) См. § 1.15 для употребляемых здесь обозначений.

совпадает с множеством решений уравнения $G\mathbf{x} = \mathbf{c}$, *где* G *задается формулой* (8.4.3).

Преобразовав линейное матричное уравнение к уравнению вида $G\mathbf{x} = \mathbf{c}$, мы можем применить теперь результаты из § 1.16 к задаче получения условия для существования и единственности решения. В явном виде оно нам не понадобится.

Обратимся теперь к некоторым частным случаям. Самая простая нетривиальная возможность — это уравнение $AX = C$. Но это уравнение эквивалентно n уравнениям $AX_{*j} = C_{*j}$ для $j = 1, 2, \ldots, n$, что охватывается анализом гл. 1. Таким образом, если A неособая, то $AX = C$ имеет единственное решение $X = A^{-1}C$, и решение существует в том и только в том случае, когда матрица $\| AC \|$ имеет ранг матрицы A.

Упр. 1. Доказать, что уравнение $AX - XA = \lambda X$ имеет нетривиальное решение в том и только в том случае, когда $\lambda = \mu_r - \mu_s$, где $\mu_1, \ldots, \mu_n$ — собственные значения для A.

8.5. Уравнение $AX + XB = C$

Применяя теорему 8.4.1 к уравнению $AX + XB = C$, получаем эквивалентное уравнение $G\mathbf{x} = \mathbf{c}$, где $G = A \otimes I + I \otimes B'$. Согласно теореме 1.16.5 уравнение для X имеет единственное решение в том и только в том случае, когда G неособая. Мы знаем также, что G неособая в том и только в том случае, когда ее собственные значения ненулевые (упр. 1 из § 2.1). Но, как мы видели в § 8.3, собственные значения матрицы $A \otimes I + I \otimes B$ в точности совпадают с n^2 числами $\lambda_r + \mu_s$, где $\lambda_1, \ldots, \lambda_n$ — собственные значения матрицы A и $\mu_1, \ldots, \mu_n$ — собственные значения матрицы B. Таким образом, получена

Теорема 8.5.1. *Уравнение* $AX + XB = C$ *имеет единственное решение* $\Longleftrightarrow$ A *и* $-B$ *не имеют общих собственных значений.*

Существует частный случай, в котором можно получить точный вид для матрицы-решения X. Предположим, что A и $-B$ не имеют общих собственных значений, и рассмотрим матрицу $Z(t)$, определяемую как решение задачи с начальным значением:

$$\frac{dZ}{dt} = AZ + ZB, \quad Z(0) = C.$$

Как указывалось в упр. 7 из § 5.9, решением этого матричного дифференциального уравнения будет

$$Z(t) = e^{At}Ce^{Bt}.$$

Предположим теперь, что $\lim\limits_{t \to \infty} e^{At}Ce^{Bt} = 0$ и что матрица $X = -\int\limits_0^\infty e^{At}Ce^{Bt}\,dt$ существует. Например, зная строение матриц e^{At} и e^{Bt} (§ 5.9), легко видеть, что эти условия выполняются, если вещественные части собственных значений для A и B все отрицательны. При этих предположениях мы можем проинтегрировать дифференциальное уравнение от $t=0$ до ∞ и получить $C=AX+XB$. Таким образом, мы имеем решение уравнения $AX+XB=C$ вида

$$X=-\int_0^\infty e^{At}Ce^{Bt}\,dt. \tag{8.5.1}$$

Ясно также, что если A и B имеют собственные значения, вещественные части которых все отрицательны, то A и $-B$ не имеют общих собственных значений, так что, согласно теореме 8.5.1, существует лишь одно решение. Таким образом, имеем:

Теорема 8.5.2. *Если все собственные значения матриц A, $B\in\mathscr{C}_{n\times n}$ имеют отрицательные вещественные части, то единственное решение X уравнения $AX+XB=C$ задается по формуле* (8.5.1).

Рассмотрим теперь случай, когда A и B имеют общие собственные значения. Тогда $G=A\otimes I+I\otimes B'$ особая и решения будут существовать при условии, что пополненная матрица $\|Gc\|$ имеет ранг матрицы G (теорема 1.16.3). Число линейно независимых решений определяется тогда размерностью d нулевого подпространства для G. Определение числа d — относительно трудный вопрос, и мы не будем этим заниматься. Теорему в полном объеме и доказательство можно найти в образцовой обзорной работе Макдаффи *), где приведено доказательство Фробениуса; это доказательство датируется 1910 г.

***Упр. 1.** Если $A=PJP^{-1}$ и $B=Q\hat{J}Q^{-1}$, где J, $\hat{J}$ — матрицы в жордановой нормальной форме, то доказать, что уравнения $AX+XB=0$ (для X) и $JY+Y\hat{J}=0$ (для Y) эквивалентны, если положить $Y=P^{-1}XQ$.

Упр. 2. Предположим (в теореме 8.5.1), что A простая и имеет некоторое собственное значение λ_i кратности α_i. Пусть μ_j — собственное значение матрицы B, для которого $\lambda_i+\mu_j=0$, и пусть μ_j имеет кратность β_j, геометрическую кратность γ_j и индекс δ_j. Если k — число линейно независимых решений уравнения $AX+XB=0$, то показать, что $k\geqslant\gamma_j$ при $\alpha_i\geqslant\beta_j$ и $k\geqslant 2$ при $\alpha_i>\delta_j$.

Упр. 3. Доказать, что уравнение $AX+XB=0$ имеет неособое решение $\Longleftrightarrow$ $I\lambda-A$ и $I\lambda+B$ имеют одинаковые инвариантные многочлены.

---

*) См. дополнение 3.

8.6. Коммутирующие матрицы

Важным частным случаем теоремы 8.5.1 является однородная задача ($C=0$), в которой $B=-A$. Таким образом, для данной матрицы A мы должны найти матрицы X, которые удовлетворяют равенству $AX=XA$, т. е. матрицы, которые коммутируют с A. Как мы видели в упр. 1 выше, если $A=PJP^{-1}$, где J — жорданова нормальная форма для A, то $AX=XA$ эквивалентно $JY=YJ$, где $Y=P^{-1}XP$.

Анализ этого редуцированного уравнения все еще остается громоздким, и мы ограничимся рассмотрением важного частного случая, в котором A простая. Предположим, что $A\in\mathscr{C}_{n\times n}$ и имеет собственные значения $\mu_1, \ldots, \mu_n$. Тогда $J=\operatorname{diag}\{\mu_1, \ldots, \mu_n\}$ и $JY=YJ$ дают

$$\mu_j y_{jk}=\mu_k y_{jk} \quad \text{для} \quad j,\ k=1,\ 2,\ \ldots,\ n.$$

При $\mu_j\neq\mu_k$ это означает, что $y_{jk}=0$, а при $\mu_j=\mu_k$ элемент y_{jk} может принимать любое значение. Без потери общности можно предположить, что существует s различных значений $\mu_1, \ldots, \mu_s$ с кратностями $\alpha_1, \ldots, \alpha_s$ и

$$J=\operatorname{diag}\{\underbrace{\mu_1, \ldots, \mu_1}_{\alpha_1 \text{ раз}},\ \mu_2, \ldots,\ \underbrace{\mu_s, \ldots, \mu_s}_{\alpha_s \text{ раз}}\}.$$

Тогда ясно, что Y имеет квазидиагональный вид

$$Y=\operatorname{diag}\{Y_1, \ldots, Y_s\},$$

где $Y_j\in\mathscr{C}_{\alpha_j\times\alpha_j}$ для $j=1,\ 2,\ \ldots,\ s$. Существует поэтому $\alpha_1^2+\alpha_2^2+\ \ldots\ +\alpha_s^2$ неопределенных элементов в Y, а так как $X=PYP^{-1}$, то *при простой матрице A линейное пространство решений уравнения $AX=XA$ имеет размерность* $\alpha_1^2+\ \ldots\ +\alpha_s^2$.

Упр. 1. Доказать последнее утверждение, рассмотрев уравнение $G\boldsymbol{y}=\boldsymbol{0}$, где $G=J\otimes I-I\otimes J$.

Упр. 2. Показать, что если A имеет различные собственные значения, то каждая коммутирующая с A матрица будет простой.

Упр. 3. Если функция f определена на спектре матрицы (простой или нет), то доказать, что $f(A)$ и A коммутируют.

Предположим теперь, что каждая матрица Y_j приводится к жордановой нормальной форме следующим образом: $Y_j=Q_jJ_jQ_j^{-1}$. Если положить $Q=\operatorname{diag}\{Q_1, \ldots, Q_s\}$, то, очевидно,

$$Y=Q\operatorname{diag}\{J_1, \ldots, J_s\}Q^{-1}$$

и, следовательно,

$$X=(PQ)\operatorname{diag}\{J_1, \ldots, J_s\}(PQ)^{-1}. \tag{8.6.1}$$

Так как $\operatorname{diag}\{J_1, \ldots, J_s\}$ имеет жорданову нормальную форму, это равенство дает некоторое представление о том, как устроены матрицы, коммутирующие с простой матрицей A. Произведем дальнейшую специализацию и рассмотрим простые матрицы, коммутирующие с A.

Теорема 8.6.1. *Простые матрицы A и X из $\mathscr{C}_{n\times n}$ коммутируют $\Longleftrightarrow$ они имеют общее множество из n линейно независимых правых собственных векторов.*

Доказательство. Так как A простая, мы знаем, что если A и X коммутируют, то X приводится к жорданову нормальному виду с помощью PQ, как отмечено в (8.6.1). Так как X тоже простая, то $\operatorname{diag}\{J_1, \ldots, J_s\}$ — диагональная матрица, так что столбцы в PQ образуют n линейно независимых правых собственных векторов для X. Мы должны лишь показать, что то же самое верно для A. Но это так и есть, ибо

$$(PQ)\,J\,(PQ)^{-1} = P \operatorname{diag}\left\{\mu_1 Q_1 Q_1^{-1}, \ \ldots, \ \mu_s Q_s Q_s^{-1}\right\} P^{-1} = PJP^{-1} = A.$$

Наоборот, если A и X имеют общее множество из n линейно независимых правых собственных векторов, то A и X, очевидно, простые и собственные векторы определяют матрицу P, для которой $A = PUP^{-1}$, $X = P\Lambda P^{-1}$, где U, Λ — диагональные матрицы. Тогда A и X коммутируют, ибо

$$AX = PU\Lambda P^{-1} = P\Lambda UP^{-1} = XA. \quad ◀$$

Упр. 4. Если A — нормальная матрица с различными собственными значениями, то доказать, что каждая матрица, которая коммутирует с A, нормальна.

Упр. 5. Доказать, что нормальные матрицы коммутируют $\Longleftrightarrow$ они имеют общее множество ортонормированных собственных векторов.

Упр. 6. Если $JY = YJ$ и $J = \lambda I_n + H_n$ (§ 4.13), то доказать, что

$$Y = \begin{Vmatrix} y_1 & y_2 & y_3 & \cdots & y_n \\ 0 & y_1 & y_2 & \cdots & y_{n-1} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \vdots & & & y_1 & y_2 \\ 0 & \cdots & & 0 & y_1 \end{Vmatrix}.$$

Упр. 7. Если A и X — простые коммутирующие между собой матрицы, то доказать, что не обязательно существует многочлен p, для которого

$$X = p(A).$$

Упр. 8. Если A — простая матрица с s различными собственными значениями, то доказать, что каждая матрица, коммутирующая с A, имеет по крайней мере s линейно независимых правых собственных векторов, общих для нее и A.

Мы докажем теперь заключительный результат, характеризующий матрицы коммутирующие с некоторой простой матрицей.

Теорема 8.6.2. *Пусть A — простая матрица с различными собственными значениями $\mu_1, \ldots, \mu_s$ и компонентными матрицами $Z_1, \ldots, Z_s$ соответственно. Матрица X коммутирует с $A \Longleftrightarrow X$ коммутирует с каждой компонентной матрицей для A, т. е. в том и только в том случае, когда $XZ_j = Z_jX$, $j = 1, 2, \ldots, s$.*

Доказательство. Предположим сначала, что $XZ_j = Z_jX$ для $j = 1, 2, \ldots, s$. Тогда

$$AX = \left(\sum_{j=1}^{s} \mu_j Z_j\right) X = X \left(\sum_{j=1}^{s} \mu_j Z_j\right) = XA.$$

Наоборот, заметим, что компонентные матрицы для A — это многочлены от A (§ 5.4), а поэтому $AX = XA$ означает, что $Z_jX = XZ_j$ для каждого j. ◀

Заметим, что последнее утверждение верно для произвольной матрицы A. Таким образом, если матрица $A \in \mathscr{C}_{n\times n}$ имеет компонентные матрицы Z_{kj} (в обозначениях § 5.4) и $AX = XA$, то $Z_{kj}X = XZ_{kj}$.

Упр. 9. Доказать последний результат (что из $AX = XA$ следует $Z_{kj}X = XZ_{kj}$), показав, что из $AX = XA$ следует

$$X(I\lambda - A)^{-1} = (I\lambda - A)^{-1} X$$

для λ, не являющегося собственным значением матрицы A, и затем воспользовавшись (5.8.2).

Упр. 10. Матрица $R \in \mathscr{C}_{n\times n}$ называется *циркулянтом*, если она имеет вид

$$R = \begin{Vmatrix} a_1 & a_2 & a_3 & \ldots & a_{n-1} & a_n \\ a_n & a_1 & a_2 & \ldots & a_{n-2} & a_{n-1} \\ a_{n-1} & a_n & a_1 & \ldots & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \cdot & & a_1 & a_2 \\ a_2 & a_3 & a_4 & \ldots & a_n & a_1 \end{Vmatrix},$$

где $a_1, \ldots, a_n$ — комплексные числа. Доказать, что множество всех циркулянтов образует n-мерное подпространство в $\mathscr{C}_{n\times n}$.

Упр. 11. Доказать, что если (в упр. 10) $a_1 = \ldots = a_{n-1} = 0$ и $a_n = 1$, то получающийся в результате циркулянт U унитарен,

$U^n = I$ и общая матрица R, являющаяся циркулянтом, может быть записана в виде

$$R = a_1 I + a_n U + a_{n-1} U^2 + \ldots + a_2 U^{n-1}.$$

Упр. 12. Воспользоваться упражнениями 4 и 11 для доказательства того, что матрица, являющаяся циркулянтом, нормальна.

Упр. 13. Доказать, что если R_1, R_2 — циркулянты, то $R_1 R_2 = R_2 R_1$ и $R_1 R_2$ — циркулянт.

Упр. 14. Пусть $\omega_1, \ldots, \omega_n$ — различные корни уравнения $z^n = 1$. Если многочлен p определен как

$$p(\lambda) = a_1 + a_n \lambda + a_{n-1} \lambda^2 + \ldots + a_2 \lambda^{n-1}$$

(ср. упр. 11), то доказать, что собственные значения циркулянта R равны $p(\omega_j)$, $j = 1, 2, \ldots, n$, и что соответствующими правыми собственными векторами будут

$$\boldsymbol{x}_j = \omega_j^{n-1} \boldsymbol{e}_1 + \omega_j^{n-2} \boldsymbol{e}_2 + \ldots + \boldsymbol{e}_n.$$

[Указание. Заметим, что $U\boldsymbol{e}_n = \boldsymbol{e}_1$, $U\boldsymbol{e}_j = \boldsymbol{e}_{j+1}$, $j = 1, 2, \ldots, n-1$.]

8.7. Теория устойчивости Ляпунова

Во введении к этой главе была определена *устойчивая* матрица. Матрицы такого типа имеют особое значение при изучении дифференциальных уравнений, что наглядно показывает следующее утверждение: *матрица A будет устойчивой в том и только в том случае, когда для каждого вектора-решения $\boldsymbol{x}(t)$ уравнения $\dot{\boldsymbol{x}} = A\boldsymbol{x}$ имеем* $\lim\limits_{t \to \infty} \boldsymbol{x}(t) = 0$. Решения дифференциального уравнения $\dot{\boldsymbol{x}} = A\boldsymbol{x}$ были подробно изучены в § 5.9. Результат становится очевидным, если эти решения рассматриваются с привлечением определения устойчивой матрицы.

В дальнейшем мы будем иметь дело с квадратичными формами. Мы будем обозначать через v квадратичную форму, матрица которой (вещественная и симметрическая, § 2.12) есть V, и подобно этому для w и W. Мы выражаем тот факт, что V (или W) положительно определенная, записью $v > 0$ (или $w > 0$) и подобно этому для неотрицательно определенных форм.

Рассмотрим некоторую квадратичную форму v, вычисленную в решении уравнения $\dot{\boldsymbol{x}} = A\boldsymbol{x}$, где $A \in \mathscr{R}_{n \times n}$. Имеем $v(\boldsymbol{x}) = \boldsymbol{x}' V \boldsymbol{x}$ и, следовательно,

$$\dot{v}(\boldsymbol{x}) = \dot{\boldsymbol{x}}' V \boldsymbol{x} + \boldsymbol{x}' V \dot{\boldsymbol{x}} = \boldsymbol{x}' A' V \boldsymbol{x} + \boldsymbol{x}' V A \boldsymbol{x} = \boldsymbol{x}' (A'V + VA) \boldsymbol{x}.$$

Если положить

$$A'V + VA = -W, \tag{8.7.1}$$

то, очевидно, W будет вещественной и симметрической и $\dot{v}(\boldsymbol{x}) = -w(\boldsymbol{x})$.

Рассмотрим теперь (8.7.1) как уравнение для V. Согласно теореме 8.5.1 при заданной матрице W единственное решение V этого уравнения существует в том и только в том случае, когда A и $-A'$ не имеют общих собственных значений. Так как собственные значения для $-A'$ — это взятые со знаком минус собственные значения для A, то справедлива

Лемма. *Уравнение* (8.7.1) *определяет взаимно однозначное соответствие между вещественными симметрическими матрицами V и W в том и только в том случае, когда A не имеет нулевых собственных значений и пар собственных значений* μ, λ, *для которых* $\mu = -\lambda$.

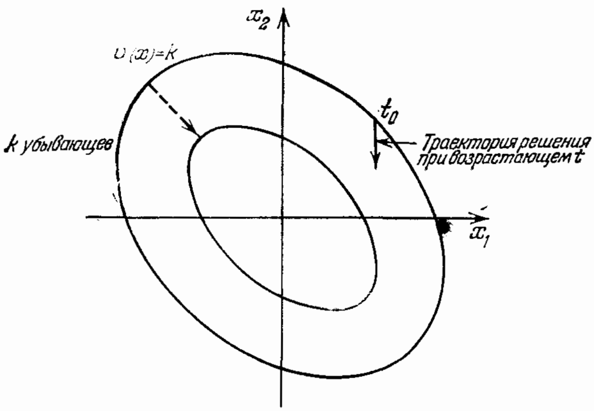

Рис. 5. Траектория для решения в случае, когда A устойчивая.

Вклад Ляпунова состоит в замечании, что при заданной определенной форме w устойчивость матрицы A может быть характеризована существованием *определенной* матрицы-решения V для (8.7.1). Таким образом, уравнение $\dot{\boldsymbol{x}} = A\boldsymbol{x}$ обладает свойством, что $\lim\limits_{t \to \infty} \boldsymbol{x}(t) \to \boldsymbol{0}$ для каждого вектора-решения $\boldsymbol{x}(t) \Longleftrightarrow$ можно найти такие положительно определенные формы w и v, что $\dot{v}(\boldsymbol{x}) = -w(\boldsymbol{x})$.

Чтобы убедиться в значении этого условия наглядно, предположим, что A имеет размеры 2×2, и рассмотрим решение

$$\boldsymbol{x}(t) = \begin{Vmatrix} x_1(t) \\ x_2(t) \end{Vmatrix},$$

в котором x_1 и x_2 — вещественнозначные функции от t. Кривыми уровня для положительно определенной формы v будут эллипсы в плоскости x_1, x_2 (рис. 5). Если $\boldsymbol{x}_0 = \boldsymbol{x}(t_0)$, то на эллипсе $v(\boldsymbol{x}) = v(\boldsymbol{x}_0)$ выбирается точка, соответствующая $\boldsymbol{x}(t_0)$. Рассмотрим теперь непрерывный путь (траекторию) решения $\boldsymbol{x}(t)$ в окрестности точки $(x_1(t_0), x_2(t_0))$. Если $w > 0$, то $\dot{v} < 0$ и путь должен входить *внутрь* эллипса $v(\boldsymbol{x}) = v(\boldsymbol{x}_0)$ при возрастающем t и, кроме того, он должен подойти на сколь угодно близкое расстояние к началу координат при достаточно большом t.

Возвращаясь к более точному изложению, мы сформулируем сейчас алгебраическую версию теоремы Ляпунова, которая представляется относительно более легкой для понимания. Более знакомый частный случай получается затем в виде теоремы 8.7.2.

Теорема 8.7.1. (i). *Если матрица $A \in \mathcal{R}_{n\times n}$ устойчивая и $W \in \mathcal{R}_{n\times n}$ положительно (неотрицательно) определенная, то существует такая вещественная положительно (положительно или неотрицательно) определенная матрица V, что $A'V + VA = -W$.*

(ii) *Пусть матрица $V \in \mathcal{R}_{n\times n}$ положительно определенная и $A \in \mathcal{R}_{n\times n}$. Определим вещественную симметрическую матрицу W равенством $A'V + VA = -W$. Если каждому отличному от других собственному значению матрицы A соответствует некоторый правый собственный вектор $\boldsymbol{a}$, для которого $\boldsymbol{a}^*W\boldsymbol{a} > 0$, то A будет устойчивой.*

Доказательство. (i) Так как A устойчива, из леммы получаем, что существует единственная матрица $V \in \mathcal{R}_{n\times n}$, для которой $A'V + VA = -W$. Кроме того, можно воспользоваться теоремой 8.5.2 и записать решение в явном виде

$$V = \int_0^\infty e^{A't} W e^{At}\, dt.$$

Следовательно,

$$v(\boldsymbol{x}) = \langle \boldsymbol{x},\ V\boldsymbol{x} \rangle = \int_0^\infty \langle e^{At}\boldsymbol{x},\ We^{At}\boldsymbol{x} \rangle\, dt.$$

Отсюда сразу же получаем, что если $w \geqslant 0$, то $v \geqslant 0$ (и возможно $v > 0$).

Но, как известно, e^{At} неособая, так что $e^{At}\boldsymbol{x} \neq \boldsymbol{0}$ для любого ненулевого $\boldsymbol{x}$ и $w > 0$ означают, что подынтегральное выражение положительно при каждом t. Таким образом, $w > 0$ означает, что $v > 0$.

(ii) Пусть μ — собственное значение матрицы A и $A\boldsymbol{a} = \mu\boldsymbol{a}$ с $\boldsymbol{a} \neq \boldsymbol{0}$. Тогда имеем также $\boldsymbol{a}^*A' = \bar{\mu}\boldsymbol{a}^*$. Умножая равенство $A'V + VA = -W$ слева и справа на $\boldsymbol{a}^*$ и $\boldsymbol{a}$ соответственно, получаем

$$\bar{\mu}\boldsymbol{a}^*V\boldsymbol{a} + \mu\boldsymbol{a}^*V\boldsymbol{a} = -\boldsymbol{a}^*W\boldsymbol{a}.$$

Таким образом,

$$2(\boldsymbol{a}^*V\boldsymbol{a})\mathcal{R}(\mu) = -\boldsymbol{a}^*W\boldsymbol{a}.$$

Согласно предположению $\boldsymbol{a}^*V\boldsymbol{a} > 0$, а также $\boldsymbol{a}^*W\boldsymbol{a} > 0$, так что $\mathrm{Re}(\mu) < 0$. Так как это верно для каждого собственного значения μ, то A будет устойчивой. ◀

Следующий результат совершенно тривиально следует из теоремы 8.7.1, но он более известен и заслуживает отдельной формулировки.

Теорема 8.7.2. *Пусть A, $W \in \mathcal{R}_{n\times n}$ и W положительно определенная. Тогда A будет устойчивой в том и только в том случае, когда уравнение $A'V + VA = -W$ имеет решением положительно определенную матрицу V.*

Доказательство. Если A устойчивая и $w > 0$, то выполняются предположения первой части теоремы 8.7.1. Если, наоборот, заданы положительно определенные V и W, то выполняются предположения части (ii) этой теоремы, откуда и следует результат.

Упр. 1. Обобщить теоремы 8.7.1 и 8.7.2 на комплексные матрицы A, рассматривая *эрмитовы* формы v и w.

При практических вычислениях выяснить вопрос о том, является ли некоторая заданная вещественная матрица A устойчивой или нет, можно при помощи решения уравнения $A'V + VA = -I$ (или $A'V + VA = -W$ для некоторого другого выбора w) относительно V и применения затем теоремы 2.14.4, чтобы узнать, будет ли V положительно определенной или нет. Таким образом, A будет устойчивой в том и только в том случае, когда следующие миноры матрицы V положительны:

$$V\begin{pmatrix}1\\1\end{pmatrix},\quad V\begin{pmatrix}1,\ 2\\1,\ 2\end{pmatrix},\ \ldots,\quad V\begin{pmatrix}1,\ 2,\ \ldots,\ n\\1,\ 2,\ \ldots,\ n\end{pmatrix}.$$

Если бы уравнение $A'V + VA = -I$ можно было решить таким образом, что элементы V явно выразились бы через элементы A, то мы получили бы прямую характеризацию устойчивых матриц через определители. Предлагаемый способ не является практическим предложением для матриц общего вида, но для относительно простых классов матриц явные критерии устойчивости могут быть получены таким образом. В следующем параграфе мы воспользуемся этой возможностью в одном особенно важном случае.

8.8. Критерий Рауса — Гурвица

Вопрос об устойчивости собственных значений матрицы всегда может быть превращен в вопрос о корнях некоторого многочлена. Для этого следует лишь заменить уравнение $A\boldsymbol{x} =$

$= \mu x$ для собственного значения на характеристическое уравнение $\det(\mu I - A) = 0$. Преимущество теоремы Ляпунова из предыдущего параграфа состоит в том, что удается избежать вычисления коэффициентов этого многочлена и какой-либо последующей потери точности. Тем не менее важно то, что мы получим необходимые и достаточные условия для того, чтобы все корни некоторого многочлена лежали в левой половине комплексной плоскости. Рассмотрим приведенный многочлен с *вещественными* коэффициентами:

$$f(z) = z^n + a_1 z^{n-1} + a_2 z^{n-2} + \ldots + a_n.$$

Матрицы Гурвица для i определяются так:

$$H_1 = a_1, \quad H_2 = \begin{Vmatrix} a_1 & 1 \\ a_3 & a_2 \end{Vmatrix}, \quad H_3 = \begin{Vmatrix} a_1 & 1 & 0 \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{Vmatrix}, \ldots,$$

$$H_n = \begin{Vmatrix} a_1 & 1 & 0 & 0 & \ldots 0 & 0 \\ a_3 & a_2 & a_1 & 1 & & \\ a_5 & a_4 & a_3 & a_2 & \cdot & \cdot \\ \cdot & & & & \cdot & \cdot \\ \cdot & & & & \cdot & \cdot \\ \cdot & & & & & \\ a_{2n-1} & & \ldots & & a_{n+1} & a_n \end{Vmatrix},$$

где $a_k = 0$ при $k > n$. Если положить также $a_0 = 1$ и $a_k = 0$ при $k < 0$, то элемент (i, j) матрицы Гурвица для f есть просто a_{2i-j}. Знаменитый критерий Рауса — Гурвица может быть сформулирован следующим образом.

Теорема 8.8.1. *Все корни многочлена f имеют отрицательные вещественные части в том и только в том случае, когда*

$$\det H_j = d_j > 0, \qquad j = 1, 2, \ldots, n.$$

Самое удовлетворительное доказательство этой теоремы использует теорию функций комплексной переменной. Однако нашей целью в этом параграфе является доказательство этой теоремы при помощи теоремы Ляпунова в виде (8.7.1). Это доказательство далеко не прямое и манипуляции с определителями довольно громоздки. Мы поэтому приводим сокращенные доказательства для необходимой последовательности лемм. Первый из нужных результатов имеет некоторое значение сам по себе и мы проведем его доказательство более подробно.

Теорема 8.8.2 (формула Орландо). *Пусть $a_1, a_2, \ldots, a_n$ — комплексные числа и многочлен $f(z) = z^n + a_1 z^{n-1} + \ldots + a_n$ имеет корни $z_1, \ldots, z_n$. Тогда определитель Гурвица d_{n-1} для f задается в виде*

$$d_{n-1} = (-1)^{n(n-1)/2} \prod_{\substack{i,\ k=1 \\ i<k}}^{n} (z_i + z_k).$$

Доказательство. Воспользуемся индукцией относительно n. В случае $n=2$ имеем хорошо известный результат для квадратичного многочлена

$$d_{n-1}=d_1=a_1=-(z_1+z_2).$$

Это есть соответствующий частный случай теоремы.

Предположим, что утверждение верно для любого приведенного многочлена f степени n, и пусть F — приведенный многочлен степени $n+1$, определяемый при некотором h по формуле

$$F(z)=(z+h)f(z)=z^{n+1}+(a_1+ha_0)z^n+$$
$$+(a_2+ha_1)z^{n-1}+\ldots+(a_n+ha_{n-1})z+ha_n.$$

Если $z_1, \ldots, z_n$ — корни для f, то $z_1, \ldots, z_n, z_{n+1}=-h$ будут корнями для F.

Построим следующий определитель при помощи «окаймления» d_n:

$$D=\det\left\|\begin{matrix} a_1 & 1 & 0 & 0 & \ldots & 0 \\ a_3 & a_2 & a_1 & 1 & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & a_{n-3} \\ a_{2n-1} & & & & a_n & a_{n-1} \\ h^n & -h^{n-1} & & \ldots & (-1)^{n-1}h & (-1)^n \end{matrix}\right\|.$$

Заменим теперь первый столбец в D на (1-й столбец) $-a_1$ (2-й столбец) $+a_2$ (3-й столбец) $+\ldots+(-1)^n a_n$ (($n+1$)-й столбец). В результате последний элемент в этом столбце станет равен $f(h)$, а все остальные элементы будут нулевые. Тогда, как нетрудно убедиться,

$$D=(-1)^n f(h)\,d_{n-1}.$$

Возвращаясь к определению D, прибавим умноженный на h столбец $r+1$ к столбцу r при $r=1, 2, \ldots, n-1$. В результате получим

$$D=\det\left\|\begin{matrix} a_1+h & 1 & 0 & 0 & \ldots & 0 \\ a_3+a_2h & a_2+ha_1 & a_1+h & 1 & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & a_n+ha_{n-1} & a_{n-1} \\ 0 & & \ldots & & 0 & (-1)^n \end{matrix}\right\|.$$

Заметим теперь, что главный $n\times n$-минор этого определителя равен определителю Гурвица Δ_n порядка n для F. Таким образом, мы имеем также $D=(-1)^n\Delta_n$. Воспользовавшись предположением индукции, при $h = -z_{n+1}$ получаем

$$\Delta_n = f(h)\, d_{n-1} = (-1)^{n(n-1)/2} \prod_{\substack{i,\ k=1 \\ i<k}}^{n} (z_i + z_k) \prod_{j=1}^{n} (h - z_j) =$$

$$= (-1)^{n(n+1)/2} \prod_{\substack{i,\ k=1 \\ i<k}}^{n+1} (z_i + z_k). \quad \blacktriangleleft$$

Введем теперь $n \times n$-матрицу

$$V = \begin{Vmatrix} a_0a_1 & 0 & a_0a_3 & \ldots \\ 0 & -a_0a_3 + a_1a_2 & 0 & \ldots \\ a_0a_3 & 0 & a_0a_5 - a_1a_4 + a_2a_3 & \ldots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{Vmatrix}, \qquad (8.8.1)$$

элементы которой задаются следующим образом:

для $i + j$ четного

$$v_{ij} = \begin{cases} \sum_{k=1}^{i} (-1)^{k+i} a_{k-1} a_{i+j-k} & \text{при} \quad j \geqslant i, \\ v_{ji} & \text{при} \quad j < i; \end{cases}$$

для $i + j$ нечетного $v_{ij} = 0$.

Для главных миноров матрицы V пишем:

$$v_j = V \begin{pmatrix} 1 & 2 & \ldots & j \\ 1 & 2 & \ldots & j \end{pmatrix}, \qquad j = 1, 2, \ldots, n.$$

Лемма 1. *Полагая* $d_0 = 1$, *будем иметь*

$$v_j = d_j d_{j-1}, \qquad j = 1, 2, \ldots, n.$$

Доказательство. Переставим строки и столбцы матрицы V симметрично, переводя нечетно занумерованные столбцы и строки в начало матрицы и четно занумерованные строки и столбцы в ее конец. Это преобразование производится при помощи перестановочной матрицы

$$P = \| \boldsymbol{e}_1 \boldsymbol{e}_3 \boldsymbol{e}_5 \ \ldots \ \vdots \ \boldsymbol{e}_2 \boldsymbol{e}_4 \ \ldots \|.$$

Обозначим получающуюся в результате матрицу через $D_n = P'VP$. В случаях $n = 2m$, $n = 2m + 1$ соответственно D_n имеет такой блочный вид:

$$D_{2m} = \begin{Vmatrix} E_m & 0 \\ 0 & F_m \end{Vmatrix}, \quad D_{2m+1} = \begin{Vmatrix} E_{m+1} & 0 \\ 0 & F_m \end{Vmatrix},$$

где индексы обозначают порядок квадратных матриц. Имеем теперь

$$\begin{aligned} v_{2m} &= \det D_{2m} = (\det E_m)(\det F_m), \\ v_{2m+1} &= \det D_{2m+1} = (\det E_{m+1})(\det F_{m+1}). \end{aligned} \tag{8.8.2}$$

Определим $2m \times m$- и $(2m+1)\times(m+1)$-матрицы так:

$$K_{2m} = \left\| \begin{matrix} 0 & & \ldots & 0 & -a_0 \\ & & & 0 & a_1 \\ \cdot & & & -a_0 & -a_2 \\ \cdot & & & a_1 & a_3 \\ \cdot & & & & \cdot \\ 0 & & & & \cdot \\ -a_0 & & & & \cdot \\ a_1 & a_3 & \ldots & & a_{2m-1} \end{matrix} \right\|,$$

$$K_{2m+1} = \left\| \begin{matrix} 0 & & \ldots & 0 & 0 & a_0 \\ & & & 0 & 0 & -a_1 \\ & & & 0 & a_0 & a_2 \\ \cdot & & & 0 & -a_1 & -a_3 \\ \cdot & & & a_0 & a_2 & a_4 \\ \cdot & & & & & \cdot \\ & & & & & \cdot \\ 0 & & & & & \cdot \\ a_0 & a_2 & \ldots & & & a_{2m} \end{matrix} \right\|.$$

Имеем

$$H_{2m}K_{2m} = \left\| \begin{matrix} 0 \\ F_m \end{matrix} \right\|, \quad H_{2m+1}K_{2m+1} = \left\| \begin{matrix} 0 \\ E_{m+1} \end{matrix} \right\|.$$

Тогда, как легко видеть,

$$d_{2m} = \det H_{2m} = \det F_m, \quad d_{2m+1} = \det H_{2m+1} = \frac{1}{a_0} \det E_{m+1}.$$

Полагая $a_0 = 1$, из (8.8.2) теперь получаем

$$v_{2m} = d_{2m-1} d_{2m}, \quad v_{2m+1} = d_{2m+1} d_{2m}$$

и в любом случае $v_j = d_j d_{j-1}$. ◀

Из леммы 1 и теоремы 2.14.4 непосредственно следует

Лемма 2. *Если $a_1, \ldots, a_n$ вещественны, то $d_j > 0$ $(j = 1, 2, \ldots, n) \Longleftrightarrow V$ положительно определенная.*

Мы введем теперь некоторую определяемую многочленом f матрицу, которая для наших теперешних целей является более удобным видом *сопровождающей матрицы*, чем тот, который использовался в § 4.11. Мы полагаем

$$A=\begin{Vmatrix} -a_1 & -a_2 & \ldots & & -a_n \\ 1 & 0 & & & \\ 0 & 1 & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ \cdot & & & & \\ & & & 1 & 0 \end{Vmatrix}$$

и замечаем, что «A устойчивая» и «все корни f имеют отрицательные вещественные части» — эквивалентные утверждения.

Лемма 3. *При определенной выше A и V из* (8.8.1) *имеем $A'V+VA=-W$, где $\frac{1}{2}W$ — неотрицательно определенная матрица*

$$\begin{Vmatrix} a_1^2 & 0 & a_1a_3 & 0 & a_1a_5 & \ldots \\ 0 & 0 & 0 & 0 & 0 & \ldots \\ a_1a_3 & 0 & a_3^2 & 0 & a_3a_5 & \ldots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \end{Vmatrix}.$$

Доказательство. Вычисление матрицы $AV'+VA$, исходя из определений A и V, приводит к первой части леммы. Чтобы убедиться, что W неотрицательно определенная, заметим, что

$$\boldsymbol{x}^*W\boldsymbol{x}=2\,|a_1x_1+a_3x_3+a_5x_5+\ \ldots\ |^2. \tag{8.8.3}$$

Лемма 4. *Если $d_j>0$, $j=1, 2, \ldots, n$, то A устойчива.*

Доказательство. Согласно лемме 2 $d_j>0$ для $j=1, 2, \ldots, n$ означает, что V положительно определенная. Тогда, воспользовавшись леммой 3 и теоремой 8.7.1, часть (ii), мы должны лишь доказать, что $\boldsymbol{a}^*W\boldsymbol{a}>0$ по крайней мере для одного правого собственного вектора, соответствующего любому заданному собственному значению μ матрицы A.

Но, как мы знаем (ср. упр. 2 из § 4.11), в качестве собственного вектора можно взять $\boldsymbol{a}$ с $\boldsymbol{a}'=\|\,\mu^{n-1}, \ldots, \mu, 1\,\|$. Согласно (8.8.3) $\boldsymbol{a}^*W\boldsymbol{a}=0$ в том и только в том случае, когда

$$a_1\mu^{n-1}+a_3\mu^{n-3}+\ \ldots\ =0.$$

Но $d_n>0$ означает, что $a_n\neq 0$ и, следовательно, $\mu\neq 0$. Таким образом, так как $f(\mu)=0$, имеем для нечетного n

$$\mu^n+a_2\mu^{n-2}+\ \ldots\ +a_{n-1}\mu=0,$$
$$a_1\mu^{n-1}+a_3\mu^{n-3}+\ \ldots\ +a_n=0,$$

откуда

$$\left\|\begin{matrix} 1 & a_2 & a_4 & \dots & a_{n-1} \\ a_1 & a_3 & a_5 & \dots & a_n \end{matrix}\right\| \left\|\begin{matrix} \mu^{n-1} \\ \cdot \\ \cdot \\ \cdot \\ \mu^2 \\ 1 \end{matrix}\right\| = \mathbf{0},$$

и для четного n

$$\mu^n + a_2\mu^{n-2} + \dots + a_n = 0,$$
$$a_1\mu^{n-1} + a_3\mu^{n-3} + \dots + a_{n-1}\mu = 0,$$

откуда

$$\left\|\begin{matrix} 1 & a_2 & \dots & a_{n-2} & a_n \\ a_1 & a_3 & \dots & a_{n-1} & 0 \end{matrix}\right\| \left\|\begin{matrix} \mu^{n} \\ \cdot \\ \cdot \\ \cdot \\ \mu^2 \\ 1 \end{matrix}\right\| = \mathbf{0}.$$

Таким образом, для n нечетного и четного имеем

$$a_1a_2 - a_3 = 0, \quad a_1a_4 - a_5 = 0, \dots$$

и вторые строка и столбец матрицы V нулевые. Это противоречит нашему выводу о том, что матрица V положительно определенная. Поэтому $\boldsymbol{a}^*W\boldsymbol{a} \neq 0$. Следовательно, A — устойчивая матрица.

Лемма 5. *Если A — устойчивая матрица, то $d_j > 0$ для $j = 1, 2, \dots, n$.*

Доказательство. Согласно первой части теоремы 8.7.1 устойчивость A и неотрицательная определенность W означают, что $v_j \geqslant 0$, $j = 1, 2, \dots, n$. Из леммы 2 тогда следует, что $d_j \geqslant 0$, $j = 1, 2, \dots, n$, и если какое-нибудь $d_j = 0$, то $d_n = 0$. Мы покажем, что предположение $d_n = 0$ приводит к противоречию, откуда будет следовать, что $d_j > 0$ для $j = 1, 2, \dots, n$.

Устойчивость A означает, что не существует нулевых корней, так что $a_n \neq 0$. Но $d_n = a_n d_{n-1}$, так что из $d_n = 0$ следует $d_{n-1} = 0$. В этом случае из формулы Орландо получаем, что существует пара корней μ_i, μ_j, для которых $\mu_i + \mu_j = 0$. Но это противоречит предположению о том, что μ_i и μ_j оба имеют отрицательные вещественные части, что и завершает доказательство.

Критерий Рауса — Гурвица, как он сформулирован в теореме 8.8.1, получается теперь объединением результатов лемм 4 и 5.

Глава 9

НЕОТРИЦАТЕЛЬНЫЕ МАТРИЦЫ

9.1. Введение

Матрица $A \in \mathscr{R}_{n\times n}$ называется *неотрицательной*, если каждый элемент в A неотрицателен. Квадратные матрицы такого типа возникают в множестве задач и, возможно, удивит то, что это определяющее свойство приводит к некоторым сильным результатам об их строении. Замечательная теорема Перрона — Фробениуса является основным результатом для неотрицательных квадратных матриц. Мы формулируем его в полной общности в виде теорем 9.2.1 и 9.2.2, но докажем лишь первую из этих теорем. Возможно, следует предупредить читателя, что, хотя конечный результат формулируется изящно просто и мы не доказываем теорему Фробениуса полностью, доказательство тем не менее наиболее громоздкое в этой книге. Несмотря на это, приводимое доказательство содержит несколько других полезных идей и, не считая теоремы Перрона — Фробениуса, другие полезные и интересные результаты получаются более просто.

При $A, B \in \mathscr{R}_{m\times n}$ будем писать $A \geqslant B\,(>B)$, если $a_{ij} \geqslant b_{ij}$ $(>b_{ij})$ для всех $1 \leqslant i \leqslant m$ и $1 \leqslant j \leqslant n$. Это согласуется с обозначениями, введенными в § 6.4. Согласно этому обозначению A будет неотрицательной $\Longleftrightarrow$ $A \geqslant 0$. Следует заметить, что из $A \geqslant 0$ н $A \neq 0$ не следует, что $A > 0$.

Вспомним определение перестановочной матрицы из § 2.12. Если P — перестановочная матрица, то непосредственно проверяется, что $P'P = I$, т. е. перестановочная матрица обязательно ортогональная. Как было отмечено в § 2.12, если $A \in \mathscr{R}_{n\times n}$, то подобное преобразование A вида $P'AP$, где P — перестановочная матрица, приводит в результате просто к перестановке элементов A, причем строки и столбцы в A переставляются соответственно. При $n \geqslant 2$ матрица A называется *приводимой*, если существует такая перестановочная матрица P, что

$$P'AP = \begin{Vmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{Vmatrix},$$

где A_{11}, A_{22} — квадратные матрицы порядка меньше n. Если такой матрицы P не существует, то P *неприводима*. Теорема Перрона — Фробениуса относится к неприводимым неотрицательным матрицам.

Понятие приводимости имеет наиболее очевидное значение при решении уравнения $A\boldsymbol{x} = \boldsymbol{c}$ относительно $\boldsymbol{x}$. Ибо если A приводима, то мы производим замену переменных, которую подсказывает равенство $(P'AP)(P'\boldsymbol{x}) = P'\boldsymbol{c}$, и получаем

$$\left\| \begin{matrix} A_{11} & A_{12} \\ 0 & A_{22} \end{matrix} \right\| \left\| \begin{matrix} \xi_1 \\ \xi_2 \end{matrix} \right\| = \left\| \begin{matrix} \gamma_1 \\ \gamma_2 \end{matrix} \right\|,$$

где

$$P'\boldsymbol{x} = \left\| \begin{matrix} \xi_1 \\ \xi_2 \end{matrix} \right\|, \quad P'\boldsymbol{c} = \left\| \begin{matrix} \gamma_1 \\ \gamma_2 \end{matrix} \right\|.$$

Задача *сводится* теперь, во-первых, к решению системы более низкого порядка $A_{22}\xi_2 = \gamma_2$ и затем к решению уравнения

$$A_{11}\xi_1 = \gamma_1 - A_{12}\xi_2.$$

Ясно также, что собственные значения A_{11} вместе с собственными значениями A_{22} составляют собственные значения A.

Интересно, что понятие приводимости никак не связано с величиной матрицы, а зависит лишь от расположения нулевых и ненулевых элементов в этой матрице. Эта идея используется в понятии направленного графа, соответствующего матрице. Мы наметим лишь первые шаги в этой теории и получим вторую характеризацию неприводимых матриц в упр. 3. Важное обобщение теоремы Гершгорина содержится в упр. 5.

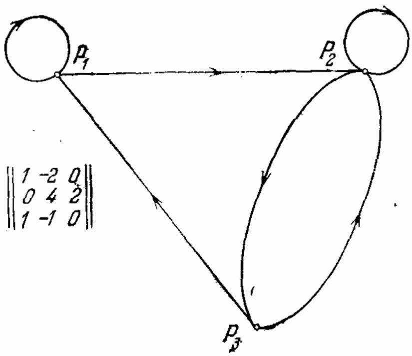

Рис. 6. Направленный граф для матрицы.

Пусть $P_1, P_2, \ldots, P_n$ — различные точки комплексной плоскости и $A \in \mathscr{C}_{n \times n}$. Для каждого ненулевого элемента a_{ij} из A соединим P_i с P_j направленной линией $\overrightarrow{P_iP_j}$. Получающаяся в результате фигура на комплексной плоскости будет *направленным графом* для A. Иллюстрация этого приведена на рис. 6 с 3×3-матрицей и ее направленным графом.

Мы будем говорить, что направленный граф *сильно связен*, если для каждой пары точек P_i, P_j существует направленный путь $\overrightarrow{P_iP_{k_1}}$, $\overrightarrow{P_{k_1}P_{k_2}}$, ..., $\overrightarrow{P_{k_{r-1}}P_j}$, соединяющий P_i с P_j.

Упр. 1. (а) Если $A \geqslant 0$, $B \geqslant C$ и AB определено, то $AB \geqslant AC$. (б) Если $A \geqslant 0$, $B > 0$ и $AB = 0$, то $A = 0$.

Упр. 2. (а) Является ли граф на рис. 6 сильно связным?

(б) Привести примеры 3×3-матриц с (и без) сильно связными графами.

Упр. 3. Доказать, что квадратная матрица неприводима $\Longleftrightarrow$ ее направленный граф сильно связен.

***Упр. 4.** Если A приводима, то доказать, что при любом положительном целом p матрица A^p будет приводимой.

Упр. 5. Пусть A — неприводимая матрица из $\mathscr{C}_{n\times n}$ и μ — собственное значение A, лежащее на окружности одного из n кругов Гершгорина (теорема 7.2.1). Доказать, что это собственное значение лежит на окружности каждого из n кругов.

Упр. 6. Если матрица A неприводима и в обозначениях теоремы 7.2.1 $|a_{jj}| \geqslant \rho_j$ для $j = 1, 2, \ldots, n$ со строгим неравенством по крайней мере при одном j, то доказать, что A неособая (ср. с упр. 3 из § 7.2).

***Упр. 7.** Описать множество всех векторов $\boldsymbol{y} \geqslant \mathbf{0}$, для которых $\boldsymbol{y} \leqslant A\boldsymbol{x}$, если

$$A = \begin{Vmatrix} 2 & 1 \\ 0 & \frac{1}{2} \end{Vmatrix} \quad \text{и} \quad \boldsymbol{x} = \begin{Vmatrix} 1 \\ 1 \end{Vmatrix}.$$

Найти наибольшее число ρ, для которого $\rho\boldsymbol{x} \leqslant A\boldsymbol{x}$.

9.2. Теорема Перрона — Фробениуса

Очевидно, если $A \geqslant 0$, то $A^p \geqslant 0$ при любом положительном целом числе p. Мы можем также ожидать, что если A имеет достаточно большую плотность ненулевых элементов, то для достаточно большого p будем иметь $A^p > 0$. Наш первый результат именно такого рода.

Лемма 1. *Если матрица $A \in \mathscr{R}_{n\times n}$ неотрицательна и неприводима, то* $(I + A)^{n-1} > 0$.

Доказательство. Рассмотрим некоторый вектор $\boldsymbol{y} \geqslant \mathbf{0}$ с $\boldsymbol{y} \neq \mathbf{0}$ и положим

$$\boldsymbol{z} = (I + A)\,\boldsymbol{y} = \boldsymbol{y} + A\boldsymbol{y}. \tag{9.2.1}$$

Так как $A \geqslant 0$, то $A\boldsymbol{y} \geqslant \mathbf{0}$ и $\boldsymbol{z}$ имеет по крайней мере столько же ненулевых (и, следовательно, положительных) элементов, сколько и $\boldsymbol{y}$. Если $\boldsymbol{y}$ еще не положительный, то мы докажем, что $\boldsymbol{z}$ будет иметь по крайней мере на один ненулевой элемент больше, чем $\boldsymbol{y}$.

Ясно, что матричные свойства, используемые в утверждении леммы, инвариантны относительно перестановочных преобразований $P'AP$. Поэтому, предполагая, что $\boldsymbol{z}$ имеет не более положительных элементов, чем $\boldsymbol{y}$, можем записать (без потери общности)

$$\boldsymbol{y} = \left\| \begin{matrix} \boldsymbol{u} \\ 0 \end{matrix} \right\|, \quad \boldsymbol{z} = \left\| \begin{matrix} \boldsymbol{v} \\ 0 \end{matrix} \right\|,$$

где $\boldsymbol{u} > \boldsymbol{0}$ и, так как $\boldsymbol{z} = \boldsymbol{y} + A\boldsymbol{y}$, $\boldsymbol{v} > \boldsymbol{0}$. Если соответственно разбить A на блоки:

$$A = \left\| \begin{matrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{matrix} \right\|,$$

то из (9.2.1) будет следовать $A_{21}\boldsymbol{u} = \boldsymbol{0}$. Так как $\boldsymbol{u} > 0$ и $A_{21} \geqslant 0$, это означает, что $A_{21} = 0$, что противоречит предположению о неприводимости A.

Повторяя это рассуждение $n - 1$ раз, получаем, что для произвольного ненулевого $\boldsymbol{y} \geqslant \boldsymbol{0}$ будет $(I + A)^{n-1}\boldsymbol{y} > \boldsymbol{0}$. Полагая $\boldsymbol{y} = \boldsymbol{e}_j$, $j = 1, 2, \ldots, n$, убеждаемся в том, что $(I + A)^{n-1} > 0$. ◀

Для ненулевой неприводимой матрицы A рассмотрим вещественнозначную функцию r, определенную на ненулевых векторах $\boldsymbol{x} \geqslant \boldsymbol{0}$ формулой

$$r(\boldsymbol{x}) = \min_{\substack{1 \leqslant i \leqslant n \\ x_i \neq 0}} \frac{(A\boldsymbol{x})_i}{x_i},$$

где $(A\boldsymbol{x})_i$ обозначает i-й элемент вектора $A\boldsymbol{x}$. Тогда $r(\boldsymbol{x}) \geqslant 0$ и при $j = 1, 2, \ldots, n$ будем иметь $r(\boldsymbol{x})x_j \leqslant (A\boldsymbol{x})_j$ с равенством при некотором j. Таким образом, $\boldsymbol{x}r(\boldsymbol{x}) \leqslant A\boldsymbol{x}$ и, кроме того, $r(\boldsymbol{x})$ — такое наибольшее число ρ, что $\rho\boldsymbol{x} \leqslant A\boldsymbol{x}$ для этого $\boldsymbol{x}$ (ср. упр. 7 из § 9.1).

Пусть $\mathscr{L}$ обозначает множество всех ненулевых неотрицательных векторов порядка n. Определим тогда

$$r = \sup_{\boldsymbol{x} \in \mathscr{L}} r(\boldsymbol{x}). \tag{9.2.2}$$

Согласно определению $r(\boldsymbol{x})$ эта функция инвариантна относительно замены $\boldsymbol{x}$ на $\alpha\boldsymbol{x}$ при любом $\alpha > 0$. Таким образом, при вычислении введенного супремума можно рассматривать лишь замкнутое множество $\mathscr{M}$ векторов $\boldsymbol{x}$, для которых $\boldsymbol{x} \geqslant \boldsymbol{0}$ и $\sum_i x_i^2 = 1$. Таким образом, $\mathscr{M} \subseteq \mathscr{L}$ и $r = \sup_{\boldsymbol{x} \in \mathscr{M}} r(\boldsymbol{x})$. Если бы функция $r(\boldsymbol{x})$ была непрерывна на $\mathscr{M}$, мы могли бы r приравнять величине $\max r(\boldsymbol{x})$, $\boldsymbol{x} \in \mathscr{M}$ (дополнение 1). Однако $r(\boldsymbol{x})$ может иметь разрывы в точках, где координаты $\boldsymbol{x}$ обращаются в нуль. Мы рассмотрим поэтому множество $\mathscr{N}$ векторов $\boldsymbol{y}$, определенных как $\boldsymbol{y} = (I + A)^{n-1}\boldsymbol{x}$, $\boldsymbol{x} \in \mathscr{M}$. Согласно лемме 1 каждый элемент из $\mathscr{N}$ будет *положительным* вектором, так что

$\mathcal{N} \subset \mathcal{L}$. Мы утверждаем, что $\mathcal{N}$ — замкнутое множество и что функция $r(\boldsymbol{y})$ непрерывна на $\mathcal{N}$.

Для любого $\boldsymbol{x} \in \mathcal{M}$

$$\boldsymbol{y} r(\boldsymbol{x}) = r(\boldsymbol{x})(I + A)^{n-1} \boldsymbol{x} \leqslant (I + A)^{n-1} A\boldsymbol{x},$$

так как $\boldsymbol{x} r(\boldsymbol{x}) \leqslant A\boldsymbol{x}$, а следовательно, $\boldsymbol{y} r(\boldsymbol{x}) \leqslant A\boldsymbol{y}$. Но $r(\boldsymbol{y})$ — наибольшее число ρ, для которого $\rho\boldsymbol{y} \leqslant A\boldsymbol{y}$, и поэтому $r(\boldsymbol{x}) \leqslant r(\boldsymbol{y})$. Таким образом, $r = \sup\limits_{\boldsymbol{x} \in \mathcal{M}} r(\boldsymbol{x}) \leqslant \max\limits_{\boldsymbol{y} \in \mathcal{N}} r(\boldsymbol{y})$. Но так как $\mathcal{N} \subset \mathcal{L}$, то

$$\max_{\boldsymbol{y} \in \mathcal{N}} r(\boldsymbol{y}) \leqslant \sup_{\boldsymbol{y} \in \mathcal{L}} r(\boldsymbol{x}) = \sup_{\boldsymbol{x} \in \mathcal{M}} r(\boldsymbol{x}).$$

Следовательно,

$$r = \max_{\boldsymbol{y} \in \mathcal{N}} r(\boldsymbol{y}) \tag{9.2.3}$$

и существует такой $\boldsymbol{y} > \boldsymbol{0}$, что $r = r(\boldsymbol{y})$.

Могут существовать и другие векторы в $\mathcal{L}$, для которых $r(\boldsymbol{x})$ принимает значение r. Любой такой вектор называется *экстремальным* для A. Таким образом, ненулевой вектор $\boldsymbol{z} \geqslant \boldsymbol{0}$ будт экстремальным для $A \Longleftrightarrow r\boldsymbol{z} = A\boldsymbol{z}$.

Наш интерес к числу r объясняется следующим результатом.

Лемма 2. *Если матрица $A \in \mathcal{R}_{n\times n}$ неотрицательна и неприводима, то число r, определяемое формулой* (9.2.2), *положительно и является собственным значением для A. Кроме того, каждый экстремальный вектор для A положителен и является правым собственным вектором для A, соответствующим собственному значению r.*

Доказательство. Пусть $\boldsymbol{u}' = \| 1, 1, \ldots, 1 \|$. Тогда $r(\boldsymbol{u}) = \min\limits_{i} \sum\limits_{k} a_{\,k} > 0$. Ибо, если какая-нибудь строка в A состоит полностью из нулей, то A будет приводимой. Так как $r \geqslant r(\boldsymbol{u})$, то $r > 0$.

Пусть $\boldsymbol{z}$ — некоторый экстремальный вектор и $\boldsymbol{x} = (I + A)^{n-1}\boldsymbol{z}$. Без потери общности можем считать, что $\boldsymbol{z} \in \mathcal{M}$. Лемма 1 означает, что $\boldsymbol{x} > \boldsymbol{0}$, и, очевидно, $\boldsymbol{x} \in \mathcal{N}$. Мы имеем также $A\boldsymbol{z} - r\boldsymbol{z} \geqslant \boldsymbol{0}$. Если $A\boldsymbol{z} - r\boldsymbol{z} \neq \boldsymbol{0}$, то $(I + A)^{n-1}(A\boldsymbol{z} - r\boldsymbol{z}) > \boldsymbol{0}$. Следовательно, $A\boldsymbol{x} - r\boldsymbol{x} > \boldsymbol{0}$, или $r\boldsymbol{x} < A\boldsymbol{x}$, откуда следует, что $r < r(\boldsymbol{x})$. Но это противоречит определению (9.2.2) числа r, так что мы должны иметь $A\boldsymbol{z} = r\boldsymbol{z}$. Таким образом, любой экстремальный собственный вектор $\boldsymbol{z}$ будет правым собственным вектором для A с собственным значением r.

Наконец, так как $A\boldsymbol{z} = r\boldsymbol{z}$, имеем $\boldsymbol{x} = (I + A)^{n-1}\boldsymbol{z} = (1 + r)^{n-1}\boldsymbol{z}$ и из $\boldsymbol{x} > \boldsymbol{0}$, $r > 0$ необходимо следует, что $\boldsymbol{z} > \boldsymbol{0}$. ◀

Мы можем теперь сформулировать и доказать первую часть теоремы Перрона — Фробениуса для неприводимых матриц.

Теорема 9.2.1. *Если матрица $A \in \mathcal{R}_{n\times n}$ неотрицательна и неприводима, то*: (i) *A имеет положительное собственное значение r, равное спектральному радиусу A*; (ii) *существует положительный правый собственный вектор, соответствующий собственному значению r*; (iii) *собственное значение r имеет* (*алгебраическую*) *кратность* 1.

Доказательство. 1. В лемме 2 было получено существование положительного собственного значения с соответствующим положительным собственным вектором. Мы покажем сначала, что другие собственные значения для A не могут превосходить r по абсолютной величине. Этим будет завершено доказательство частей (i) и (ii) теоремы.

Если $\alpha \boldsymbol{y} = A\boldsymbol{y}$ и $\boldsymbol{y} \neq 0$, то, так как $A \geqslant 0$, имеем $|\alpha||\boldsymbol{y}| = |A\boldsymbol{y}| \leqslant A|\boldsymbol{y}|$. Следовательно, $|\alpha| \leqslant r(|y|) \leqslant r$, что нам и нужно было доказать.

2. Предположим теперь, что $\boldsymbol{z}$ — произвольный правый собственный вектор, соответствующий r. Таким образом, $A\boldsymbol{z} = r\boldsymbol{z}$, $\boldsymbol{z} \neq \mathbf{0}$ и, как и выше, $r|\boldsymbol{z}| \leqslant A|\boldsymbol{z}|$.

Это означает, что $|z|$ — экстремальный вектор и, согласно лемме 2, $|z| > 0$. Таким образом, $z_i \neq 0$, $i = 1, 2, \ldots, n$. Это означает, что размерность правого собственного подпространства для r равна 1. В противном случае мы смогли бы найти два линейно независимых правых собственных вектора $\boldsymbol{z}_1$, $\boldsymbol{z}_2$ и затем определить числа α, β так, чтобы $\alpha\boldsymbol{z}_1 + \beta\boldsymbol{z}_2$ имел нулевую координату. Это доказывает, что *геометрическая* кратность для r равна 1. Для завершения доказательства мы используем этот факт, чтобы показать, что если $\Delta(\lambda)$ — характеристический многочлен матрицы A, то $\Delta^{(1)}(r)$, производная от $\Delta(r)$ в r, отлична от нуля.

3. Пусть $B(\lambda) = (I\lambda - A)^{\vee}$. Так как $\mathcal{N}(Ir - A)$ имеет размерность 1, то ранг матрицы $Ir - A$ равен $n - 1$ (теорема 1.16.2), так что существует по крайней мере один ненулевой минор в $Ir - A$ порядка $n - 1$. Следовательно, $B(r) \neq 0$.

Но каждый ненулевой столбец в $B(r)$ является правым собственным вектором для A, соответствующим r (упр. 3 из § 4.5), откуда получаем, что каждый ненулевой столбец в $B(r)$ состоит из ненулевых элементов, которые все имеют один и тот же знак. Однако, применяя проведенное рассуждение к A', то же самое свойство получаем для строк в $B(r)$. Эти результаты будут согласованы в том и только в том случае, когда *каждый* элемент в $B(r)$ ненулевой и одного и того же знака.

Так как $B(\lambda)(I\lambda - A) = \Delta(\lambda)I$, то имеем

$$B^{(1)}(\lambda)(I\lambda - A) + B(\lambda) = \Delta^{(1)}(\lambda)I.$$

Полагая $\lambda = r$ и умножая справа полученное равенство на правый собственный вектор z, получаем

$$B(r)z = \Delta^{(1)}(r)z. \tag{9.2.4}$$

Так как $B(r)$ и z одновременно имеют ненулевые элементы и эти элементы в каждом из них одного и того же знака, то из (9.2.4) следует, что $\Delta^{(1)}(r) \neq 0$. Это завершает доказательство теоремы.

Нетрудно показать, что $B(r) > 0$. Это следует из того, что так как r — наибольший вещественный корень для $\Delta(\lambda)$, то $\Delta^{(1)}(r) > 0$. Применяя (9.2.4), получаем результат.

Изложенные выше обобщения результатов Перрона были опубликованы Фробениусом в 1912 г. При этом им были доказаны также следующие результаты, дающие более детальную информацию о строении матрицы A в случае, когда существует несколько собственных значений с модулем r. Этот результат формулируется здесь для полноты и доказательство его не приводится.

Теорема 9.2.2. *Пусть матрица $A \in \mathscr{R}_{n\times n}$ неотрицательна и неприводима и имеет собственные значения $\mu_1, \mu_2, \ldots, \mu_n$. Если имеется h и только h собственных значений $\mu_1 = r, \mu_2, \ldots, \mu_h$ с модулем r и если $\omega_1, \ldots, \omega_h$ — различные корни степени h из единицы, то $\mu_j = \omega_j r$, $j = 1, 2, \ldots, h$.*

Кроме того, n точек комплексной плоскости, соответствующих $\mu_1, \ldots, \mu_n$, инвариантны относительно вращений около начала координат с углами, кратными $2\pi/h$. При $h > 1$ существует такая перестановочная матрица P, что $P'AP$ имеет симметрично блочный вид

$$\begin{Vmatrix} 0 & A_{12} & 0 & \ldots & 0 \\ 0 & 0 & A_{23} & \ldots & 0 \\ \cdot & & \cdot & & \\ \cdot & & & \cdot & \\ \cdot & & & & \cdot \\ 0 & 0 & \ldots & 0 & A_{h-1,h} \\ A_{h1} & 0 & \ldots & 0 & 0 \end{Vmatrix}.$$

Первоначальные результаты Перрона были опубликованы в 1907 г. и могут быть сформулированы теперь в виде следствия теорем 9.2.1 и 9.2.2.

Следствие. *Положительная квадратная матрица A имеет вещественное и положительное собственное значение r, которое имеет алгебраическую кратность 1 и превосходит модули всех других собственных значений матрицы A. Собственному значению r соответствует положительный правый собственный вектор.*

Последняя часть теоремы 9.2.2 означает, что при $A > 0$ мы должны иметь $h = 1$, и, следовательно, r превосходит модули всех других собственных значений матрицы A.

Упр. 1. Если $A > 0$ и G — компонентная матрица для A, соответствующая r, то доказать, что $G > 0$.

Упр. 2. Если A неприводима и неотрицательна и $C(\lambda)$ — приведенная присоединенная матрица для $I\lambda - A$, то $C(r) > 0$.

Упр. 3. Если $A \in \mathcal{R}_{n\times n}$ неприводима и D — неособая неотрицательная диагональная матрица, то $(D+A)^{n-1} > 0$.

***Упр. 4.** Если $A \in \mathcal{R}_{n\times n}$ неприводима и неотрицательна и $\sigma_j = \sum_{k=1}^{n} a_{jk}$, то доказать, что $\min_j \sigma_j \leqslant r \leqslant \max_j \sigma_j$ и что равенство в обоих случаях выполняется в том и только в том случае, когда $\sigma_1 = \sigma_2 = \ldots = \sigma_n$.

Решение. Прибавляя все предшествующие столбцы в $\det(Ir - A)$ к последнему, получаем

$$\det\begin{Vmatrix} r-a_{11} & -a_{12} & \ldots & -a_{1,n-1} & r-\sigma_1 \\ -a_{21} & r-a_{22} & \ldots & -a_{2,n-1} & r-\sigma_2 \\ \vdots & \vdots & & \vdots & \vdots \\ & & & r-a_{n-1,n-1} & r-\sigma_{n-1} \\ -a_{n1} & -a_{n2} & \ldots & -a_{n,n-1} & r-\sigma_n \end{Vmatrix} = 0.$$

Разлагая этот определитель по последнему столбцу и обозначая (i, j)-элемент в $B(r) = (Ir - A)^{\vee}$ через $b_{ij}(r)$, получим

$$\sum_{j=1}^{n} (r - \sigma_j)\, b_{nj}(r) = 0.$$

Но, как мы видели, $B(r) > 0$, так что $b_{nj}(r) > 0$ для $j = 1, 2, \ldots$ $\ldots, n$. Отсюда следует, что либо $\sigma_1 = \sigma_2 = \ldots = \sigma_n$, либо существуют такие индексы k, l, что $r < \sigma_k$ и $r > \sigma_l$. Следовательно, результат доказан.

Упр. 5. Воспользовавшись упр. 4, без рассмотрения характеристического многочлена доказать, что следующая матрица имеет максимальное вещественное собственное значение 7 и что существует собственное значение, равное 4:

$$\begin{Vmatrix} 3 & 0 & 1 & 0 \\ 1 & 6 & 1 & 1 \\ 2 & 0 & 2 & 0 \\ 1 & 3 & 1 & 4 \end{Vmatrix}.$$

9.3. Приводимые матрицы

Мы рассмотрим теперь вопрос о возможности устранения предположения о том, что A неприводима, из условия теоремы 9.2.1. Заключения (i) и (ii) этой теоремы приводят к результатам для приводимых матриц, потому что такую матрицу можно представить в виде предела последовательности положительных (и, следовательно, неприводимых) матриц. Нужные нам выводы будут следовать тогда из непрерывной зависимости доминирующего собственного значения (и его нормализованного правого собственного вектора) от матричных элементов.

В противоположность этому, как мы знаем по опыту из теории возмущений в главе 7, структура матрицы (как она определяется ее элементарными делителями) не зависит непрерывно от матричных элементов. По этой причине заключение (iii) теоремы 9.2.1 не приводит к полезным результатам в приводимом случае.

Теорема 9.3.1. *Если матрица $A \in \mathscr{R}_{n\times n}$ неотрицательна, то:*

(i) *A имеет вещественное собственное значение r, равное спектральному радиусу A;*

(ii) *существует неотрицательный правый собственный вектор, соответствующий собственному значению r.*

Доказательство. Определим матрицу $D(t) \in \mathscr{R}_{n\times n}$ следующим образом:

$$d_{ij}(t) = \begin{cases} a_{ij} & \text{при} \quad a_{ij} > 0, \\ t & \text{при} \quad a_{ij} = 0. \end{cases}$$

Тогда $D(t) > 0$ при $t > 0$ и $D(0) = A$. Пусть $\rho(t)$ — максимальное вещественное собственное значение $D(t)$ при $t > 0$. Все собственные значения $D(t)$ будут непрерывными функциями от t (§§ 7.2 и 7.5), а так как $\rho(t)$ равно спектральному радиусу $D(t)$ при $t > 0$, то $\lim\limits_{t\to +0} \rho(t) = r$ будет собственным значением матрицы $D(0) = A$, причем это собственное значение равно спектральному радиусу A.

Мы знаем также (лемма 1 из § 7.9), что существует нормализованный правый собственный вектор $\boldsymbol{x}(t)$, который соответствует $\rho(t)$ и зависит непрерывно от t при $t \geqslant 0$. Согласно теореме 9.2.1 $\boldsymbol{x}(t) > \boldsymbol{0}$ при $t > 0$ и, следовательно, $\boldsymbol{x}(t) \geqslant \boldsymbol{0}$ в $t = 0$. ◀

Упр. 1. Что случится, если попытаться обобщить результаты теоремы 9.2.2 на приводимые матрицы при помощи рассуждения о непрерывности?

Упр. 2. Если $A \in \mathscr{R}_{n\times n}$ неотрицательна и $\sigma_j = \sum\limits_{k=1}^{n} a_{jk}$, то доказать, что

$$\min_j \sigma_j \leqslant r \leqslant \max_j \sigma_j.$$

[Указание. К результату упр. 4 из § 9.2 применить рассуждение о непрерывности.]

Упр. 3. Если $A \in \mathcal{R}_{n\times n}$ и $a_{jk} > 0$ при $j \neq k$, то доказать, что собственное значение ρ матрицы A с наибольшей вещественной частью вещественно и имеет кратность 1. Доказать также, что существует *положительный* правый собственный вектор для A, соответствующий ρ. Что можно сказать, если мы имеем лишь $a_{jk} \geqslant 0$ при $j \neq k$?

9.4. Примитивные и импримитивные матрицы

В теоремах 9.2.1 и 9.2.2 выяснено, что структура неотрицательной неприводимой матрицы A зависит от числа h собственных значений, модули которых равны спектральному радиусу A. В частности, удобно различать случаи $h = 1$ и $h > 1$. Таким образом, неприводимая неотрицательная матрица называется *примитивной* или *импримитивной* соответственно тому, будет ли $h = 1$ или $h > 1$. Число h называется *индексом импримитивности.*

Как было отмечено в § 9.2, при $A > 0$ имеем $h = 1$, так что каждая положительная матрица обязательно примитивна.

Индекс импримитивности находится без труда, если известен характеристический многочлен. Предположим, что

$$c(\lambda) = \lambda^n + a_1\lambda^{n_1} + a_2\lambda^{n_2} + \ldots + a_t\lambda^{n_t} \tag{9.4.1}$$

— характеристический многочлен, причем $a_1, a_2, \ldots, a_t$ ненулевые и $n > n_1 > n_2 > \ldots > n_t \geqslant 0$.

Если λ_0 — произвольное собственное значение и $\omega_1, \ldots, \omega_h$ — корни степени h из единицы, то, как мы видели в теореме 9.2.2, $\omega_j\lambda_0$, $j = 1, 2, \ldots, h$, все будут собственными значениями матрицы A. Таким образом, $c(\lambda)$ делится на

$$(\lambda - \omega_1\lambda_0)(\lambda - \omega_2\lambda_0) \ldots (\lambda - \omega_h\lambda_0) = \lambda^h - \lambda_0^h$$

без остатка. Следовательно, для некоторого многочлена g и целого числа k имеем $c(\lambda) = g(\lambda^h)\lambda^k$.

Сравнивая это равенство с (9.4.1), убеждаемся в том, что h будет общим делителем разностей $n - n_j$, $j = 1, 2, \ldots, t$. Предположим, что l — наибольший общий делитель этих разностей. Проводя тогда рассуждение в обратном направлении, получаем, что спектр матрицы A инвариантен относительно вращений на углы, кратные $2\pi/l$. Если $h < l$, то $2\pi/l < 2\pi/h$ и инвариантность при вращениях на углы, кратные $2\pi/l$, не согласуется с определением числа h. Следовательно, $h = l$ и при записи характеристического многочлена в виде (9.4.1) получаем, что *индекс импримитивности h равен наибольшему общему делителю разностей $n - n_j$, $j = 1, 2, \ldots, t$.*

Например, если $c(\lambda) = \lambda^{10} + a_1\lambda^7 + a_2\lambda$, $a_1, a_2 \neq 0$, то $h = 3$, и если $c(\lambda) = \lambda^{10} + a_1\lambda^7 + a_2\lambda + a_3$, $a_1, a_2, a_3 \neq 0$, то $h = 1$.

Следующий результат дает несколько другую характеризацию примитивных матриц.

Теорема 9.4.1. *Неотрицательная матрица A примитивна в том и только в том случае, когда существует такое положительное целое число p, что $A^p > 0$.*

Доказательство. Предположим сначала, что $A^p > 0$ и, следовательно, что A^p неприводима. Как мы видели (в упр. 4 из § 9.1), из приводимости A следовала бы приводимость A^p, а следовательно, $A^p > 0$ означает, что A неприводима. Если бы A имела индекс импримитивности $h > 1$, то, так как собственные значения для A^p являются p-ми степенями собственных значений для A, A^p также имела бы индекс нмпримитивности $h > 1$. Но это противоречило бы теореме Перрона, примененной к положительной матрице A^p. Следовательно, $h = 1$ и A примитивна.

Наоборот, предположим, что A примитивна и, следовательно, что A имеет доминирующее вещественное собственное значение r кратности 1 (теоремы 9.2.2 и 9.2.1). Воспользовавшись теоремой 5.4.1 и положив $r = \lambda$, будем иметь

$$A^p = r^p Z_{11} + \sum_{k=2}^{s} \sum_{j=1}^{m_k} f_k^{(j-1)} Z_{kj},$$

где $f(\lambda) = \lambda^p$ и $f_k^{(j-1)}$ — $(j-1)$-я производная от f в λ_k. Из этого равенства легко получаем, что, так как r — доминирующее собственное значение,

$$\lim_{p\to\infty} \frac{A^p}{r^p} = Z_{11}. \tag{9.4.2}$$

Из упр. 3 в § 5.4 имеем также

$$Z_{11} = \frac{C(r)}{\psi^{(1)}(r)},$$

и минимальный многочлен ψ задается в виде

$$\psi(\lambda) = (\lambda - r) \prod_{j=2}^{s} (\lambda - \lambda_j)^{m_j}.$$

Так как r — наибольший вещественный корень для ψ и ψ — приведенный многочлен, отсюда следует, что $\psi^{(1)}(r) > 0$. Так как $C(r) > 0$ (упр. 2 из § 9.2), то $Z_{11} > 0$. Из (9.4.2) теперь следует, что $A^p > 0$ при некотором достаточно большом p.

Упр. 1. Если A примитивна и p — положительное целое число, то доказать, что A^p примитивна.

Упр. 2. Если $A \geqslant 0$ неприводима и $\varepsilon > 0$, то доказать при помощи рассмотрения собственных значений, что $\varepsilon I + A$ примитивна. Сравнить этот результат с леммой 1 из § 9.2.

9.5. Стохастические матрицы

Матрица $P \in \mathscr{R}_{n\times n}$ называется *стохастической*, если $P \geqslant 0$ и

$$\sum_{j=1}^{n} p_{ij} = 1, \qquad i = 1, 2, \ldots, n. \tag{9.5.1}$$

Матрицы такого типа возникают в задачах теории вероятностей, которые будут рассмотрены в § 9.6. Для последующих приложений в этом параграфе будут получены некоторые из их наиболее полезных свойств.

Теорема 9.5.1. *Неотрицательная матрица P будет стохастической в том и только в том случае, когда она имеет собственное значение* 1 *с правым собственным вектором, задаваемым в виде $\boldsymbol{u}' = \| 1, 1, \ldots, 1 \|$. Кроме того, спектральный радиус стохастической матрицы равен* 1.

Доказательство. Если P стохастическая, то условие (9.5.1) может быть, очевидно, записано в виде $P\boldsymbol{u} = \boldsymbol{u}$. Следовательно, 1 будет собственным значением с правым собственным вектором $\boldsymbol{u}$.

Наоборот, если $P\boldsymbol{u} = \boldsymbol{u}$, то выполняется (9.5.1) и, следовательно, P стохастическая.

Для последней части теоремы следует воспользоваться упр. 2 из § 9.3, чтобы показать, что *доминирующее* вещественное значение матрицы P равно 1. ◀

В следующей теореме мы убедимся, что широкий класс неотрицательных матриц может быть при помощи простого преобразования сведен к стохастическим матрицам. Заметим, что некоторые неотрицательные матрицы исключаются предположением о том, что $\boldsymbol{z}$ *положителен*; ср. с теоремой 9.3.1.

Теорема 9.5.2. *Пусть A — неотрицательная матрица с максимальным вещественным собственным значением r. Если существует положительный правый собственный вектор $\boldsymbol{z}$, соответствующий r, и мы положили $Z = \operatorname{diag}\{z_1, \ldots, z_n\}$, то*

$$A = rZPZ^{-1},$$

где P — стохастическая матрица.

Доказательство. Пусть $P = r^{-1}Z^{-1}AZ$. Мы должны лишь доказать, что P стохастическая. Так как $A\boldsymbol{z} = r\boldsymbol{z}$, то

$$\sum_{j=1}^{n} a_{ij} z_j = r z_i; \qquad i = 1, 2, \ldots, n.$$

По определению $p_{ij} = r^{-1} z_i^{-1} a_{ij} z_j$, и, следовательно,

$$\sum_{j=1}^{n} p_{ij} = r^{-1} z_i^{-1} \sum_{j=1}^{k} a_{ij} z_j = 1.$$

Таким образом, P стохастическая, и теорема доказана.

Мы рассмотрим, наконец, последовательность неотрицательных матриц P, P^2, P^3, ..., где P стохастическая. Нас будут интересовать условия, при которых существует

$$P^{\infty} = \lim_{p \to \infty} P^p.$$

Мы уже знаем, что $P^{\infty} \neq 0$, ибо P имеет собственное значение, равное 1, и $P^r \to 0$ при $r \to \infty \Longleftrightarrow$ все собственные значения P по модулю меньше 1 (упр. 5 из § 7.1).

Теорема 9.5.3. *Если P — неприводимая стохастическая матрица, то матрица $P^{\infty} = \lim_{p \to \infty} P^p$ существует $\Longleftrightarrow$ P примитивна.*

Доказательство. 1. Как мы знаем, P имеем максимальное вещественное собственное значение, равное 1. Предположим, что $\mu_2, \ldots, \mu_h$ — собственные значения P с модулем 1 и $\mu_{h+1}, \ldots, \mu_s$ — различные собственные значения P с модулем, меньшим 1. Положим $f(\lambda) = \lambda^p$ и обозначим через $f_k^{(j-1)}$ $(j-1)$-ю производную от f в μ_k. Напомним, что, так как P неприводима, собственное значение 1 не кратное (теорема 9.2.1, часть (iii). Опять воспользовавшись теоремой 5.4.1, получаем

$$P^p = Z_{11} + \sum_{k=2}^{h} \sum_{j=1}^{m_k} f_k^{(j-1)} Z_{kj} + \sum_{k=h+1}^{s} \sum_{j=1}^{m_k} f_k^{(j-1)} Z_{kj}. \qquad (9.5.2)$$

Для достаточно большого p имеем

$$f_k^{(j-1)} = \frac{p!}{(p-j+1)!} \mu_k^{p-j+1} \qquad (9.5.3)$$

и, так как $|\mu_k| < 1$ для $k = h+1, \ldots, s$, отсюда следует, что последний член в (9.5.2) стремится к нулевой матрице при $p \to \infty$.

2. Если предположить теперь, что P примитивна, то второй член справа в (9.5.2) будет отсутствовать, и ясно, что P^{∞} существует и, кроме того, $P^{\infty} = Z_{11}$.

3. Наоборот, предположим, что известно существование P^{∞}. Тогда из (9.5.2) получаем, что существует

$$Q = \lim_{p \to \infty} \sum_{k=2}^{h} \sum_{j=1}^{m_k} f_k^{(j-1)} Z_{kj}.$$

$m_2 + m_3 + \ldots + m_h$ матриц Z_{kj}, входящих в это суммирование, являются линейно независимыми элементами в $\mathscr{C}_{n \times n}$ (теорема 5.4.1) и порождают, скажем, подпространство $\mathscr{S}$ в $\mathscr{C}_{n \times n}$. Так как Q тоже должна принадлежать $\mathscr{S}$, то существуют такие числа $\alpha_{k,j-1}$, что $Q=\sum_{k=2}^{h}\sum_{j=1}^{m_k}\alpha_{k,j-1}Z_{kj}$. Отсюда следует, что для каждых k и j $\alpha_{k,j-1}=\lim_{p\to\infty} f_k^{(j-1)}$. Но для $k=2, 3, \ldots, h$ имеем $|\mu_k|=1$ $(\mu_k\neq 1)$ и из (9.5.2) следует, что предел справа не существует. Таким образом, получено противоречие, которое может быть разрешено лишь при условии, что из существования P^∞ следует равенство $h=1$. ◀

Следует заметить, что предположение о том, что P неприводима, несущественно. Можно доказать, что собственное значение 1 матрицы P имеет лишь линейные элементарные делители, даже если P приводима. Таким образом, (9.5.2) все еще выполняется и проведенное выше рассуждение применимо. Доказательство более общего результата требует более глубокого рассмотрения приводимых неотрицательных матриц, чем это входит в наши намерения.

Отметим также, что, воспользовавшись упр. 3 из § 5.4, можем написать

$$P^\infty=\frac{C(1)}{\psi^{(1)}(1)},$$

где $C(\lambda)$, $\psi(\lambda)$ — приведенная присоединенная матрица и минимальный многочлен для P. Если $B(\lambda)=(I\lambda-P)^\vee$ и $c(\lambda)$ — характеристический многочлен матрицы P, то многочленное тождество

$$\frac{B(\lambda)}{c(\lambda)}=\frac{C(\lambda)}{\psi(\lambda)}$$

означает, что можно также написать

$$P^\infty=\frac{B(1)}{c^{(1)}(1)}.$$

Еще одна возможность возникает, если отождествить Z_{11} с сопутствующей матрицей G_1 из теоремы 2.5.1. Таким образом, в случае, когда P неприводима, существуют положительные правый и левый собственные векторы $\boldsymbol{x}$, $\boldsymbol{y}$ соответственно с собственным значением 1, для которых

$$\boldsymbol{y}'\boldsymbol{x}=1 \quad \text{и} \quad P^\infty=\boldsymbol{x}\boldsymbol{y}'. \tag{9.5.4}$$

9.6. Цепи Маркова

Неотрицательные матрицы возникают во многих физических задачах. Мы коснемся вкратце одной задачи теории вероятностей, которая использует некоторые из полученных ранее в этой главе теорем.

Для простоты мы рассмотрим физическую систему, которая может быть полностью описана n различными состояниями. Обозначим эти состояния через $s_1, s_2, \ldots, s_n$. Мы предположим, что систему возможно наблюдать в моменты времени $t_0 < t_1 < < t_2 < \ldots$ и отмечать ее состояние в каждый из этих моментов. Мы предположим, кроме того, что абсолютная вероятность нахождения системы в состоянии s_i в момент времени t_r для $i = 1, 2, \ldots, n$ и $r = 0, 1, 2, \ldots$ равна $p_i(t_r)$. Так как система должна быть в одном из n состояний в момент времени t_r, то имеем

$$\sum_{i=1}^{n} p_i(t_r) = 1, \qquad r = 0, 1, 2, \ldots \tag{9.6.1}$$

Предполагается, что при всех i и j мы можем определить вероятность того, что система в момент времени t_r переходит в состояние s_i после того, как она в момент времени t_{r-1} находилась в состоянии s_j. Таким образом, мы предполагаем, что известны *условные вероятности* $q_{ij}(t_r \mid t_{r-1})$, $i, j = 1, 2, \ldots, n$, перехода в момент времени t_r в состояние s_i, если задано, что в момент времени t_{r-1} система находилась в состоянии s_j. Правила теории вероятностей тогда означают, что

$$p_i(t_r) = \sum_{j=1}^{n} q_{ij}(t_r \mid t_{r-1})\, p_j(t_{r-1})$$

или в матричном виде

$$\boldsymbol{p}_r = Q(t_r, t_{r-1})\, \boldsymbol{p}_{r-1}, \qquad r = 1, 2, \ldots, \tag{9.6.2}$$

где положено

$$\boldsymbol{p}'_s = \| p_1(t_s),\ p_2(t_s),\ \ldots,\ p_n(t_s) \|, \qquad s = 0, 1, 2, \ldots,$$

и $Q(t_r, t_{r-1})$ — матрица из $\mathcal{R}_{n\times n}$ с элементами $q_{ij}(t_r \mid t_{r-1})$.

Так как система должна перейти в одно из n состояний в момент времени t_r после того, как она была в состоянии s_j в момент времени t_{r-1}, то

$$\sum_{i=1}^{n} q_{ij}(t_r \mid t_{r-1}) = 1. \tag{9.6.3}$$

Наконец, предположим, что известно начальное состояние системы. Так, если начальное состояние есть s_i, то $\boldsymbol{p}_0 = \boldsymbol{e}_i$ — вектору с единицей на месте i и нулями на остальных местах. Из (9.6.2) получаем, что в момент времени t_r

$$\boldsymbol{p}_r = Q(t_r, t_{r-1})\, Q(t_{r-1}, t_{r-2}) \ldots Q(t_1, t_0)\, \boldsymbol{p}_0. \tag{9.6.4}$$

Только что описанный процесс известен как *цепь Маркова*. Если условные вероятности $q_{ij}(t_r|t_{r-1})$ не зависят от t или одинаковы в каждый момент $t_0, t_1, t_2, \ldots$, то процесс описывается как *однородная* цепь Маркова. Мы ограничимся рассмотрением этого относительно простого случая. Отметим, что теперь можно записать $Q(t_r, t_{r-1}) = Q$, $r = 1, 2, \ldots$; равенство (9.6.4) переходит в $\boldsymbol{p}_r = Q^r\boldsymbol{p}_0$ и (9.6.3) означает, что матрица Q, соответствующая однородной цепи Маркова, такова, что Q' стохастическая. Для удобства мы полагаем $P = Q'$.

Во многих задачах, использующих цепи Маркова, нам интересно знать предельное поведение последовательности $\boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{p}_2, \ldots$ Это приводит нас к необходимости рассматривать вопрос о существовании $\lim\limits_{r\to\infty} Q^r$ или $\lim\limits_{r\to\infty} P^r$. Если P (или Q) неприводима, то, согласно теореме 9.5.3, эти пределы существуют $\Longleftrightarrow P$ примитивна. Так как $P = Q'$, то Q имеет собственное значение 1 и правый и левый собственные векторы $\boldsymbol{x}$, $\boldsymbol{y}$ из (9.5.4) будут левым и правым собственными веикторами для Q соответственно. Таким образом, существуют правый и левый собственные векторы $\boldsymbol{\xi}(=\boldsymbol{y})$ и $\boldsymbol{\eta}(=\boldsymbol{x})$ для Q с собственным значением 1, для которых $\boldsymbol{\eta}'\boldsymbol{\xi} = 1$ и $Q^\infty = \boldsymbol{\xi}\boldsymbol{\eta}'$. Имеем тогда $\boldsymbol{p}_\infty = (\boldsymbol{\eta}'\boldsymbol{p}_0)\boldsymbol{\xi}$, так что, каков бы ни был начальный вектор $\boldsymbol{p}_0$, предельный вектор будет пропорционален $\boldsymbol{\xi}$. Но $\boldsymbol{p}_r$ удовлетворяет условию (9.6.1) при каждом r, так что $\boldsymbol{p}_\infty$ удовлетворяет тому же условию. Таким образом, если определить положительный вектор $\boldsymbol{\xi}$ так, что $\sum\limits_{i=1}^{n} \xi_i = 1$, то $\boldsymbol{p}_\infty = \boldsymbol{\xi}$ при каждом возможном выборе $\boldsymbol{p}_0$.

Тот факт, что $\boldsymbol{\xi} > \boldsymbol{0}$, означает, что можно закончить любым из n состояний $s_1, s_2, \ldots, s_n$. Мы доказали, что: *если Q — матрица условных вероятностей, соответствующая некоторой однородной цепи Маркова, и Q неприводима, то предельные абсолютные вероятности $p_i(t_\infty)$ существуют $\Longleftrightarrow Q$ примитивна. Кроме того, если Q примитивна, то $p_i(t_\infty) > 0$ при $i = 1, 2, \ldots, n$ и эти вероятности являются компонентами (должным образом нормализованного) правого собственного вектора для Q, соответствующего собственному значению* 1.

ДОПОЛНЕНИЕ 1

НЕКОТОРЫЕ ТЕОРЕМЫ ИЗ АНАЛИЗА

В этом дополнении мы обсуждаем без доказательства некоторые важные теоремы из анализа, которыми мы воспользовались несколько раз в основной части книги. Эти результаты формулируются в настолько общем виде, насколько это нужно для наших приложений, хотя они могут быть доказаны так же легко и в более общем контексте. Приведем сначала некоторые определения.

Напомним, что любая векторная норма в $\mathscr{C}_n$ может быть использована в качестве меры величины расстояния (§ 6.2). Для заданной векторной нормы h на $\mathscr{C}_n$ мы определяем *окрестность* $N_r(\boldsymbol{x})$ точки $\boldsymbol{x} \in \mathscr{C}_n$ как множество точек $\boldsymbol{y}$ в $\mathscr{C}_n$, для которых $h(\boldsymbol{y}-\boldsymbol{x}) \leqslant r$ (т. е. это есть *шар* с центром в $\boldsymbol{x}$ и радиусом r, см. упр. 12 из § 6.2). Если $\mathscr{S}$ — некоторое множество точек в $\mathscr{C}_n$, $\boldsymbol{x} \in \mathscr{C}_n$ будет *предельной точкой* для $\mathscr{S}$, если каждая окрестность для $\boldsymbol{x}$ содержит некоторую точку $\boldsymbol{p} \in \mathscr{S}$ с $\boldsymbol{p} \neq \boldsymbol{x}$. Множество $\mathscr{S}$ называется *замкнутым*, если каждая предельная точка для $\mathscr{S}$ принадлежит $\mathscr{S}$, и $\mathscr{S}$ *ограничено*, если существуют такие $K \geqslant 0$ и $\boldsymbol{x} \in \mathscr{C}_n$, что $h(\boldsymbol{y}-\boldsymbol{x}) < K$ для всех $\boldsymbol{y} \in \mathscr{S}$. Мы можем сформулировать теперь первый результат:

Теорема 1. *Пусть X — замкнутое и ограниченное множество вещественных чисел и M, m — верхняя и нижняя грани для чисел x в M. Тогда $M, m \in X$.*

Доказательство является непосредственным выводом из определений.

Рассмотрим теперь некоторую функцию f со значениями в $\mathscr{C}_m$, определенную на $\mathscr{S} \subseteq \mathscr{C}_n$. Мы говорим, что f *непрерывна* в $\boldsymbol{x} \in \mathscr{S}$, если для заданного $\varepsilon > 0$ существует такое $\delta > 0$, что из $\boldsymbol{y} \in N_\delta(\boldsymbol{x})$ следует

$$h(f(\boldsymbol{y}) - f(\boldsymbol{x})) < \varepsilon.$$

Мы говорим, что f непрерывна на $\mathscr{S}$, если f непрерывна в каждой точке $\mathscr{S}$. Множество $f(\mathscr{S}) \subseteq \mathscr{C}_m$ определяется условием: $\boldsymbol{a} \in f(\mathscr{S})$ в том и только в том случае, когда существует такая точка $\boldsymbol{b} \in \mathscr{C}_n$, что $\boldsymbol{a} = f(\boldsymbol{b})$.

Теорема 2. *Пусть f — определенная выше функция. Если $\mathscr{S}$ — замкнутое и ограниченное множество в $\mathscr{C}_n$ и f непрерывна на $\mathscr{S}$, то $f(\mathscr{S})$ будет замкнутым и ограниченным подмножеством в $\mathscr{C}_m$.*

В этой и следующей теоремах можно заменить $\mathscr{C}_m$ на $\mathscr{R}_m$, $\mathscr{C}_n$ на $\mathscr{R}_n$ или произвести обе замены одновременно. Последняя теорема получается в качестве следствия теорем 1 и 2.

Теорема 3. *Пусть f — непрерывная вещественнозначная функция, определенная на замкнутом и ограниченном подмножестве $\mathscr{S}$ в $\mathscr{C}_n$. Если*

$$M = \sup_{\boldsymbol{x} \in \mathscr{S}} f(\boldsymbol{x}), \quad m = \inf_{\boldsymbol{x} \in \mathscr{S}} f(\boldsymbol{x}),$$

то существуют такие векторы $\boldsymbol{y}, \boldsymbol{z} \in \mathscr{S}$, что

$$M = f(\boldsymbol{y}), \quad m = f(\boldsymbol{z}).$$

В этом утверждении $\sup\limits_{\boldsymbol{x} \in \mathscr{C}} f(\boldsymbol{x})$ обозначает супремум, или верхнюю грань чисел $f(\boldsymbol{x})$, где $\boldsymbol{x} \in \mathscr{S}$. Подобно этому $\inf\limits_{\boldsymbol{x} \in \mathscr{S}} f(\boldsymbol{x})$ обозначает инфимум, или нижнюю грань.

Согласно теореме 2, при предположениях теоремы 3 $f(\mathscr{S})$ будет замкнутым и ограниченным множеством вещественных чисел и результат будет следовать из теоремы 1. Можно сказать, что при предположениях теоремы 3 f имеет максимум и минимум на $\mathscr{S}$, и мы можем записать

$$\sup_{\boldsymbol{x} \in \mathscr{S}} f(\boldsymbol{x}) = \max_{\boldsymbol{x} \in \mathscr{S}} f(\boldsymbol{x}), \quad \inf_{\boldsymbol{x} \in \mathscr{S}} f(\boldsymbol{x}) = \min_{\boldsymbol{x} \in \mathscr{S}} f(\boldsymbol{x}).$$

Мы приведем теперь три простых примера, показывающих, что три условия: f непрерывна на $\mathscr{S}$, $\mathscr{S}$ ограничено и $\mathscr{S}$ замкнуто — все необходимы.

(а) Пусть $f(x) = x$, $\mathscr{S}$ — интервал $(0, 1)$ и $x \in \mathscr{S}$. Здесь $\mathscr{S}$ ограничено, но не замкнуто, и f непрерывна на $\mathscr{S}$. Очевидно, что $M = 1$, $m = 0$ и f не принимает этих значений на $\mathscr{S}$.

(б) При $f(x) = 1/x$ и множестве $\mathscr{S}$ всех вещественных чисел, не меньших 1, имеем, что f непрерывна на $\mathscr{S}$ и $\mathscr{S}$ замкнуто, но не ограничено. В этом случае $f(\mathscr{S})$ — полуоткрытый интервал $(0, 1]$ и $\inf\limits_{\boldsymbol{x} \in \mathscr{S}} f(\boldsymbol{x}) = 0$ не принадлежит $f(\mathscr{S})$.

(в) При

$$f(x) = \begin{cases} x, & 0 \leqslant x < 1, \\ 0, & x = 1, \end{cases}$$

имеем $\mathscr{S} = [0, 1]$, что замкнуто и ограничено, но f разрывна на $\mathscr{S}$. Здесь $f(\mathscr{S}) = [0, 1]$ и $\sup\limits_{x \in \mathscr{S}} f(x) = 1$ не принадлежит $f(\mathscr{S})$.

Литература к дополнению 1

Buck R. C., Advanced Calculus, McGraw-Hill, N. Y., 1956.
Rudin W., Principles of Mathematical Analysis, McGraw-Hill, N. Y., 1964.

ДОПОЛНЕНИЕ 2

ОБОБЩЕННАЯ ОБРАТНАЯ МАТРИЦА

В этом дополнении мы введем обобщение для обращения матриц. Это обобщенное обращение имеет некоторые важные свойства обычного обращения, но определено не только для особых квадратных матриц, но и для прямоугольных матриц. Прежде чем читать это дополнение, следует быть хорошо знакомым с содержанием § 2.11.

Пусть $A \in \mathscr{C}_{m\times n}$ и имеет ранг r. Выберем некоторый базис $f_1, f_2, \ldots, f_r$ для образа, или пространства столбцов, матрицы A и положим $F = \| f_1, f_2, \ldots, f_r \|$. Заметим, что столбцами в F можно взять любые r линейно независимых столбцов в A. Тогда каждый столбец из A будет линейной комбинацией столбцов из F. Таким образом, можно записать $A_{*j} = F\rho_j$ при некотором векторе $\rho_j \in \mathscr{C}_r$, $j = 1, 2, \ldots, n$, и существует такая $n \times r$-матрица R, что

$$A = FR^* \tag{1}$$

(j-м столбцом в R^* является ρ_j). Из этого построения легко получаем, что ранг R также равен r (упр. 2 из § 1.15). Кроме того, так как $A^* = RF^*$, убеждаемся в том, что столбцы из R образуют базис для $\mathscr{R}(A^*)$ — образа матрицы A^*. Вспоминая (2.11.1), видим, что если положить

$$P = F(F^*F)^{-1}F^*, \quad Q = R(R^*R)^{-1}R^*, \tag{2}$$

то $P \in \mathscr{C}_{m\times m}$ *будет проекцией на* $\mathscr{R}(A)$ *и* $Q \in \mathscr{C}_{n\times n}$ *будет проекцией на* $\mathscr{R}(A^*)$.

Мы определяем теперь $n \times m$-матрицу

$$A^I = R(R^*R)^{-1}(F^*F)^{-1}F^*. \tag{3}$$

Объединяя это определение с (1), непосредственно убеждаемся в том, что

$$AA^I = F(F^*F)^{-1}F^* = P, \quad A^IA = R(R^*R)^{-1}R^* = Q. \tag{4}$$

Матрица A^I называется *обобщенной обратной* для A. Основание для этой терминологии состоит в том, что P ведет себя как единичная матрица, если ее применять к векторам из $\mathscr{R}(A)$, т. е. при $\boldsymbol{x} \in \mathscr{R}(A)$ имеем $P\boldsymbol{x} = I\boldsymbol{x} = \boldsymbol{x}$. Из равенства $AA^I = P$ тогда видим, что A^I обладает свойством, аналогичным свойству матрицы A^{-1}, когда она определена. Кроме того, если A неособая, то $\mathscr{R}(A) = \mathscr{C}_n$, $P = I$ и $A^I = A^{-1}$. Подобное же рассуждение применимо к Q и $\mathscr{R}(A^*)$.

Умножая равенства (4) справа на A и A^I соответственно, получаем следующие результаты, характеристические для обратной к неособой матрице:

$$AA^IA = A, \quad A^IAA^I = A^I. \tag{5}$$

Мы можем теперь доказать, что определение (3) обобщенной обратной приводит к *единственной* матрице, удовлетворяющей равенствам (4) и (5).

Теорема 1. *Если $A \in \mathscr{C}_{m\times n}$, P — проекция на $\mathscr{R}(A)$ и Q — проекция на $\mathscr{R}(A^*)$, то существует одна и лишь одна матрица A^I, для которой*

$$AA^I = P, \quad A^IA = Q$$

и

$$AA^IA = A, \quad A^IAA^I = A^I.$$

Доказательство. Мы уже видели, что матрица A^I, определяемая формулой (3), обладает требуемыми свойствами. Предположим, что A_1^I и A_2^I обе удовлетворяют этим четырем равенствам.

Так как $A_1^IA = Q$, то можно написать

$$A_1^IAA_2^I = QA_2^I.$$

Кроме того, $A_2^IA = Q$ и $A_2^IAA_2^I = A_2^I$ означают, что $QA_2^I = A_2^I$. Следовательно,

$$A_1^IAA_2^I = A_2^I. \tag{6}$$

Но $AA_2^I = P$ приводит к

$$A_1^IAA_2^I = A_1^IP$$

и $AA_1^I = P$, $A_1^IAA_1^I = A_1^I$ означают, что $A_1^IP = A_1^I$. Следовательно,

$$A_1^IAA_2^I = A_1^I.$$

Сравнивая это с (6), убеждаемся в том, что $A_1^I = A_2^I$. ◀

Обобщенная обратная имеет красивое приложение при изучении решений, или возможно приближенных решений, уравнения $A\boldsymbol{x} = \boldsymbol{c}$, где $A \in \mathscr{C}_{m\times n}$. Как мы видели в теореме 1.16.3, решение этого уравнения существует $\Longleftrightarrow \boldsymbol{c} \in \mathscr{R}(A)$. Если это так, то вектор $\boldsymbol{x}_0 = A^I\boldsymbol{c}$ будет решением уравнения $A\boldsymbol{x} = \boldsymbol{c}$, ибо $\boldsymbol{c} \in \mathscr{R}(A)$ означает, что $P\boldsymbol{c} = \boldsymbol{c}$, а поэтому

$$A\boldsymbol{x}_0 = AA^I\boldsymbol{c} = P\boldsymbol{c} = \boldsymbol{c}.$$

Конечно, x_0 может быть или не быть *единственным* решением (теорема 1 16.5). Если он является единственным решением, то $A^I = A^{-1}$ и задача не имеет большого интереса в данном контексте.

Введем обозначение $\|x\| = \langle x, x\rangle^{1/2}$ для длины вектора x (§ 1.3) и заметим, что x будет решением уравнения $Ax = c \Longleftrightarrow \Longleftrightarrow \|Ax - c\| = 0$.

В случае $c \notin \mathcal{R}(A)$ можно определить *приближенное решение* уравнения $Ax = c$ (или решение в смысле наименьшего квадратичного отклонения) как вектор $x = x_0$, для которого $\|Ax - c\|$ минимизируется. Мы намереваемся показать, что $x_0 = A^I c$ будет таким вектором. Во всяком случае это стационарное свойство может быть использовано для характеризации решений в случае $c \in \mathcal{R}(A)$. Действительно, мы докажем, что $x_0 = A^I c$ будет *либо* единственным вектором (приближенным решением или решением, смотря какой случай представился), для которого $\|Ax - c\|$ минимизируется, *либо* единственным вектором, для которого $\|Ax - c\|$ и $\|x\|$ обе минимизируются. Нам понадобится следующая

Лемма. *Пусть $A \in \mathscr{C}_{m \times n}$ и P, Q, A^I определены выше. Тогда для произвольных векторов $g \in \mathscr{C}_n$ и $h \in \mathscr{C}_m$ имеем*:

$$\text{(i)} \qquad \|Ag + (I - P)h\|^2 = \|Ag\|^2 + \|(I - P)h\|^2$$

и

$$\text{(ii)} \qquad \|A^I h + (I - Q)g\|^2 = \|A^I h\|^2 + \|(I - Q)g\|^2.$$

Доказательство. Так как P — проекция на $\mathcal{R}(A)$, то $I - P$ будет проекцией на ортогональное дополнение к $\mathcal{R}(A)$ (упр. 5 из § 2.11 и теорема 2.11.1). Таким образом, в формуле (i) Ag и $(I - P)h$ — ортогональные векторы из $\mathscr{C}_m$. Результат следует теперь из обобщения теоремы Пифагора (упр. 3 из § 1.3).

Для получения формулы (ii) применяем то же самое рассуждение к A^I, замечая, что $A^{II} = A$ и что Q — проекция на $\mathcal{R}(A^I)$. ◀

Теорема 2. *Пусть $A \in \mathscr{C}_{m \times n}$, $c \in \mathscr{C}_m$ и $x_0 = A^I c$. Тогда каждый вектор $x \in \mathscr{C}_n$, $x \neq x_0$, удовлетворяет одному из двух условий*:

$$\text{(а)} \qquad \|Ax - c\| > \|Ax_0 - c\|$$

или

$$\text{(б)} \qquad \|Ax - c\| = \|Ax_0 - c\| \text{ и } \|x\| > \|x_0\|.$$

Доказательство. 1. Для любого $x \in \mathscr{C}_n$ запишем

$$Ax - c = A(x - A^I c) + (I - AA^I)(-c).$$

Так как $AA^I = P$, можно применить часть (i) леммы и получить

$$\| Ax - c \|^2 = \| A(x - x_0) \|^2 + \| Ax_0 - c \|^2 \geqslant \| Ax_0 - c \|^2$$

с равенством только в том случае, когда $Ax = Ax_0$. Остается лишь рассмотреть этот случай и получить условие (б).

2. В случае равенства имеем $Ax = Ax_0$ и, так как $x_0 = A^I c$,

$$A^I A x = A^I A x_0 = (A^I A A^I) c.$$

Но из теоремы 1 имеем $A^I A A^I = A^I$, так что

$$A^I A x = A^I c = x_0.$$

Таким образом, можно написать

$$x = x_0 + (x - x_0) = A^I c + (I - A^I A) x.$$

Применяя часть (ii) леммы, получаем

$$\| x \|^2 = \| x_0 \|^2 + \| x - x_0 \|^2.$$

Следовательно, $\| x \| \geqslant \| x_0 \|$ с равенством в том и только в том случае, когда $x = x_0$. ◀

Упр. 1. Найти обобщенную обратную к матрице

$$A = \begin{Vmatrix} 1 & 1 \\ 2 & 0 \\ -1 & 3 \end{Vmatrix}$$

и показать, что единственным приближенным решением уравнения $Ax = c$, определяемым теоремой 2, будет

$$\frac{1}{14}\begin{Vmatrix} 1 \\ 10 \end{Vmatrix}, \quad \text{если } c = \begin{Vmatrix} 1 \\ 0 \\ 2 \end{Vmatrix}.$$

Упр. 2. Обобщить понятие «приближенное решение» и затем теорему 2 с тем, чтобы включить в рассмотрение уравнение $AX = C$, где A, X, C — комплексные матрицы, порядки которых ограничены лишь условием, что написанное уравнение имеет смысл.

Литература к дополнению 2

Grevill T. N. E., The pseudoinverse of a rectangular matrix, SIAM Rev. 1 (1959), 38—43.

Penrose R. A., A generalized inverse for matrices, Proc. Cambridge Phil. Soc. 51 (1955), 406—413.

ДОПОЛНЕНИЕ 3

РЕКОМЕНДАЦИИ ДЛЯ ДАЛЬНЕЙШЕГО ЧТЕНИЯ

Справочные работы общего характера:

Гантмахер Ф. Р., Теория матриц, «Наука», М., 1967.

MacDuffee C. C., The theory of matrices, Chelsea, N. Y., 1946.

Маркус М., Минк Х., Обзор по теории матриц и матричным неравенствам, «Наука», М., 1972.

Работы, относящиеся в основном к материалу глав 1—4:

Finkbeiner D. T., Introduction to matrices and linear transformation, Freeman, San Francisco, 1966.

Халмош П., Конечномерные векторные пространства, Физматгиз, М., 1963.

Hamburger H. L. and Grimshaw M. E., Linear transformations in n-dimensional vector space, Cambridge Univ. Press, London — N. Y., 1956.

Hohn F. E., Elementary matrix algebra, Macmillan, N. Y., 1964.

Mirsky L., An introduction to linear algebra, Oxford Univ. Press (Clarendon), London — N. Y., 1961.

Perlis S, Theory of matrices, Addison-Wesley, Mass., 1958.

Глава 5:

Гантмахер Ф. Р., Теория матриц, «Наука», М., 1967.

Rinehart R. F., The equivalence of definitions of a matrix function, Amer. Math. Monthly **62** (1955), 395—414.

Глава 6:

Householder A. S., The theory of matrices in numerical analysis, Ginn (Blaisdell), Boston, 1964.

Фаддеев Д. К., Фаддеева Н. Н., Вычислительные методы линейной алгебры, Физматгиз, М., 1963.

Ostrowski A. M., Über Normen von Matrizen, Math. Z. **63** (1955), 2—18.

Глава 7:

Householder A. S., The theory of matrices in numerical analysis, Ginn (Blaisdell), Boston, 1964.

Kato T., Perturbation theory for linear operators, Springer, N. Y., 1966.

Wilkinson J. H., The algebraic eigenvalue problem, Oxford Univ. Press (Clarendon), London — N. Y., 1965.

Глава 8:

Беллман Р., Введение в теорию матриц, «Наука», М., 1976.

Гантмахер Ф. Р., Теория матриц, «Наука», М., 1967.

McDuffee C. C., The theory of matrices, Chelsea, N. Y., 1959.

Ostrowski A. M. and Schneider H., Some theorems on the inertia of general matrices, J. Math. Anal. Appl. **4** (1962), 72—84.

Глава 9:

Беллман Р., Введение в теорию матриц, «Наука», М., 1976.

Гантмахер Ф. Р., Теория матриц, «Наука», М., 1967.

Varga R. S., Matrix iterative analysis, Prentice-Hall, Englewood Cliffs, N. J., 1962.

АЛФАВИТНЫЙ УКАЗАТЕЛЬ