С.-Петербургский государственный университет

В. Н. МАЛОЗЁМОВ

ЛИНЕЙНАЯ АЛГЕБРА
БЕЗ ОПРЕДЕЛИТЕЛЕЙ.
КВАДРАТИЧНАЯ ФУНКЦИЯ

Учебное пособие

САНКТ-ПЕТЕРБУРГ
ИЗДАТЕЛЬСТВО С.-ПЕТЕРБУРГСКОГО
УНИВЕРСИТЕТА
1997

УДК 512.64(07)

Малозёмов В.Н. Линейная алгебра без определителей. Квадратичная функция.

В книге представлены начальные сведения из линейной алгебры. Изложение базируется на понятии линейной независимости векторов и на свойствах квадратичной функции.

Вводится и изучается понятие обратной матрицы. Устанавливается критерий совместности системы линейных уравнений. Описывается эффективный алгоритм построения QR-разложения прямоугольной матрицы. Доказываются общие теоремы о существовании и характеризации точки минимума выпуклой квадратичной функции на аффинном многообразии. Они позволяют получить спектральное разложение симметричной матрицы и сингулярное разложение произвольной прямоугольной матрицы. Рассматриваются две конкретные экстремальные задачи: об ортогональном проектировании точки на подпространство и о наилучшем приближении прямоугольной матрицы матрицами меньшего ранга. Последняя задача имеет прямое отношение к проблеме сжатия информации.

Книга ориентирована на инженеров и студентов технических вузов, желающих повысить свою математическую квалификацию.

Библиогр. 9 назв. Ил. 5.

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Эта небольшая книга написана для инженеров и студентов технических вузов, заинтересованных в повышении своей математической квалификации. В ней представлены те начальные сведения из линейной алгебры, которые, по мнению автора, наиболее часто используются в приложениях.

К базовым понятиям линейной алгебры относится понятие линейной независимости векторов и связанное с ним понятие ранга матрицы. В главе 1 показано, как далеко можно продвинуться, опираясь только на эти понятия. Итог получается в определенном смысле неожиданный. Удается ввести и исследовать понятие обратной матрицы, установить критерий совместности системы линейных уравнений (в форме условия ортогональности вектора, стоящего в правой части системы, любому решению однородной сопряженной системы), описать эффективный алгоритм вычисления ранга прямоугольной матрицы и попутно построить для нее QR-разложение. Ведущую роль в главе 1 играет утверждение о линейной зависимости линейных комбинаций векторов. Оно названо основной леммой линейной алгебры.

В главе 2 развивается вариационный подход к задаче о собственных числах и собственных векторах симметричной матрицы. Предварительно вводится квадратичная функция и подробно изучаются ее свойства. Доказаны общие теоремы о существовании и характеризации точки минимума выпуклой квадратичной функции на аффинном многообразии. Доказательства достаточно элементарны. Они опираются на теорему о совместности системы линейных уравнений. С использованием общих теорем получено спектральное разложение симметричной матрицы и, как следствие, сингулярное разложение произвольной прямоугольной матрицы.

Рассматриваются две конкретные экстремальные задачи: об ортогональном проектировании точки на подпространство и о наилучшем приближении прямоугольной матрицы матрицами меньшего ранга. Решение первой задачи проводится в рамках стандартной схемы. Указана явная формула для проекции, детально исследованы свойства матрицы ортогонального проектирования. Вторая задача имеет прямое отношение к проблеме сжатия информации. В книге дается изящное решение этой задачи, опирающееся, по существу, на все предыдущие результаты. Отмечается, как связана матрица наилучшего приближения с сингулярным разложением исходной матрицы.

Следует еще раз подчеркнуть, что книга содержит лишь начальные сведения из линейной алгебры. Впрочем, для многих читателей их будет вполне достаточно. Расширить свои знания (прежде всего за счет теории определителей) можно, обращаясь к более полным учебным пособиям. Назовем два из них:

- Воеводин В.В. Линейная алгебра. Изд.2-е. М., 1980. 400 с.
- Беклемишев Д.В. Дополнительные главы линейной алгебры. М., 1983. 335 с.

Автор глубоко признателен друзьям и коллегам В.А. Даугавет, А.Б. Певному, И.В. Романовскому, П.В. Яковлеву, которые своими советами, замечаниями и участием помогали в работе над книгой.

ГЛАВА 1. ЛИНЕЙНАЯ АЛГЕБРА БЕЗ ОПРЕДЕЛИТЕЛЕЙ

Основные обозначения

$\mathcal{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ — множество целых чисел;

$m{:}n$ — множество целых чисел от m до n включительно, т.е. $m{:}n = \{m, m+1, \ldots, n\}$; здесь m и n — целые числа;

$M, N, \ldots$ — произвольные индексные множества (конечные наборы целых чисел);

$|M|$ — количество элементов, содержащихся в M;

$\mathcal{R}$ — множество вещественных чисел;

$\mathcal{R}^N$ — линейное пространство векторов $x = x[N]$ с компонентами $x[j]$, $j \in N$;

$e_j = e_j[N]$, где $j \in N$, — j-й орт (вектор, у которого $e_j[j] = 1$ и $e_j[k] = 0$ при $k \neq j$);

$\mathbf{0} = \mathbf{0}[N]$ — нулевой вектор;

$\langle x, y\rangle = x[N] \times y[N] = \sum_{j \in N} x[j] \times y[j]$ — скалярное произведение векторов x и y;

$A = A[M, N]$ — матрица с элементами $A[i, j]$, $i \in M$, $j \in N$;

$A^\top = A^\top[N, M]$ — транспонированная матрица, у которой

$$A^\top[j, i] = A[i, j], \quad j \in N, \quad i \in M;$$

$A[M_1, N_1]$, где $M_1 \subset M$, $N_1 \subset N$, — подматрица матрицы $A[M, N]$;

$A[M, j]$ — j-й столбец матрицы $A[M, N]$;

$A[i, N]$ — i-я строка матрицы $A[M, N]$;

$v = Ax = A[M, N] \times x[N]$ — вектор с компонентами $v[i] = A[i, N] \times x[N]$, $i \in M$;

$y = uA = u[M] \times A[M, N]$ — вектор с компонентами $y[j] = u[M] \times A[M, j]$, $j \in N$;

$C = AB = A[M, N] \times B[N, Q]$ — произведение матриц A и B (матрица с элементами $C[i, k] = A[i, N] \times B[N, k]$, $i \in M$, $k \in Q$);

$E = E[N, N]$ — единичная матрица, у которой $E[j, j] = 1$ при всех $j \in N$ и $E[j, k] = 0$ при $j \neq k$;

$\mathcal{R}^{M,N}$ — линейное пространство матриц $A = A[M, N]$.

Когда мы говорим об $\mathcal{R}^{M,N}$ как о линейном пространстве, то имеем в виду, что вместе с матрицами A_1, A_2 ему принадлежит матрица $A = \alpha_1 A_1 + \alpha_2 A_2$ с элементами

$$A[i, j] = \alpha_1 A_1[i, j] + \alpha_2 A_2[i, j],$$

где α_1, α_2 — произвольные вещественные числа.

Вектор, в зависимости от обстоятельств, можно рассматривать либо как матрицу, состоящую из одного столбца, либо как матрицу, состоящую из одной строки.

Равенство матриц $A_1[M, N] = A_2[M, N]$ понимается как поэлементное равенство: $A_1[i, j] = A_2[i, j]$ при всех $i \in M$, $j \in N$.

1.1. Свойства векторов и матриц (индексная техника)

Приведем основные свойства векторов и матриц.

1. Пусть $N_1 \subset N$, $N_2 = N \setminus N_1$. Тогда

$$c[N] \times x[N] = c[N_1] \times x[N_1] + c[N_2] \times x[N_2].$$

Это следует из определения скалярного произведения.

2. Если $N_1 \subset N$ и $N_2 = N \setminus N_1$, то

$$A[M, N] \times x[N] = A[M, N_1] \times x[N_1] + A[M, N_2] \times x[N_2]. \tag{1.1}$$

Аналогично, если $M_1 \subset M$ и $M_2 = M \setminus M_1$, то

$$u[M] \times A[M, N] = u[M_1] \times A[M_1, N] + u[M_2] \times A[M_2, N]. \tag{1.2}$$

Проверим, например, равенство (1.1). При любом $i \in M$ имеем

$$A[i, N] \times x[N] = \sum\nolimits_{j \in N} A[i, j] \times x[j] = \left(\sum\nolimits_{j \in N_1} + \sum\nolimits_{j \in N_2}\right) A[i, j] \times x[j] =$$
$$= A[i, N_1] \times x[N_1] + A[i, N_2] \times x[N_2].$$

Таким образом, у векторов, стоящих в левой и правой частях равенства (1.1), все компоненты равны. По определению, равны и сами эти векторы.

Аналогично проверяется равенство (1.2).

Если учесть, что индексное множество есть объединение всех его элементов, то в качестве следствия из (1.1) и (1.2) получаем важные формулы

$$Ax = \sum\nolimits_{j \in N} A[M, j] \times x[j], \tag{1.3}$$

$$uA = \sum\nolimits_{i \in M} u[i] \times A[i, N]. \tag{1.4}$$

Они означают, что вектор Ax равен линейной комбинации столбцов $A[M, j]$ матрицы A с коэффициентами $x[j]$, а вектор uA — линейной комбинации строк $A[i, N]$ матрицы A с коэффициентами $u[i]$.

3. Справедливо равенство

$$u[M] \times (A[M, N] \times x[N]) = (u[M] \times A[M, N]) \times x[N],$$

которое коротко можно записать так:

$$\langle u, Ax\rangle = \langle uA, x\rangle. \tag{1.5}$$

Действительно,

$$\begin{aligned}\langle u, Ax\rangle &= \sum\nolimits_{i\in M} u[i] \times (A[i, N] \times x[N]) = \\ &= \sum\nolimits_{i\in M} u[i] \times \left(\sum\nolimits_{j\in N} A[i, j] \times x[j]\right) = \\ &= \sum\nolimits_{i\in M} \left(\sum\nolimits_{j\in N} u[i] \times A[i, j] \times x[j]\right),\end{aligned}$$

$$\begin{aligned}\langle uA, x\rangle &= \sum\nolimits_{j\in N} (u[M] \times A[M, j]) \times x[j]) = \\ &= \sum\nolimits_{j\in N} \left(\sum\nolimits_{i\in M} u[i] \times A[i, j]\right) \times x[j] = \\ &= \sum\nolimits_{j\in N} \left(\sum\nolimits_{i\in M} u[i] \times A[i, j] \times x[j]\right).\end{aligned}$$

В правых частях этих равенств стоят повторные суммы, различающиеся только порядком суммирования. Они равны. Значит, равны и левые части.

Отметим, что векторы uA и $A^\top u$ имеют одинаковые компоненты. Поэтому $\langle uA, x\rangle = \langle A^\top u, x\rangle$. Равенство (1.5) можно переписать в виде

$$\langle u, Ax\rangle = \langle A^\top u, x\rangle. \tag{1.6}$$

4. Справедливо равенство

$$C[P, M] \times (A[M, N] \times B[N, Q]) = (C[P, M] \times A[M, N]) \times B[N, Q],$$

которое коротко можно записать так:

$$C(AB) = (CA)B. \tag{1.7}$$

Доказательство. Поскольку j-й столбец матрицы AB равен произведению $A[M, N] \times B[N, j]$, то для элемента с индексами (i, j) матрицы $C(AB)$ получаем представление

$$C[i, M] \times (A[M, N] \times B[N, j]).$$

Аналогично записывается выражение для элемента с индексами (i,j) матрицы $(CA)B$:

$$(C[i,M] \times A[M,N]) \times B[N,j].$$

Согласно (1.5) эти элементы равны при всех $i \in M$ и $j \in N$. По определению, равны и сами матрицы $C(AB)$ и $(CA)B$.

Формула (1.7) характеризует *ассоциативность* операции умножения матриц.

5. Справедливо равенство

$$A[M,N] \times B[N,Q] = \sum\nolimits_{j \in N} A[M,j] \times B[j,Q]. \qquad (1.8)$$

Оно проверяется непосредственным сравнением элементов с индексами (i,k) матриц, стоящих в левой и правой частях (1.8).

При всей своей простоте формула (1.8) весьма содержательна. Она указывает на то, что произведение AB можно представить в виде суммы матриц, каждая из которых является произведением столбца матрицы A на соответствующую строку матрицы B.

ЗАДАЧИ

1.1. Пусть $x, x_1, x_2, \ldots, x_p$ — векторы из $\mathcal{R}^N$ и

$$x = \sum_{k=1}^{p} \alpha_k x_k,$$

где $\alpha_1, \ldots, \alpha_p$ — вещественные числа. Докажите, что для матрицы $A \in \mathcal{R}^{M,N}$ справедливо равенство

$$Ax = \sum_{k=1}^{p} \alpha_k (Ax_k).$$

1.2. Пусть $A, A_1, A_2, \ldots, A_s$ — матрицы из $\mathcal{R}^{M,N}$ и

$$A = \sum_{k=1}^{s} \beta_k A_k,$$

где $\beta_1, \ldots, \beta_s$ — вещественные числа. Докажите, что для матриц $B \in \mathcal{R}^{N,Q}$ и $C \in \mathcal{R}^{P,M}$ справедливы равенства

$$AB = \sum_{k=1}^{s} \beta_k (A_k B), \qquad CA = \sum_{k=1}^{s} \beta_k (CA_k).$$

1.3. Пусть $C[M,Q] = A[M,N] \times B[N,Q]$. Докажите, что $C^\top = B^\top A^\top$.

1.4. Пусть $N_1 \subset N$. Докажите, что

$$E[N_1, N] \times x[N] = x[N_1], \qquad x[N] \times E[N, N_1] = x[N_1].$$

1.2. Основная лемма линейной алгебры

К фундаментальным понятиям линейной алгебры относится понятие линейной независимости векторов.

Определение. Векторы $x_1, x_2, \ldots, x_p$ из $\mathcal{R}^N$ называются *линейно независимыми*, если равенство

$$\sum_{k=1}^{p} \alpha_k x_k = \mathbf{0} \tag{2.1}$$

справедливо только тогда, когда все коэффициенты α_k равны нулю.

Векторы $x_1, x_2, \ldots, x_p$ *линейно зависимы*, если существуют числа $\alpha_1, \alpha_2, \ldots, \alpha_p$, не все равные нулю, такие, что выполняется (2.1). Пусть, например, $\alpha_1 \neq 0$. Тогда

$$x_1 = -\sum_{k=2}^{p} \left(\frac{\alpha_k}{\alpha_1}\right) x_k.$$

Это означает, что в случае линейной зависимости один из векторов $x_1, x_2, \ldots, x_p$ можно представить в виде линейной комбинации остальных.

При $p = 1$ вектор x_1 линейно независим лишь тогда, когда он отличен от нулевого.

Наличие нулевого вектора в системе $x_1, x_2, \ldots, x_p$ делает эту систему линейно зависимой.

Упражнение. Пусть заданы векторы $x_1, x_2, \ldots, x_p, x_{p+1}, \ldots, x_{p+r}$. Докажите утверждения:

- Если система векторов $x_1, x_2, \ldots, x_p$ линейно зависима, то линейно зависима и расширенная система $x_1, x_2, \ldots, x_{p+r}$.

- Если система векторов $x_1, x_2, \ldots, x_{p+r}$ линейно независима, то линейно независима и подсистема $x_1, x_2, \ldots, x_p$.

ОСНОВНАЯ ЛЕММА ЛИНЕЙНОЙ АЛГЕБРЫ. *Допустим, что векторы $y_1, y_2, \ldots, y_q$ являются линейными комбинациями векторов $x_1, x_2, \ldots, x_p$ из $\mathcal{R}^N$, т.е.*

$$y_1 = c_{11}x_1 + c_{12}x_2 + \cdots + c_{1p}x_p,$$

$$y_2 = c_{21}x_1 + c_{22}x_2 + \cdots + c_{2p}x_p,$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \tag{2.2}$$

$$y_q = c_{q1}x_1 + c_{q2}x_2 + \dots + c_{qp}x_p.$$

Тогда при $q > p$ *векторы* $y_1,\ y_2,\ \dots,\ y_q$ *линейно зависимы.*

Д о к а з а т е л ь с т в о проводится индукцией по p. При $p = 1$ условие леммы можно представить в виде

$$y_k = c_k x_1, \quad k \in 1\colon q; \quad q \geq 2. \tag{2.3}$$

Покажем, что векторы y_1, y_2 линейно зависимы. Если $c_1 = 0$, то $y_1 = \mathbf{0}$. В этом случае линейная зависимость векторов y_1, y_2 очевидна. Пусть $c_1 \neq 0$. Согласно (2.3)

$$y_2 - \left(\frac{c_2}{c_1}\right) y_1 = \mathbf{0}.$$

Коэффициент при y_2 равен 1. Значит, векторы y_1, y_2 линейно зависимы. Линейно зависима и вся система $y_1, y_2, \dots, y_q$.

Сделаем индукционный переход от $p - 1$ к p. Рассмотрим соотношения (2.2) при $q > p$. Возможны два случая.

1. Все коэффициенты $c_{11}, c_{21}, \dots, c_{q1}$ при x_1 равны нулю. Тогда $y_1, y_2, \dots, y_q$ суть линейные комбинации $x_2, \dots, x_p$. По индукционному предположению векторы $y_1, y_2, \dots, y_q$ линейно зависимы.

2. Среди коэффициентов $c_{11}, c_{21}, \dots, c_{q1}$ имеется ненулевой. Пусть это будет c_{11}. Введем векторы

$$\begin{aligned} z_2 &= y_2 - \tfrac{c_{21}}{c_{11}}\, y_1, \\ &\dots\dots\dots\dots\dots \\ z_q &= y_q - \tfrac{c_{q1}}{c_{11}}\, y_1. \end{aligned} \tag{2.4}$$

   Эти векторы являются линейными комбинациями $x_2, \dots, x_p$, и их количество $q-1$ больше $p-1$. По индукционному предположению они линейно зависимы. Это значит, что найдутся числа $\gamma_2, \dots, \gamma_q$, не все равные нулю, такие, что

$$\gamma_2 z_2 + \ldots + \gamma_q z_q = \mathbf{0}.$$

   Согласно (2.4) $\gamma_1 y_1 + \gamma_2 y_2 + \cdots \gamma_q y_q = \mathbf{0}$, где

$$\gamma_1 = -\frac{c_{21}}{c_{11}}\gamma_2 - \dots - \frac{c_{q1}}{c_{11}}\gamma_q.$$

   Коэффициенты $\gamma_1, \gamma_2, \dots, \gamma_q$ не все равны нулю. Следовательно, векторы $y_1, y_2, \dots, y_q$ линейно зависимы. ■

Отметим, что лемма остается справедливой, если $\mathcal{R}^N$ заменить любым линейным пространством. Однако, здесь она будет использоваться только для векторов из $\mathcal{R}^N$.

Обозначим $n = |N|$.

ТЕОРЕМА 2.1. *В линейном пространстве* $\mathcal{R}^N$ *наибольшее число линейно независимых векторов равно* n.

Доказательство. Прежде всего предъявим n линейно независимых векторов в $\mathcal{R}^N$. Это орты $e_j = e_j[N]$, $j \in N$, у которых $e_j[j] = 1$ и $e_j[k] = 0$ при $k \neq j$. Их линейная независимость проверяется так. Предположим, что

$$\sum\nolimits_{j\in N} \alpha_j\; e_j = \mathbf{0}. \tag{2.5}$$

Компонента с индексом k вектора, стоящего в левой части (2.5), равна $\sum_{j\in N} \alpha_j e_j[k] = \alpha_k$. Согласно (2.5), $\alpha_k = 0$ при всех $k \in N$. Линейная независимость ортов e_j, $j \in N$, установлена.

Остается проверить, что любые $n+1$ векторов $y_1, y_2, \ldots, y_{n+1}$ из $\mathcal{R}^N$ линейно зависимы. Это следует из представления

$$y_s = \sum\nolimits_{j\in N} y_s\,[j]\; e_j, \qquad s \in 1\colon n+1,$$

и основной леммы линейной алгебры. ∎

Наибольшее число линейно независимых векторов, принадлежащих линейному пространству X, называется *размерностью* X и обозначается $\dim X$. Согласно теореме 2.1 $\dim \mathcal{R}^N = |N|$.

Орты $\{e_j\}$ образуют *базис* пространства $\mathcal{R}^N$ в том смысле, что любой вектор $x \in \mathcal{R}^N$ можно единственным образом представить в виде их линейной комбинации:

$$x = \sum\nolimits_{j\in N} x[j]\; e_j.$$

Базис $\{e_j\}$ не единственный.

ТЕОРЕМА 2.2. *Любые* n *линейно независимых векторов* $u_1, u_2, \ldots, u_n$ *из* $\mathcal{R}^N$ *образуют базис* $\mathcal{R}^N$.

Доказательство. Проверим, что любой вектор $x \in \mathcal{R}^N$ можно единственным образом представить в виде

$$x = \sum_{j=1}^{n} \alpha_j u_j. \tag{2.6}$$

Векторы $x, u_1, \ldots, u_n$ принадлежат $\mathcal{R}^N$, и их количество равно $n+1$. Согласно теореме 2.1 они линейно зависимы. Это значит, что найдутся числа $\gamma, \gamma_1, \ldots, \gamma_n$, не все равные нулю, такие, что

$$\gamma x + \gamma_1 u_1 + \ldots + \gamma_n u_n = \mathbf{0}. \tag{2.7}$$

Коэффициент γ отличен от нуля. В противном случае, в силу линейной независимости векторов $u_1, \ldots, u_n$, и остальные коэффициенты $\gamma_1, \ldots, \gamma_n$ равнялись бы нулю. Но это противоречит предположению.

Из (2.7) получаем

$$x = -\sum_{j=1}^{n} \left(\frac{\gamma_j}{\gamma}\right) u_j,$$

что соответствует (2.6) с $\alpha_j = -\frac{\gamma_j}{\gamma}$.

Установим единственность представления (2.6). Допустим, что имеется еще одно представление

$$x = \sum_{j=1}^{n} \beta_j u_j. \tag{2.8}$$

Вычитая (2.8) из (2.6), приходим к равенству

$$\mathbf{0} = \sum_{j=1}^{n} (\alpha_j - \beta_j)\ u_j.$$

Из него следует в силу линейной независимости векторов $u_1, \ldots, u_n$, что $\alpha_j = \beta_j$ при всех $j \in 1\colon n$. ■

Отметим, что число векторов, образующих базис пространства $\mathcal{R}^N$, совпадает с размерностью $\mathcal{R}^N$.

ЗАДАЧИ

2.1. Пусть N_0 является собственным подмножеством индексного множества N. Докажите утверждения:

- Если векторы $x_1[N], x_2[N], \ldots, x_p[N]$ линейно зависимы, то линейно зависимы и векторы $x_1[N_0], x_2[N_0], \ldots, x_p[N_0]$.
- Если векторы $x_1[N_0], x_2[N_0], \ldots, x_p[N_0]$ линейно независимы, то линейно независимы и векторы $x_1[N], x_2[N], \ldots, x_p[N]$.

2.2. Предположим, что векторы $x_1, x_2, \ldots, x_p$, принадлежащие $\mathcal{R}^N$, линейно независимы и любой вектор x из $\mathcal{R}^N$ можно представить в виде их линейной комбинации. Докажите, что $p = |N|$.

2.3. Докажите, что систему линейно независимых векторов x_1, x_2, $\ldots$, x_p из $\mathcal{R}^N$ при $p < n$ можно дополнить векторами $x_{p+1}, \ldots, x_n$ так, что система $x_1, x_2, \ldots, x_n$ будет линейно независимой.

1.3. Ранг матрицы

В этом параграфе будем считать для определенности, что $N = \{1, 2, \ldots, n\}$, или в других обозначениях $N = 1\colon n$.

ЛЕММА 3.1. *Если столбцы квадратной матрицы* $A = A[N,N]$ *линейно независимы, то линейно независимы и ее строки.*

Доказательство. Допустим, вопреки утверждению, что строки матрицы A линейно зависимы. Тогда одну из этих строк можно представить в виде линейной комбинации остальных. Например,

$$A[1,N] = \sum_{i=2}^{n} \beta_i \, A\, [i,N]. \tag{3.1}$$

Матрица A без первой строки имеет $n-1$ строк и n столбцов. Поскольку количество столбцов больше их размерности, то по теореме 2.1 столбцы матрицы $A[2\colon n, 1\colon n]$ линейно зависимы. Значит, найдутся числа $\gamma_1, \dots, \gamma_n$, не все равные нулю, такие, что

$$\sum_{j=1}^{n} \gamma_j \, A\, [i,j] = 0, \qquad i \in 2\colon n. \tag{3.2}$$

Вместе с тем согласно (3.1)

$$\sum_{j=1}^{n} \gamma_j A[1,j] = \sum_{j=1}^{n} \gamma_j \sum_{i=2}^{n} \beta_i A[i,j] = \sum_{i=2}^{n} \beta_i \sum_{j=1}^{n} \gamma_j A[i,j] = 0.$$

Получили, что соотношение (3.2) выполняется при всех $i \in 1\colon n$. Это противоречит линейной независимости столбцов матрицы A. ■

Отметим, что лемма 3.1 остается справедливой, если матрицу $A[N,N]$ заменить матрицей $A[M,N]$ с $|M| = |N|$.

Рассмотрим произвольную ненулевую матрицу $A = A[M,N]$, где $M = 1\colon m$, $N = 1\colon n$. Обозначим через q *наибольшее* число линейно независимых столбцов этой матрицы. Систему, состоящую из q линейно независимых столбцов, назовем *базисом столбцов* матрицы A. Согласно определению q добавление к базисным столбцам небазисного делает систему линейно зависимой. Отсюда следует, что любой небазисный столбец можно представить в виде линейной комбинации базисных.

Обозначим через p *наибольшее* число линейно независимых строк матрицы A. Систему, состоящую из p линейно независимых строк, назовем *базисом строк*. Отметим, что любую небазисную строку можно представить в виде линейной комбинации базисных.

ТЕОРЕМА 3.1. *У произвольной ненулевой матрицы* $A = A[M,N]$ *наибольшее число* q *линейно независимых столбцов и наибольшее число* p *линейно независимых строк совпадают.*

Это общее значение называется *рангом* матрицы A и обозначается $\operatorname{rank} A$.

Доказательство. Пусть $\{A[M,j]\}$, $j \in N_*$, $|N_*| = q$, — базис столбцов и $\{A[i,N]\}$, $i \in M_*$, $|M_*| = p$, — базис строк матрицы A.

Предположим сначала, что $p < q$. Выделим подматрицу $A[M_*, N_*]$. У нее строк меньше, чем столбцов. По теореме 2.1 столбцы этой матрицы линейно зависимы. Значит, найдутся числа γ_j, $j \in N_*$, не все равные нулю, такие, что

$$\sum\nolimits_{j\in N_*} \gamma_j\, A\, [i,j] = 0, \qquad i \in M_*. \tag{3.3}$$

Зафиксируем $i_0 \in M \setminus M_*$. В силу определения базиса строк

$$A[i_0, N] = \sum\nolimits_{i\in M_*} \beta_i\, A\, [i, N].$$

Отсюда и из (3.3) следует, что

$$\sum_{j\in N_*} \gamma_j A[i_0, j] = \sum_{j\in N_*} \gamma_j \sum_{i\in M_*} \beta_i A[i,j] = \sum_{i\in M_*} \beta_i \sum_{j\in N_*} \gamma_j A[i,j] = 0.$$

Получили, что соотношение (3.3) выполняется при всех $i \in M$. Но это противоречит линейной независимости столбцов $A[M, j]$, $j \in N_*$.

Установлено, что $p \geq q$. Аналогично проверяется неравенство $q \geq p$. В результате приходим к равенству $p = q$. ■

В теореме рассматривались ненулевые матрицы. У нулевой матрицы ранг равен нулю по определению.

ТЕОРЕМА 3.2. *Пусть* $A = A[M, N]$, $B = B[N, Q]$ *и* $C = AB$. *Тогда*

$$\operatorname{rank} C \leq \min\{\operatorname{rank} A, \operatorname{rank} B\}. \tag{3.4}$$

Доказательство. Если C — нулевая матрица, то неравенство тривиально. Поэтому будем считать, что C, а значит, A и B — ненулевые матрицы.

Согласно (1.3), k-й столбец матрицы C, равный $A[M, N] \times B[N, k]$, можно представить в виде линейной комбинации столбцов матрицы A, точнее, в виде линейной комбинации базисных столбцов матрицы A в количестве $\operatorname{rank} A$. По основной лемме линейной алгебры число линейно независимых столбцов матрицы C не больше $\operatorname{rank} A$. Значит, $\operatorname{rank} C \leq \operatorname{rank} A$.

Далее, согласно (1.4), i-ю строку матрицы C, равную $A[i, N] \times B[N, Q]$, можно представить в виде линейной комбинации базисных строк матрицы B в количестве $\operatorname{rank} B$. По основной лемме линейной алгебры $\operatorname{rank} C \leq \operatorname{rank} B$. Объединяя два полученных неравенства, приходим к (3.4). ■

Отметим, что неравенство (3.4) может выполняться как строгое. Пусть, например,

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

В этом случае $C = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $\text{rank } C = 0$, в то время как $\text{rank } A = \text{rank } B = 2$.

ТЕОРЕМА 3.3. *Для суммы матриц* $A = A[M,N]$ *и* $B = B[M,N]$ *справедливо неравенство*

$$\text{rank } (A+B) \leq \text{rank } A + \text{rank } B. \tag{3.5}$$

Доказательство. Можно считать, что обе матрицы A и B ненулевые. Обозначим через $A_1, \ldots, A_n$ и $B_1, \ldots, B_n$ столбцы матриц A и B соответственно. Совокупность столбцов $A_1, \ldots, A_n$, $B_1, \ldots, B_n$ образует матрицу D. Согласно основной лемме линейной алгебры
$\text{rank } (A+B) \leq \text{rank } D$, поскольку каждый столбец $A_j + B_j$ матрицы $A+B$ можно представить в виде линейной комбинации базисных столбцов матрицы D.

Вместе с тем ясно, что $\text{rank } D \leq \text{rank } A + \text{rank } B$. Окончательно получаем

$$\text{rank } (A+B) \leq \text{rank } D \leq \text{rank } A + \text{rank } B. \quad \blacksquare$$

Простейший пример

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

показывает, что неравенство (3.5) может обращаться в равенство.

ЗАДАЧА

3.1. Квадратная матрица $A = A[N,N]$ называется *диагональной*, если $A[i,j] = 0$ при $i \neq j$. Докажите, что ранг диагональной матрицы равен числу ненулевых диагональных элементов $A[i,i]$.

1.4. Обратная матрица

Квадратная матрица $A = A[N,N]$ называется *обратимой*, если существует матрица $B = B[N,N]$, такая, что

$$AB = E, \quad BA = E,$$

где $E = E[N,N]$ — единичная матрица. Матрица B называется *обратной* и обозначается $B = A^{-1}$.

Обратная матрица, если она существует, единственна. Действительно, пусть нашлась еще одна матрица B_0, такая, что $AB_0 = E, B_0A = E$. Умножая равенство $AB = E$ слева на B_0 и пользуясь ассоциативностью операции умножения матриц, получаем $(B_0A)B = B_0$ или $B = B_0$.

Непосредственно из определения следует, что

- единичная матрица E обратима и $E^{-1} = E$;
- если матрица A обратима, то обратима и обратная матрица A^{-1}; при этом $(A^{-1})^{-1} = A$.

ТЕОРЕМА 4.1. *Для того чтобы матрица $A = A[N, N]$ была обратимой, необходимо и достаточно, чтобы ее столбцы были линейно независимы.*

Доказательство. Необходимость. Пусть матрица A обратима. Тогда $AA^{-1} = E$. Согласно теореме 3.2

$$\text{rank } E \le \min\{\text{rank } A,\ \text{rank } A^{-1}\}. \tag{4.1}$$

По определению ранга, $\text{rank } A \le |N|$, $\text{rank } A^{-1} \le |N|$. Кроме того, $\text{rank } E = |N|$. Неравенство (4.1) будет непротиворечивым только тогда, когда $\text{rank } A = |N|$, $\text{rank } A^{-1} = |N|$. Последнее означает, что столбцы матриц A и A^{-1} линейно независимы.

Достаточность. Предположим, что столбцы матрицы A линейно независимы. Согласно теореме 2.2 они образуют базис пространства $\mathcal{R}^N$.

Рассмотрим матричное уравнение $AB = E$ и перепишем его в виде равенства столбцов $AB_j = e_j$, $j \in N$. Здесь e_j — j-й орт в $\mathcal{R}^N$ и B_j — j-й столбец неизвестной матрицы B. Каждое из последних уравнений имеет единственное решение. В качестве компонент вектора B_j нужно взять коэффициенты разложения орта e_j по базису, составленному из столбцов матрицы A. Тем самым найдено единственное решение B уравнения $AB = E$.

Покажем, что равенство $BA = E$ выполняется автоматически. Как отмечалось при доказательстве необходимости, из условия $AB = E$ следует, что $\text{rank } B = |N|$. Это значит, что столбцы B_j матрицы B линейно независимы и образуют базис пространства $\mathcal{R}^N$. В частности,

$$e_k = \sum_{j \in N} \alpha_{kj} B_j, \quad k \in N. \tag{4.2}$$

Умножим равенство $AB = E$ слева на B. Получим $(BA)B = B$, или $(BA)B_j = B_j$ при всех $j \in N$. На основании последних равенств и (4.2) заключаем, что при всех $k \in N$

$$(BA)e_k = \sum_{j \in N} \alpha_{kj} (BA) B_j = \sum_{j \in N} \alpha_{kj} B_j = e_k.$$

Перепишем этот результат в матричной форме: $(BA)E = E$. Равенство $BA = E$, а с ним и теорема, доказаны. ■

ТЕОРЕМА 4.2. *Пусть у прямоугольной матрицы $A = A[M, N]$ столбцы линейно независимы. Тогда матрица $A^\top A$ обратима.*

Доказательство. Рассмотрим линейную комбинацию $(A^\top A)u$ столбцов матрицы $A^\top A$ и предположим, что $(A^\top A)u = \mathbf{0}$. Покажем, что $u = \mathbf{0}$. Это будет означать, что столбцы матрицы $A^\top A$ линейно независимы.

Согласно (1.6) имеем

$$0 = \langle (A^\top A)u, u\rangle = \langle Au, Au\rangle = \sum_{i\in M}((Au)[i])^2.$$

Отсюда следует, что $Au = \mathbf{0}$. По условию столбцы матрицы A линейно независимы. Значит, $u = \mathbf{0}$.

Установлена линейная независимость столбцов квадратной матрицы $A^\top A$. По теореме 4.1 матрица $A^\top A$ обратима. ■

ТЕОРЕМА 4.3. *Если матрица $A = A[M, N]$ имеет ранг r, то ранг матриц $A^\top A$ и $AA^\top$ также равен r.*

Доказательство. Можно считать, что $r \geq 1$. Обозначим через B матрицу, составленную из базисных столбцов матрицы A, и пусть $G = B^\top B$. По теореме 4.2 матрица G обратима, а согласно теореме 4.1 ее столбцы линейно независимы. Тем более линейно независимыми будут соответствующие столбцы матрицы $A^\top A$ (рис. 1).

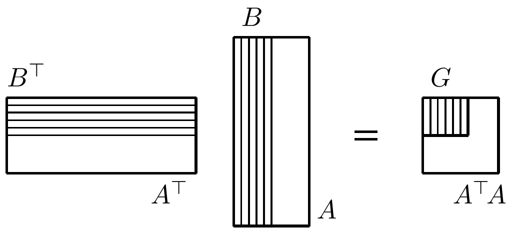

Рисунок 1.

Этих столбцов r штук. Значит, $\operatorname{rank} A^\top A \geq r$. Вместе с тем по теореме 3.2 $\operatorname{rank} A^\top A \leq r$. Приходим к равенству $\operatorname{rank} A^\top A = r$.

Поскольку $AA^\top = (A^\top)^\top A^\top$, то $\operatorname{rank} AA^\top = \operatorname{rank} A^\top = r$. Теорема доказана. ■

ТЕОРЕМА 4.4. *Даны матрицы $A = A[M, N]$, $B = B[M, M]$, $C = C[N, N]$, причем известно, что матрицы B и C обратимы. Тогда* $\operatorname{rank} BAC = \operatorname{rank} A$.

Доказательство. Согласно теореме 3.2

$$\operatorname{rank} A \geq \operatorname{rank} AC \geq \operatorname{rank} ACC^{-1} = \operatorname{rank} A.$$

Значит, $\operatorname{rank} AC = \operatorname{rank} A$. Далее

$$\operatorname{rank} AC \geq \operatorname{rank} BAC \geq \operatorname{rank} B^{-1}BAC = \operatorname{rank} AC.$$

Значит, $\operatorname{rank} BAC = \operatorname{rank} AC$. Имеем

$$\operatorname{rank} BAC = \operatorname{rank} AC = \operatorname{rank} A.$$

Теорема доказана. ■

Понятие обратной матрицы можно ввести не только для матрицы $A[N, N]$, но и для матрицы $A[M, N]$ при условии, что $|M| = |N|$.

Квадратная матрица $A[M, N]$ с $|M| = |N|$ называется обратимой, если существует матрица $B[N, M]$, такая, что $A[M, N] \times B[N, M] = E[M, M]$ и $B[N, M] \times A[M, N] = E[N, N]$. Матрица $B[N, M]$ называется обратной и обозначается $B[N, M] = A^{-1}[N, M]$.

ЗАДАЧИ

4.1. Пусть матрица A обратима. Докажите, что транспонированная матрица $A^\top$ также обратима и $(A^\top)^{-1} = (A^{-1})^\top$.

4.2. Известно, что матрицы $A = A[N, N]$ и $B = B[N, N]$ обратимы. Докажите, что обратимо и произведение AB; при этом $(AB)^{-1} = B^{-1}A^{-1}$.

4.3. Пусть $C[N, N] = A[N, N] \times B[N, N]$. Докажите, что из обратимости матрицы C следует обратимость матриц A и B.

4.4. Матрица $A = A[1 : n, 1 : n]$ называется *верхней треугольной*, если $A[i, j] = 0$ при $i > j$. Докажите, что верхняя треугольная матрица A обратима тогда и только тогда, когда все ее диагональные элементы $A[i, i]$ отличны от нуля.

4.5. Обозначим через M_*, N_* множества индексов базисных строк и базисных столбцов прямоугольной матрицы $A[M, N]$. Докажите, что квадратная матрица $A[M_*, N_*]$ обратима.

4.6. Пусть $A = A[N, N]$ — обратимая матрица и $B = xy^\top$, где $x = x[N, 1]$ — вектор-столбец, $y^\top = y^\top[1, N]$ — вектор-строка. Докажите формулу Шермана–Моррисона

$$(A + B)^{-1} = A^{-1} - \beta A^{-1}BA^{-1},$$

где $\beta = (1 + \alpha)^{-1}$, $\alpha = \langle y, A^{-1}x \rangle$, при условии, что $\alpha \neq -1$.

4.7. Пусть $A = \begin{bmatrix} B & C \\ D & H \end{bmatrix}$, где B и H — квадратные матрицы, причем B обратима. Обозначим $F = H - DB^{-1}C$. Выведите формулу Фробениуса: если существует F^{-1}, то

$$A^{-1} = \begin{bmatrix} B^{-1} + B^{-1}CF^{-1}DB^{-1} & -B^{-1}CF^{-1} \\ -F^{-1}DB^{-1} & F^{-1} \end{bmatrix}.$$

1.5. Скелетное разложение матрицы

ТЕОРЕМА 5.1. *Для того чтобы ненулевая матрица* $A = A[M,N]$ *имела ранг* r, *необходимо и достаточно, чтобы существовало представление*

$$A[M,N] = X[M,R] \times Y^\top[R,N], \tag{5.1}$$

где $|R| = r$ *и* $\operatorname{rank} X = \operatorname{rank} Y^\top = r$.

Доказательство. Необходимость. Пусть матрица A имеет ранг r. В качестве X возьмем матрицу, составленную из базисных столбцов матрицы A. Тогда коэффициенты разложения k-го столбца матрицы A по базису ее столбцов определят k-й столбец матрицы $Y^\top$.

Ранг X по определению равен r. Матрица $Y^\top$ содержит единичную подматрицу размером $r \times r$, поэтому и ее ранг равен r.

Достаточность. Допустим, что для матрицы $A = A[M,N]$ существует представление (5.1), в котором матрицы $X[M,R]$, $Y[N,R]$ имеют ранг $r = |R|$. По теореме 3.2 $\operatorname{rank} A \le r$.

Возьмем r линейно независимых столбцов матрицы $Y^\top$. Они образуют квадратную обратимую подматрицу B. По теореме 4.4 $\operatorname{rank} XB = \operatorname{rank} X = r$, т.е. все r столбцов матрицы XB линейно независимы. Эти столбцы являются также столбцами матрицы $A = XY^\top$ (рис. 2). По определению ранга $\operatorname{rank} A \ge r$.

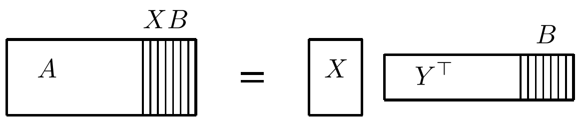

Рисунок 2.

Полученное неравенство вместе с указанным ранее обратным неравенством $\operatorname{rank} A \le r$ приводят к равенству $\operatorname{rank} A = r$. ■

Следствие. *Для того чтобы матрица* $A = A[M,N]$ *имела ранг* 1 *(была одноранговой), необходимо и достаточно, чтобы существовало представление*

$$A = xy^\top,$$

где $x = x[M,1]$ — *ненулевой вектор-столбец и* $y^\top = y^\top[1,N]$ — *ненулевая вектор-строка.*

Представление (5.1) называется *скелетным разложением* матрицы A.

1.6. Системы линейных уравнений

Пусть $A = A[M,N]$ —прямоугольная матрица. Введем линейное множество

$$\mathcal{L} = \{x \in \mathcal{R}^N \mid Ax = \mathbf{0}\}.$$

Оно называется *ядром* матрицы A.

ТЕОРЕМА 6.1. *Справедлива формула*

$$\dim \mathcal{L} = |N| - \operatorname{rank} A.$$

Доказательство. Для нулевой матрицы A теорема тривиальна. Предположим, что A — ненулевая матрица. Обозначим через Q множество индексов ее базисных столбцов, так что $|Q| = \operatorname{rank} A$. Зафиксируем $k \in N \setminus Q$. По определению базиса столбцов система уравнений

$$A[M,Q] \times x_k[Q] = -A[M,k] \tag{6.1}$$

имеет единственное решение. Доопределим x_k, положив

$$x_k[j] = \begin{cases} 1 & \text{при } j = k, \\ 0 & \text{при остальных } j \in N \setminus Q. \end{cases}$$

Получим $A[M,N] \times x_k[N] = \mathbf{0}[M]$. Тем самым, построено $|N \setminus Q| = |N| - \operatorname{rank} A$ элементов x_k подпространства $\mathcal{L}$. Эти элементы линейно независимы, что следует из их задания на индексах $j \in N \setminus Q$. Покажем, что $\{x_k\}$ — базис $\mathcal{L}$.

Возьмем произвольный элемент z из $\mathcal{L}$ и рассмотрим вектор

$$z_0 = \sum_{k \in N \setminus Q} z[k]\ x_k.$$

Проверим, что $z_0 = z$. Равенство $z_0[N \setminus Q] = z[N \setminus Q]$ справедливо в силу определения x_k. Далее, согласно (6.1)

$$A[M,Q] \times z_0[Q] = \sum_{k \in N \setminus Q} z[k] A[M,Q] \times x_k[Q] =$$

$$= -\sum_{k \in N \setminus Q} z[k]\, A[M,k] = -A[M, N \setminus Q] \times z[N \setminus Q]. \tag{6.2}$$

По условию $A[M,N] \times z[N] = \mathbf{0}[M]$, что можно переписать в виде

$$A[M,Q] \times z[Q] = -A[M, N \setminus Q] \times z[N \setminus Q]. \tag{6.3}$$

Правые части в (6.2) и (6.3) равны. Значит, равны и левые, т.е.

$$A[M,Q] \times (z_0[Q] - z[Q]) = \mathbf{0}[M].$$

В силу линейной независимости столбцов матрицы $A[M,Q]$ получаем $z_0[Q] = z[Q]$.

Установлено, что $z_0 = z$. Это означает, что любой элемент $z \in \mathcal{L}$ допускает представление

$$z = \sum_{k \in N \setminus Q} z[k]\; x_k.$$

По основной лемме линейной алгебры в $\mathcal{L}$ не может быть более $|N \setminus Q|$ линейно независимых элементов. Но столько, $|N \setminus Q|$, есть — это $\{x_k\}$. По определению размерности линейного множества $\dim \mathcal{L} = |N \setminus Q| = |N| - \operatorname{rank} A$. Теорема доказана. ■

Векторы u, v из $\mathcal{R}^M$ называются *ортогональными*, если $\langle u, v\rangle = 0$. Возьмем линейное множество (подпространство) $\mathcal{H}$ пространства $\mathcal{R}^M$. *Ортогональным дополнением* к $\mathcal{H}$ называется множество вида

$$\mathcal{H}^\perp = \{u \in \mathcal{R}^M \mid \langle u, v\rangle = 0 \text{ для всех } v \in \mathcal{H}\}.$$

ТЕОРЕМА 6.2. *Пусть* $A = A[M,N]$ — *прямоугольная матрица и* $\mathcal{H} = \{v \in \mathcal{R}^M \mid A^\top v = \mathbf{0}\}$. *Тогда*

$$\mathcal{H}^\perp = \{u = Ax \mid x \in \mathcal{R}^N\}. \tag{6.4}$$

Доказательство. Для нулевой матрицы A теорема тривиальна. Ортогональным дополнением к $\mathcal{H} = \mathcal{R}^M$ является $\mathcal{H}^\perp = \{\mathbf{0}\}$. Действительно, ненулевой вектор u не может принадлежать $\mathcal{H}^\perp$, поскольку для вектора $v = u$ из $\mathcal{H}$ будет $\langle u, v\rangle = \langle u, u\rangle > 0$.

В дальнейшем считаем, что A — ненулевая матрица. Обозначим через $\mathcal{P}$ множество, стоящее в правой части (6.4), и покажем сначала, что $\mathcal{P} \subset \mathcal{H}^\perp$. Возьмем $u \in \mathcal{P}$. Тогда для любого $v \in \mathcal{H}$ согласно (1.6) получим

$$\langle u, v\rangle = \langle Ax, v\rangle = \langle x, A^\top v\rangle = 0.$$

Это значит, что $u \in \mathcal{H}^\perp$.

Проверим обратное включение $\mathcal{H}^\perp \subset \mathcal{P}$. Возьмем $u \in \mathcal{H}^\perp$. Выделим подматрицу $B = A[M,Q]$, составленную из базисных столбцов матрицы A, и покажем, что $u = Bx$ при некотором $x = x[Q]$.

Анализ. Если такой вектор x существует, то справедливо равенство $B^\top u = B^\top Bx$. По теореме 4.2 матрица $B^\top B$ обратима, поэтому

$$x = (B^\top B)^{-1} B^\top u. \tag{6.5}$$

Продолжение доказательства. Вычислим x по формуле (6.5) и проверим, что $u = Bx$. Обозначим $v = u - Bx$. Согласно (6.5) $B^\top v = B^\top u - B^\top Bx = \mathbf{0}$. Подробнее

$$A^\top[Q,M] \times v[M] = \mathbf{0}[Q].$$

Поскольку строки матрицы $A^\top[Q,M]$ образуют базис строк матрицы $A^\top[N,M]$, то
$$A^\top[N,M] \times v[M] = \mathbf{0}[M].$$
Это означает, что $v \in \mathcal{H}$.

По условию $u \in \mathcal{H}^\perp$. Имеем
$$0 = \langle u, v\rangle = \langle v + Bx, v\rangle = \langle v, v\rangle + \langle x, B^\top v\rangle = \langle v, v\rangle.$$
Следовательно, $v = \mathbf{0}$, или $u = Bx$, или
$$u[M] = A[M,Q] \times x[Q].$$
Доопределим вектор x, положив $x[j] = 0$ при $j \in N \setminus Q$. Получим $u = Ax$, т.е. $u \in \mathcal{P}$. Включение $\mathcal{H}^\perp \subset \mathcal{P}$, а с ним и равенство $\mathcal{H}^\perp = \mathcal{P}$, установлены. ■

ТЕОРЕМА 6.3. *Система линейных уравнений $Ax = b$ с прямоугольной матрицей $A = A[M,N]$ имеет решение (совместна) тогда и только тогда, когда любое решение однородной сопряженной системы $A^\top v = \mathbf{0}$ ортогонально вектору b.*

Доказательство. Обозначим $\mathcal{H} = \{v \in \mathcal{R}^M \mid A^\top v = \mathbf{0}\}$. Если $\langle b, v\rangle = 0$ при любом $v \in \mathcal{H}$, то $b \in \mathcal{H}^\perp$. По теореме 6.2 вектор b допускает представление $b = Ax$, так что система $Ax = b$ совместна.

Наоборот, пусть x_0 удовлетворяет системе $Ax = b$. Для любого $v \in \mathcal{H}$ имеем
$$\langle b, v\rangle = \langle Ax_0, v\rangle = \langle x_0, A^\top v\rangle = 0.$$
Теорема доказана. ■

1.7. Вычисление ранга матрицы и QR-разложение

В линейной алгебре вектор u обычно фигурирует как вектор-столбец. Для соответствующей вектор-строки используется обозначение $u^\top$.

Пусть $A = A[M,N]$ — ненулевая прямоугольная матрица ранга r. Рассмотрим матрицу
$$F = A - xy^\top,$$
где $x = x[M,1]$ — вектор-столбец, а $y^\top = y^\top[1,N]$ — вектор-строка. По теореме 3.2 ранг матрицы $xy^\top$ равен нулю или единице. Согласно теореме 3.3
$$\mathrm{rank}\ F \le \mathrm{rank}\ A + 1,$$
$$\mathrm{rank}\ A = \mathrm{rank}\ (F + xy^\top) \le \mathrm{rank}\ F + 1.$$
Поскольку $\mathrm{rank}\ A = r$, то
$$r - 1 \le \mathrm{rank}\ F \le r + 1. \tag{7.1}$$

ТЕОРЕМА 7.1. *Справедливы утверждения:*

(I) rank $F = r - 1$ *тогда и только тогда, когда вектор* x *можно представить в виде линейной комбинации столбцов матрицы* A, *а вектор* $y^\top$ — *в виде линейной комбинации строк матрицы* A, *т.е.*

$$x = Av_0, \quad y^\top = u_0^\top A, \tag{7.2}$$

при этом должно выполняться равенство

$$\langle u_0, Av_0 \rangle = 1; \tag{7.3}$$

(II) rank $F = r + 1$ *тогда и только тогда, когда обе системы линейных уравнений* $Av = x$ *и* $A^\top u = y$ *несовместны;*

(III) rank $F = r$ *во всех остальных случаях.*

Доказательство. Начнем с утверждения (I).

Достаточность. Согласно (7.2)

$$F = A - Av_0 u_0^\top A. \tag{7.4}$$

Пусть $n = |N|$. По теореме 6.1 система линейных однородных уравнений $Av = \mathbf{0}$ имеет $n - r$ линейно независимых решений $v_1, \ldots, v_{n-r}$. В силу (7.3), $Av_0 \neq \mathbf{0}$, поэтому линейно независимой будет расширенная система векторов $v_0, v_1, \ldots, v_{n-r}$.

Учитывая (7.4), получаем $Fv_j = \mathbf{0}$ при $j \in 1\colon n - r$. Кроме того, согласно (7.3)

$$Fv_0 = Av_0 - Av_0(u_0^\top Av_0) = Av_0 - Av_0\langle u_0, Av_0 \rangle = \mathbf{0}.$$

Значит, система линейных однородных уравнений $Fv = \mathbf{0}$ имеет $n - r + 1$ линейно независимых решений $v_0, v_1, \ldots, v_{n-r}$.

Обозначим $\mathcal{H} = \{v \in \mathcal{R}^N \mid Fv = \mathbf{0}\}$. Вспоминая определение размерности линейного множества и снова привлекая теорему 6.1, приходим к соотношениям

$$n - r + 1 \leq \dim \mathcal{H} = n - \operatorname{rank} F.$$

Отсюда следует, что rank $F \leq r - 1$. Обратное неравенство указано в (7.1). Таким образом, rank $F = r - 1$.

Необходимость . Пусть теперь rank $F = r - 1$. Поскольку rank $A = r$, то найдутся векторы v_0 и u_0, такие, что

$$Fv_0 = \mathbf{0}, \qquad Av_0 \neq \mathbf{0},$$

$$u_0^\top F = \mathbf{0}, \qquad u_0^\top A \neq \mathbf{0}.$$

Действительно, ядро $\{v \mid Fv = \mathbf{0}\}$ матрицы F имеет размерность $n - r + 1$. Если $v_1, \ldots, v_{n-r+1}$ — базис ядра, то $Av_j \neq \mathbf{0}$ при некотором $j \in 1\colon n - r + 1$. Иначе оказалось бы, что размерность множества $\mathcal{L} = \{v \mid Av = \mathbf{0}\}$ не меньше $n - r + 1$, в то время как по теореме 6.1

$\dim \mathcal{L} = n - r$. Соответствующее v_j и обозначим через v_0. Аналогично устанавливается существование вектора u_0 со свойствами $F^\top u_0 = \mathbf{0}$, $A^\top u_0 \neq \mathbf{0}$.

Имеем

$$\begin{aligned} Av_0 &= (F + xy^\top)v_0 = \langle y, v_0\rangle x, \\ u_0^\top A &= u_0^\top (F + xy^\top) = \langle u_0, x\rangle y^\top. \end{aligned} \tag{7.5}$$

По построению векторы Av_0 и $u_0^\top A$ ненулевые. Значит, $\langle y, v_0\rangle \neq 0$, $\langle u_0, x\rangle \neq 0$. Можно считать, что

$$\langle y, v_0\rangle = 1, \qquad \langle u_0, x\rangle = 1. \tag{7.6}$$

Для этого вектор v_0 следует заменить на $v_0/\langle y, v_0\rangle$, а вектор u_0 — на $u_0/\langle u_0, x\rangle$. Согласно (7.5), (7.6) получаем

$$Av_0 = x, \qquad u_0^\top A = y^\top,$$

$$\langle u_0, Av_0\rangle = \langle u_0, x\rangle = 1.$$

Утверждение (I) доказано.

Переходим к утверждению (II).

Д о с т а т о ч н о с т ь. Пусть системы $Av = x$ и $A^\top u = y$ несовместны. По доказанному $\operatorname{rank} F \neq r - 1$. Допустим, что $\operatorname{rank} F = r$. Рассмотрим векторы $v_1, \ldots, v_{n-r}$, образующие базис подпространства $\mathcal{H} = \{v \mid Fv = \mathbf{0}\}$. В силу определения F имеем

$$Av_j = \langle y, v_j\rangle x, \qquad j \in 1 \colon n - r.$$

Поскольку система $Av = x$ несовместна, то $\langle y, v_j\rangle = 0$ при всех $j \in 1 \colon n - r$. Условие $\operatorname{rank} A = r$ гарантирует, что векторы $v_1, \ldots, v_{n-r}$ образуют базис подпространства $\mathcal{L} = \{v \mid Av = \mathbf{0}\}$.

Система $A^\top u = y$ также несовместна. По теореме 6.3 найдется решение v_0 однородной сопряженной системы $Av = \mathbf{0}$, такое, что $\langle y, v_0\rangle \neq 0$. Вектор v_0 принадлежит $\mathcal{L}$, поэтому его можно представить в виде

$$v_0 = \sum_{j=1}^{n-r} t_j v_j.$$

Но тогда

$$\langle y, v_0\rangle = \sum_{j=1}^{n-r} t_j \langle y, v_j\rangle = 0.$$

Получили противоречие.

Показано, что ранг F отличен от $r - 1$ и r. Согласно (7.1) $\operatorname{rank} F = r + 1$.

Н е о б х о д и м о с т ь . Пусть $\operatorname{rank} F = r+1$. Поскольку $\operatorname{rank} A = r$, то найдутся векторы v_0 и u_0 со свойствами

$$Av_0 = \mathbf{0}, \qquad Fv_0 \neq \mathbf{0},$$

$$A^\top u_0 = \mathbf{0}, \qquad F^\top u_0 \neq \mathbf{0}.$$

Имеем $Fv_0 = -\langle y, v_0\rangle x$, $F^\top u_0 = -\langle x, u_0\rangle y$. В частности, $\langle y, v_0\rangle \neq 0$, $\langle x, u_0\rangle \neq 0$.

Система $Av = x$ несовместна, так как вектор u_0 удовлетворяет однородной сопряженной системе $A^\top u = \mathbf{0}$ и $\langle x, u_0\rangle \neq 0$. По аналогичной причине несовместна и система $A^\top u = y$. Утверждение (II) доказано.

Справедливость утверждения (III) следует из (7.1). ■

Теорема 7.1 дает основу для построения алгоритма вычисления ранга ненулевой прямоугольной матрицы $A = A[M, N]$. Возьмем ненулевой столбец $A[M, j_0] =: a$. Обозначим $\|a\| = \langle a, a\rangle^{1/2}$. Величина $\|a\|$ называется нормой вектора a. Проверим, что векторы

$$x = \frac{1}{\|a\|}a, \qquad y^\top = x^\top A$$

удовлетворяют условиям (7.2), (7.3).

Прежде всего отметим, что $\langle x, x\rangle = \frac{1}{\|a\|^2}\langle a, a\rangle = 1$. Извлекая квадратный корень, получаем $\|x\| = 1$. Вектор x со свойством $\|x\| = 1$ называется единичным или нормированным. Вектор y — ненулевой, у него $y[j_0] = \|a\|$. Далее

$$x = \frac{1}{\|a\|}A[M, j_0] = \frac{1}{\|a\|}Ae_{j_0} = A\left(\frac{1}{\|a\|}e_{j_0}\right) =: Av_0,$$

$$y^\top = x^\top A =: u_0^\top A,$$

$$\langle u_0, Av_0\rangle = \langle x, x\rangle = 1.$$

Согласно теореме 7.1 $\operatorname{rank}\,(A - xy^\top) = \operatorname{rank} A - 1$.

При данном выборе векторов x и y справедливо представление

$$F := A - xy^\top = A - xx^\top A = (E - xx^\top)A. \tag{7.7}$$

Матрица $P = E - xx^\top$ есть матрица проектирования на подпространство $\mathcal{L}$, ортогональное вектору x (рис. 3).

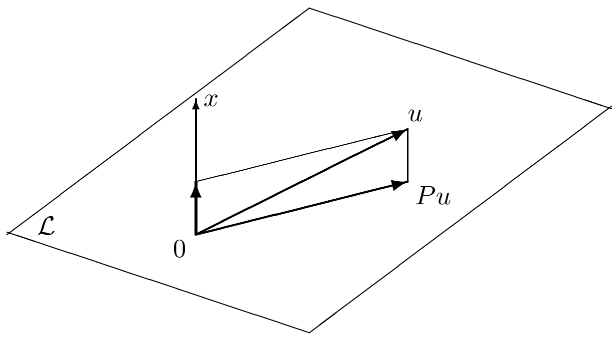

Рисунок 3.

Действительно, $Pv = v$ для всех $v \in \mathcal{L}$. Кроме того, $Px = \mathbf{0}$, поскольку $\langle x, x\rangle = 1$. Значит, столбцы матрицы F являются проекциями соответствующих столбцов матрицы A на подпространство $\mathcal{L}$. В частности, столбец $A[M, j_0]$ преобразуется в нулевой столбец матрицы F.

Последовательное применение указанного приема приводит к следующему алгоритму.

Алгоритм вычисления ранга прямоугольной матрицы.

1. Полагаем $q = 0,\ F_0 = A$.
2. Если $F_q = \mathbf{0}$, то процесс закончен и $\operatorname{rank} A = q$.
3. Иначе полагаем $q := q + 1$.
4. Берем ненулевой столбец $f_q = F_{q-1}[M, j_q]$ матрицы F_{q-1} и строим векторы
$$x_q = \frac{1}{\|f_q\|} f_q, \qquad y_q^\top = x_q^\top F_{q-1}.$$
5. Вычисляем $F_q = F_{q-1} - x_q y_q^\top$.
6. Переходим к пункту 2.

В результате работы алгоритма получим $q = \operatorname{rank} A$ и

$$A = \sum_{k=1}^{q} x_k\ y_k^\top. \tag{7.8}$$

Поясним это. Алгоритм строит последовательность матриц

$$\begin{aligned} &F_0 = A, \\ &F_k = F_{k-1} - x_k y_k^\top, \quad k \in 1\colon q, \end{aligned}$$

таких, что $\mathrm{rank}\ F_k = \mathrm{rank}\ F_{k-1} - 1$, и $F_q = \mathbf{0}$. Поскольку $0 = \mathrm{rank}\ F_q = \mathrm{rank}\ F_{q-1} - 1 = \ldots = \mathrm{rank}\ F_0 - q = \mathrm{rank}\ A - q$, то $\mathrm{rank}\ A = q$. По условию $\mathrm{rank}\ A = r$. Значит, $q = r$. Кроме этого основного результата, мы получаем новое разложение (7.8) матрицы $A = A[M, N]$. Исследуем его свойства.

Покажем, что векторы $x_1, \ldots, x_q$ попарно ортогональны. Согласно (7.7)

$$x_1^\top F_1 = x_1^\top (F_0 - x_1 y_1^\top) = x_1^\top F_0 - y_1^\top = \mathbf{0}.$$

Это значит, что каждый столбец матрицы F_1 ортогонален x_1. В частности, $\langle x_1, x_2 \rangle = 0$.

Допустим, что векторы $x_1, \ldots, x_k$ попарно ортогональны, и проверим их ортогональность x_{k+1}. Вектор x_{k+1} с точностью до множителя равен ненулевому столбцу матрицы F_k. При $j \in 1\colon k$ имеем

$$x_j^\top F_k = x_j^\top \left(F_{j-1} - \sum_{i=j}^{k} x_i y_i^\top \right) = x_j^\top F_{j-1} - y_j^\top = \mathbf{0}.$$

Отсюда, в частности, следует, что $\langle x_j, x_{k+1} \rangle = 0$ при $j \in 1\colon k$. Ортогональность векторов $x_1, \ldots, x_q$ установлена.

Теперь отметим такой факт. Если на некотором шаге алгоритма столбец текущей матрицы F_{k-1} перешел в нулевой, то в дальнейшем он останется нулевым. Это связано с тем, что согласно (7.7) шаг алгоритма эквивалентен линейному преобразованию матрицы F_k.

Если обозначить $P_k = E - x_k x_k^\top$, то

$$F_k = P_k F_{k-1}.$$

Как следствие,

$$\mathbf{0} = F_q = P_q F_{q-1} = \cdots = P_q P_{q-1} \cdots P_1 A,$$

т.е.

$$P_q P_{q-1} \cdots P_1 A = \mathbf{0}.$$

Эта формула поясняет, что описанный алгоритм есть алгоритм последовательного проектирования.

По сути алгоритма у матрицы $F_k = F_{k-1} - x_k y_k^\top$, по крайней мере, на один нулевой столбец больше, чем у F_{k-1}. У вектора $y_k^\top$ компоненты, соответствующие нулевым столбцам матрицы F_{k-1}, равны нулю (нулевые столбцы переходят в нулевые). Компонента $y_k^\top[j_k]$, соответствующая выбираемому ненулевому столбцу f_k матрицы F_{k-1}, отлична от нуля (равна $\|f_k\|$). Поскольку столбец f_k переходит в нулевой, то $y_{k+1}^\top[j_k] = 0$. Ясно, что количество нулевых компонент у $y_{k+1}^\top$ больше, чем у $y_k^\top$. Точнее

- у вектора $y_2^\top$, по крайней мере, одна нулевая компонента,

- у вектора $y_3^\top$, по крайней мере, две нулевые компоненты,

   ....

- у вектора $y_q^\top$, по крайней мере, $q-1$ нулевых компонент.

Перепишем (7.8) в матричной форме

$$A = XY^\top. \tag{7.9}$$

Здесь X — матрица, составленная из ортонормированных столбцов $x_1, \ldots, x_q$, а $Y^\top$ — матрица, составленная из строк $y_1^\top, \ldots, y_q^\top$. По теореме 3.2 строки матрицы $Y^\top$ линейно независимы. Соотношение (7.9) называется QR-разложением прямоугольной матрицы $A = A[M, N]$.

ЗАДАЧИ

7.1. Пусть $w_1, w_2, \ldots, w_k$ — линейно независимые векторы из $\mathcal{R}^N$, причем $\|w_1\| = 1$. Введем матрицу проектирования $P_1 = E - w_1 w_1^\top$. Докажите, что векторы $P_1 w_2, \ldots, P_1 w_k$ линейно независимы.

7.2. Пусть $A = A[1\colon n, 1\colon n]$ — обратимая матрица. Докажите, что ее можно представить в виде $A = XY^\top$, где $X = X[1\colon n, 1\colon n]$ — матрица с ортонормированными столбцами и $Y^\top = Y^\top[1\colon n, 1\colon n]$ — верхняя треугольная матрица, у которой $Y^\top[i,j] = 0$ при $i > j$ и $Y^\top[j,j] > 0$ при всех $j \in 1\colon n$.

7.3. Докажите, что разложение обратимой матрицы $A = A[1\colon n, 1\colon n]$, указанное в предыдущей задаче, единственно.

ЛИТЕРАТУРА К ГЛАВЕ 1

- Воеводин В.В., Кузнецов Ю.А. Матрицы и вычисления. М., 1984. 320 с.

- Хорн Р., Джонсон Ч. Матричный анализ. М., 1989. 685 с.

- Даугавет В.А., Яковлев П.В. Среднеквадратическая аппроксимация прямоугольной матрицы матрицами меньшего ранга //Журнал вычислительной математики и математической физики. 1989. Т.29. ц 10. С.1466–1479.

ГЛАВА 2. КВАДРАТИЧНАЯ ФУНКЦИЯ

2.1. Симметричные матрицы

Квадратная матрица $D = D[N, N]$ называется *симметричной*, если $D^\top = D$. Распишем это равенство по элементам: $D^\top[i, j] = D[i, j]$, $i, j \in N$. Поскольку $D^\top[i, j] = D[j, i]$, то симметричность матрицы D означает, что $D[j, i] = D[i, j]$ при всех $i, j \in N$.

ЛЕММА 1.1. *Квадратная матрица $D = D[N, N]$ симметрична тогда и только тогда, когда*

$$\langle Dx, y\rangle = \langle x, Dy\rangle \quad \text{при всех } \ x, y \in \mathcal{R}^N. \tag{1.1}$$

Доказательство. Если матрица D симметрична, то равенство (1.1) следует из формулы (1.6) главы 1. Наоборот, допустив справедливость (1.1), при $x = e_i, \ y = e_j$ получим

$$\langle Dx, y\rangle = D[j, i], \quad \langle x, Dy\rangle = D[i, j],$$

так что $D[j, i] = D[i, j]$ при всех $i, j \in N$. ■

Симметричная матрица D называется *неотрицательно определенной*, если $\langle Dx, x\rangle \geq 0$ при всех $x \in \mathcal{R}^N$, и *положительно определенной*, если $\langle Dx, x\rangle > 0$ при всех ненулевых $x \in \mathcal{R}^N$.

ЛЕММА 1.2. *Симметричная положительно определенная матрица D имеет обратную матрицу D^{-1}, которая также является симметричной и положительно определенной.*

Доказательство. Однородная система $Dx = \mathbf{0}$ имеет только нулевое решение. Действительно, если бы $Dx_0 = \mathbf{0}$ при некотором $x_0 \neq \mathbf{0}$, то получили бы $\langle Dx_0, x_0\rangle = 0$, что противоречит положительной определенности матрицы D. Установлено, что столбцы матрицы D линейно независимы. По теореме 4.1 главы 1 матрица D имеет обратную D^{-1}.

В силу симметричности D

$$\langle D^{-1}x, y\rangle = \langle D^{-1}x, DD^{-1}y\rangle = \langle x, D^{-1}y\rangle,$$

откуда следует симметричность D^{-1}.

Возьмем $x \neq \mathbf{0}$ и обозначим $y = D^{-1}x$. Очевидно, что $y \neq \mathbf{0}$. Поскольку

$$\langle D^{-1}x, x\rangle = \langle D^{-1}x, DD^{-1}x\rangle = \langle y, Dy\rangle > 0,$$

то D^{-1} — положительно определенная матрица. ■

Используя симметричную положительно определенную матрицу D, введем обобщенное скалярное произведение

$$\langle x, y\rangle_D = \langle Dx, y\rangle$$

и обобщенную норму $\|x\|_D = \langle Dx, x\rangle^{1/2}$. Обобщенная норма обладает следующими свойствами:

(I) $\|x\|_D = 0$ тогда и только тогда, когда $x = \mathbf{0}$;

(II) $\|tx\|_D = |t| \cdot \|x\|_D$ при любом $t \in \mathcal{R}$;

(III) $\langle x, y\rangle_D \le \|x\|_D \cdot \|y\|_D$;

(IV) $\|x + y\|_D \le \|x\|_D + \|y\|_D$.

Свойства (I) и (II) очевидны. Проверим (III).

Если хотя бы один из векторов x, y равен нулю, то неравенство (III) выполняется как равенство. Поэтому будем считать, что $x \ne \mathbf{0}$ и $y \ne \mathbf{0}$. Пусть сначала $\|x\|_D = 1$, $\|y\|_D = 1$. Тогда

$$\|x - y\|_D^2 = \langle D(x - y), x - y\rangle = 2 - 2\langle x, y\rangle_D.$$

Перепишем это соотношение в виде

$$\langle x, y\rangle_D = 1 - {}^1\!/\!_2\|x - y\|_D^2.$$

В случае произвольных ненулевых векторов x, y получаем

$$\left\langle \frac{x}{\|x\|_D}, \frac{y}{\|y\|_D} \right\rangle_D = 1 - \frac{1}{2}\left\| \frac{x}{\|x\|_D} - \frac{y}{\|y\|_D} \right\|_D^2.$$

Значит,

$$\langle x, y\rangle_D = \left(1 - \frac{1}{2}\left\| \frac{x}{\|x\|_D} - \frac{y}{\|y\|_D} \right\|_D^2\right) \|x\|_D \cdot \|y\|_D. \tag{1.2}$$

Отсюда очевидным образом следует неравенство (III). Более того, поскольку $-\langle x, y\rangle_D = \langle -x, y\rangle_D \le \|x\|_D \cdot \|y\|_D$, то

$$|\langle x, y\rangle_D| \le \|x\|_D \cdot \|y\|_D.$$

Что касается неравенства (IV), то после возведения в квадрат и элементарных преобразований оно переходит в неравенство (III).

Отметим, что согласно (1.2) неравенства (III) и (IV) выполняются как равенства при ненулевых x, y тогда и только тогда, когда

$$\frac{x}{\|x\|_D} = \frac{y}{\|y\|_D}.$$

Обычная евклидова норма соответствует обобщенной норме при $D = E$. Евклидова норма вектора x обозначается $\|x\|$.

ЗАДАЧИ

1.1. Докажите, что произведение AB симметричных матриц симметрично тогда и только тогда, когда матрицы A и B перестановочны, т.е. когда $AB = BA$.

1.2. Пусть у матрицы $A = A[M, N]$ столбцы линейно независимы. Докажите, что матрица $D = A^\top A$ является симметричной и положительно определенной.

1.3. Докажите, что для элементов симметричной положительно определенной матрицы $D = D[N, N]$ выполняется неравенство

$$D[i, j] < {}^1\!/_2\,(D[i, i] + D[j, j]) \quad \text{при } i \neq j.$$

1.4. Пусть $D[N, N]$ — симметричная положительно определенная матрица и $N_0 \subset N$. Докажите, что подматрица $D[N_0, N_0]$ также симметрична и положительно определена.

1.5. Докажите, что симметричная положительно определенная матрица $D = D[1 : n, 1 : n]$ допускает представление $D = LL^\top$, где L – нижняя треугольная матрица с положительными диагональными элементами, т.е. $L[i, j] = 0$ при $i < j$ и $L[i, i] > 0$ при всех $i \in 1 : n$ (разложение Холесского).

2.2. Выпуклые квадратичные функции

Пусть $D = D[N, N]$ — симметричная матрица. Функция вида

$$Q(x) = {}^1\!/_2\langle Dx, x\rangle + \langle c, x\rangle + \alpha$$

называется *квадратичной.*

ЛЕММА 2.1. *Справедливо разложение*

$$Q(x + h) - Q(x) = \langle Dx + c, h\rangle + {}^1\!/_2\langle Dh, h\rangle. \tag{2.1}$$

Доказательство. В силу симметричности матрицы D имеем

$$Q(x+h) = {}^1\!/_2\langle D(x+h), x+h\rangle + \langle c, x+h\rangle + \alpha = Q(x) + {}^1\!/_2\langle Dh, h\rangle + \langle Dx+c, h\rangle.$$

Это равносильно (2.1). ■

Следствие. При всех x, h из $\mathcal{R}^N$ и $t \in \mathcal{R}$ справедливо равенство

$$Q(x + th) - Q(x) = t\ [Q(x + h) - Q(x)] - {}^1\!/_2\, t\ (1 - t)\langle Dh, h\rangle. \tag{2.2}$$

Действительно, согласно (2.1)

$$Q(x + th) - Q(x) = t\ \langle Dx + c, h\rangle + {}^1\!/_2\, t^2\langle Dh, h\rangle.$$

Вместе с тем, из (2.1) следует, что

$$\langle Dx + c, h\rangle = Q(x+h) - Q(x) - {}^1\!/_2\langle Dh, h\rangle.$$

Объединяя две последние формулы, приходим к (2.2).

Функция Q называется *выпуклой на* $\mathcal{R}^N$, если

$$Q(tx_1 + (1-t)x_0) \le t\,Q(x_1) + (1-t)\,Q(x_0) \tag{2.3}$$

для всех x_0, x_1 из $\mathcal{R}^N$ и $t \in [0,1]$.

ТЕОРЕМА 2.1. *Для того чтобы квадратичная функция Q была выпуклой на $\mathcal{R}^N$, необходимо и достаточно, чтобы матрица D была неотрицательно определенной.*

Доказательство. Необходимость . Пусть Q — выпуклая функция. Перепишем неравенство (2.3) в виде

$$Q(x_0 + t(x_1 - x_0)) \le Q(x_0) + t\,[Q(x_1) - Q(x_0)]. \tag{2.4}$$

Зафиксируем x_0 и положим $x_1 = x_0 + h$. При $t \in (0,1)$ имеем

$$Q(x_0 + th) - Q(x_0) - t\,[Q(x_0 + h) - Q(x_0)] \le 0. \tag{2.5}$$

В силу (2.2) получаем $-{}^1\!/_2\,t\,(1-t)\langle Dh, h\rangle \le 0$, откуда следует, что

$$\langle Dh, h\rangle \ge 0 \quad \text{при всех } \ h \in \mathcal{R}^N. \tag{2.6}$$

Неотрицательная определенность матрицы D установлена.

Достаточность. Зафиксируем x_0, x_1 из $\mathcal{R}^N$ и $t \in [0,1]$. Положим $h = x_1 - x_0$. Из (2.2) и (2.6) следует неравенство (2.5), равносильное (2.4). Последнее, в свою очередь, характеризует функцию Q как выпуклую. ■

Рассмотрим случай, когда D — положительно определенная матрица. Введем величину

$$\lambda = \min_{\|x\|=1} \langle Dx, x\rangle.$$

Минимум непрерывной функции $\langle Dx, x\rangle$ на ограниченном замкнутом множестве $S = \{x \in \mathcal{R}^N \mid \|x\| = 1\}$ достигается. Обозначим точку минимума через x_0. В силу положительной определенности матрицы D имеем $\lambda = \langle Dx_0, x_0\rangle > 0$.

Согласно выбору λ

$$\langle Dx, x\rangle \ge \lambda\|x\|^2 \quad \text{при всех } \ x \in \mathcal{R}^N.$$

Воспользуемся соотношением (2.2) при $x = x_0$ и $h = x_1 - x_0$. Для $t \in [0,1]$ получим

$$Q(tx_1 + (1-t)x_0) \le t\,Q(x_1) + (1-t)\,Q\,(x_0) - {}^1\!/_2\lambda\,t\,(1-t)\|x_1 - x_0\|^2. \tag{2.7}$$

Неравенство (2.7) характеризует функцию Q как *сильно выпуклую.*

Установим еще один критерий выпуклости квадратичной функции.

ТЕОРЕМА 2.2. *Для того чтобы квадратичная функция Q была выпуклой на $\mathcal{R}^N$, необходимо и достаточно, чтобы при всех x и h из $\mathcal{R}^N$ выполнялось неравенство*

$$Q(x+h) - Q(x) \geq \langle Dx + c, h \rangle. \tag{2.8}$$

Доказательство. Необходимость . Выпуклость функции Q согласно теореме 2.1 гарантирует неотрицательную определенность матрицы D. Теперь (2.8) следует из (2.1).

Достаточность. На основании (2.1) и (2.8) заключаем, что $\langle Dh, h \rangle \geq 0$ при всех $h \in \mathcal{R}^N$, т.е. что D — неотрицательно определенная матрица. Остается сослаться на теорему 2.1. ■

Неравенство (2.8) можно переписать в эквивалентном виде

$$Q(x_1) - Q(x_0) \geq \langle Dx_0 + c, x_1 - x_0 \rangle. \tag{2.9}$$

Рассмотрим аффинное многообразие

$$\Omega = \{x \in \mathcal{R}^N \mid A[M,N] \times x[N] = b[M]\}.$$

Будем считать, что оно не пустое. Квадратичную функцию Q будем называть выпуклой на Ω, если неравенство (2.3) выполняется при всех x_0, x_1 из Ω и $t \in [0,1]$.

Положим $\Omega_0 = \{h \mid Ah = \mathbf{0}\}$.

ТЕОРЕМА 2.3. *Для того чтобы квадратичная функция Q была выпуклой на Ω, необходимо и достаточно, чтобы матрица D была неотрицательно определенной на Ω_0.*

Доказательство. Необходимость . Возьмем произвольные точки $x_0 \in \Omega$ и $h \in \Omega_0$. Очевидно, что точка $x_1 = x_0 + h$ принадлежит Ω. Так как функция Q выпукла на Ω, то согласно (2.2) при $x = x_0$ и $t = 1/2$ получаем $-1/8\langle Dh, h \rangle \leq 0$, или $\langle Dh, h \rangle \geq 0$ при всех $h \in \Omega_0$. Это означает, что матрица D неотрицательно определена на Ω_0.

Достаточность. Пусть точки x_0, x_1 принадлежат Ω. Тогда разность $h = x_1 - x_0$ принадлежит Ω_0. Неравенство (2.3) для $t \in [0,1]$ следует из соотношения (2.2) при $x = x_0$ и неотрицательной определенности матрицы D на Ω_0. ■

ТЕОРЕМА 2.4. *Для того чтобы квадратичная функция Q была выпуклой на Ω, необходимо и достаточно, чтобы выполнялось неравенство* (2.8) *при всех $x \in \Omega$ и $h \in \Omega_0$.*

Доказательство. Необходимость . Согласно теореме 2.3 выпуклость Q на Ω влечет неотрицательную определенность матрицы D на Ω_0. Неравенство (2.8) при $x \in \Omega$ и $h \in \Omega_0$ следует теперь из (2.1).

Д о с т а т о ч н о с т ь. Возьмем $x \in \Omega$, $h \in \Omega_0$. На основании (2.1) и (2.8) получаем $\langle Dh, h\rangle \geq 0$. Остается сослаться на теорему 2.3. ■

Отметим, что в случае неотрицательной определенности матрицы D на множестве $\Omega_0 = \{h \,|\, Ah = \mathbf{0}\}$ квадратичная функция Q будет выпуклой на аффинном многообразии $\Omega = x_0 + \Omega_0 = \{x = x_0 + h \,|\, h \in \Omega_0\}$ при любом фиксированном $x_0 \in \mathcal{R}^N$.

ЗАДАЧИ

2.1. Докажите выпуклость на $\mathcal{R}^N$ квадратичной функции

$$Q(x) = {}^1\!/_2 \|Ax - b\|^2,$$

где $A = A[M, N]$ — произвольная матрица.

2.2. Пусть $Q(x)$ — квадратичная функция. Докажите, что функция

$$F(y) = Q(x_0 + By),$$

где $B = B[N, M]$, также является квадратичной.

2.3. Пусть $Q(x) = {}^1\!/_2\langle Gx, x\rangle$, где G — произвольная квадратная матрица. Укажите такую симметричную матрицу D, что $Q(x) = {}^1\!/_2\langle Dx, x\rangle$.

2.3. Минимизация квадратичной функции

Точка x_* называется *точкой минимума* функции Q на $\mathcal{R}^N$, если $Q(x) \geq Q(x_*)$ при всех $x \in \mathcal{R}^N$.

ЛЕММА 3.1. *Если квадратичная функция Q ограничена снизу на $\mathcal{R}^N$, то она выпукла на $\mathcal{R}^N$ и система линейных уравнений*

$$Dx = -c \tag{3.1}$$

имеет решение.

Д о к а з а т е л ь с т в о. Проверим вначале совместность системы (3.1). Возьмем произвольное решение u_0 однородной сопряженной системы $D^\top u = \mathbf{0}$. В силу симметричности матрицы D

$$Du_0 = \mathbf{0}. \tag{3.2}$$

Покажем, что $\langle c, u_0\rangle = 0$. Тогда по теореме 6.3 главы 1 система (3.1) будет совместной.

Зафиксируем точку $x_0 \in \mathcal{R}^N$. Согласно лемме 2.1 и (3.2) при $t \in \mathcal{R}$ получаем

$$Q(x_0 + tu_0) - Q(x_0) = t\langle Dx_0 + c, u_0\rangle + {}^1\!/_2 t^2\langle Du_0, u_0\rangle =$$

$$= t[\langle c, u_0\rangle + \langle x_0, Du_0\rangle] = t\langle c, u_0\rangle.$$

Поскольку функция Q ограничена снизу, то необходимо $\langle c, u_0\rangle = 0$. Совместность системы (3.1) установлена.

Обозначим через x_* решение системы (3.1). Согласно лемме 2.1 при любых $h \in \mathcal{R}^N$ и $t \in \mathcal{R}$ имеем

$$Q(x_* + th) - Q(x_*) = t\langle Dx_* + c, h\rangle + {}^1\!/_2\, t^2\langle Dh, h\rangle = {}^1\!/_2\, t^2\langle Dh, h\rangle.$$

В силу ограниченности снизу функции Q отсюда следует, что $\langle Dh, h\rangle \geq 0$. По теореме 2.1 функция Q выпукла на $\mathcal{R}^N$. ■

ТЕОРЕМА 3.1. *Для того чтобы квадратичная функция Q имела точку минимума на $\mathcal{R}^N$, необходимо и достаточно, чтобы она была ограничена снизу на $\mathcal{R}^N$.*

Доказательство. Необходимость тривиальна.

Достаточность. Если функция Q ограничена снизу, то по лемме 3.1 она выпукла и существует точка x_*, удовлетворяющая системе (3.1). Покажем, что x_* — точка минимума. Согласно (2.9)

$$Q(x) - Q(x_*) \geq \langle Dx_* + c, x - x_*\rangle = 0,$$

или $Q(x) \geq Q(x_*)$ при всех $x \in \mathcal{R}^N$. Существование точки минимума установлено. ■

ТЕОРЕМА 3.2. *Пусть Q— выпуклая квадратичная функция. Для того чтобы минимальное значение Q на $\mathcal{R}^N$ достигалось в точке x_*, необходимо и достаточно, чтобы $Dx_* = -c$.*

Доказательство. Необходимость . Пусть x_* — точка минимума. Положим $h_* = Dx_* + c$. Согласно лемме 2.1

$$0 \leq Q(x_* + th_*) - Q(x_*) = t\langle Dx_* + c, Dx_* + c\rangle + {}^1\!/_2\, t^2\langle Dh_*, h_*\rangle.$$

Получили, что квадратичная функция

$$q(t) = t\, \|Dx_* + c\|^2 + {}^1\!/_2\, t^2\, \langle Dh_*, h_*\rangle$$

неотрицательна при всех $t \in \mathcal{R}$. Но тогда $\|Dx_* + c\| = 0$. Это очевидно при $\langle Dh_*, h_*\rangle = 0$ и следует из неположительности дискриминанта при $\langle Dh_*, h_*\rangle \neq 0$. Равенство же $\|Dx_* + c\| = 0$ выполняется только тогда, когда $Dx_* = -c$.

Достаточность. Если Q — выпуклая функция и $Dx_* = -c$, то согласно (2.9)

$$Q(x) - Q(x_*) \geq \langle Dx_* + c, x - x_*\rangle = 0,$$

или $Q(x) \geq Q(x_*)$ при всех $x \in \mathcal{R}^N$. Это значит, что x_* — точка минимума. ■

Рассмотрим задачу минимизации квадратичной функции на аффинном многообразии:

$$\begin{gathered} Q(x) := {}^1\!/_2\langle Dx, x\rangle + \langle c, x\rangle + \alpha \to \min, \\ Ax = b, \end{gathered} \tag{3.3}$$

где A — произвольная прямоугольная матрица. Вектор x, удовлетворяющий ограничению $Ax = b$ задачи (3.3), называется ее *планом.* Множество планов обозначим Ω. Будем считать, что Ω — непустое множество.

Положим $\Omega_0 = \{h \mid Ah = \mathbf{0}\}$.

ЛЕММА 3.2. *Если квадратичная функция Q ограничена снизу на Ω, то она выпукла на Ω и система линейных уравнений*

$$\begin{array}{lclcl} Dx & - A^\top u & = & -c, \\ Ax & & = & b \end{array} \tag{3.4}$$

имеет решение.

Доказательство. Проверим сначала совместность системы (3.4). Возьмем произвольное решение $\{v_0, z_0\}$ однородной сопряженной системы. По определению

$$\begin{pmatrix} D & A^\top \\ -A & \mathbf{0} \end{pmatrix} \begin{pmatrix} v_0 \\ z_0 \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

или $Dv_0 + A^\top z_0 = \mathbf{0}$, $Av_0 = \mathbf{0}$. Зафиксируем $x_0 \in \Omega$. Тогда

$$Q(x_0 + tv_0) - Q(x_0) = t\langle Dx_0 + c, v_0\rangle + {}^1\!/_2 t^2\langle Dv_0, v_0\rangle =$$

$$= t\langle c, v_0\rangle + t\langle x_0, Dv_0\rangle - {}^1\!/_2 t^2\langle A^\top z_0, v_0\rangle =$$

$$= t\langle c, v_0\rangle - t\langle x_0, A^\top z_0\rangle - {}^1\!/_2 t^2\langle z_0, Av_0\rangle =$$

$$= t\langle c, v_0\rangle - t\langle Ax_0, z_0\rangle = -t[\langle -c, v_0\rangle + \langle b, z_0\rangle].$$

Поскольку функция Q ограничена снизу на Ω и точка $x_0 + tv_0$ принадлежит Ω при всех $t \in \mathcal{R}$, то необходимо $\langle -c, v_0\rangle + \langle b, z_0\rangle = 0$. По теореме 6.3 главы 1 система (3.4) имеет решение.

Обозначим через $\{x_*, u_*\}$ решение системы (3.4) и возьмем $h \in \Omega_0$. Имеем

$$Q(x_* + th) - Q(x_*) = t\langle Dx_* + c, h\rangle + {}^1\!/_2 t^2\langle Dh, h\rangle.$$

Однако при $h \in \Omega_0$

$$\langle Dx_* + c, h\rangle = \langle A^\top u_*, h\rangle = \langle u_*, Ah\rangle = 0.$$

Поэтому

$$Q(x_* + th) - Q(x_*) = {}^1\!/_2 t^2\langle Dh, h\rangle.$$

В силу ограниченности снизу функции Q на Ω получаем $\langle Dh, h\rangle \geq 0$ при всех $h \in \Omega_0$, т.е. матрица D является неотрицательно определенной на Ω_0. По теореме 2.3 функция Q выпукла на Ω. ■

ТЕОРЕМА 3.3. *Для того чтобы задача* (3.3) *имела решение, необходимо и достаточно, чтобы функция Q была ограниченной снизу на Ω.*

Доказательство. Необходимость тривиальна.

Достаточность. Если квадратичная функция Q ограничена снизу на Ω, то по лемме 3.2 она выпукла на Ω и существует решение $\{x_*, u_*\}$ системы (3.4). Покажем, что x_* — точка минимума функции Q на Ω.

Возьмем произвольную точку x из Ω. Вектор $h = x - x_*$ принадлежит Ω_0. Согласно(2.9)

$$Q(x) - Q(x_*) \geq \langle Dx_* + c, x - x_*\rangle = \langle A^\top u_*, h\rangle = \langle u_*, Ah\rangle = 0.$$

Таким образом, $Q(x) \geq Q(x_*)$ при всех $x \in \Omega$. Существование точки минимума установлено. ■

ТЕОРЕМА 3.4. *Предположим, что квадратичная функция Q выпукла на Ω. Для того чтобы минимальное значение Q на Ω достигалось в точке $x_* \in \Omega$, необходимо и достаточно, чтобы нашелся вектор u_*, такой, что*

$$Dx_* + c = A^\top u_*. \tag{3.5}$$

Доказательство. Необходимость . Пусть x_* — точка минимума. Нужно установить, что система линейных уравнений

$$A^\top u = Dx_* + c \tag{3.6}$$

имеет решение. Возьмем произвольное решение v_* однородной сопряженной системы $Av = \mathbf{0}$ и покажем, что $\langle Dx_* + c, v_*\rangle = 0$.

Вектор $x_* + tv_*$ принадлежит Ω при любом $t \in \mathcal{R}$. Поэтому

$$0 \leq Q(x_* + tv_*) - Q(x_*) = t\langle Dx_* + c, v_*\rangle + {}^1\!/_2\, t^2\langle Dv_*, v_*\rangle.$$

Получили, что квадратичная функция

$$q(t) = t\langle Dx_* + c, v_*\rangle + {}^1\!/_2\, t^2\langle Dv_*, v_*\rangle$$

неотрицательна при всех $t \in \mathcal{R}$. Но тогда $\langle Dx_* + c, v_*\rangle = 0$. По теореме 6.3 главы 1 система (3.6) имеет решение.

Достаточность. Пусть Q — выпуклая на Ω функция и выполнено соотношение (3.5). Возьмем произвольный вектор x из Ω. Разность $h = x - x_*$ принадлежит Ω_0. Согласно (2.9)

$$Q(x) - Q(x_*) \geq \langle Dx_* + c, x - x_*\rangle = \langle A^\top u_*, h\rangle = \langle u_*, Ah\rangle = 0,$$

или $Q(x) \geq Q(x_*)$ при всех $x \in \Omega$. Значит, x_* — точка минимума. ■

ЗАДАЧИ

3.1. Пусть D — симметричная неотрицательно определенная матрица. Докажите, что из соотношения $\langle Dx_0, x_0\rangle = 0$ следует $Dx_0 = \mathbf{0}$.

3.2. Докажите, что для симметричной положительно определенной матрицы D справедливо соотношение

$$\min_{y\in\mathcal{R}^N}\{\langle Dy, y\rangle - 2\langle x, y\rangle\} = -\langle D^{-1}x, x\rangle.$$

3.3. Пусть Q — выпуклая квадратичная функция и x_* — ее точка минимума на $\mathcal{R}^N$. Докажите, что при всех $x \in \mathcal{R}^N$

$$Q(x) - Q(x_*) = {}^1\!/_2\langle Dx + c, x - x_*\rangle.$$

2.4. Обобщенная проблема собственных чисел

В этом параграфе удобно считать, что $N = 1\colon n$.

ТЕОРЕМА 4.1. *Пусть $D = D[N, N]$ и $B = B[N, N]$ — симметричные матрицы, причем B положительно определена. Тогда существуют вещественные числа $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ и ненулевые векторы $x_1, x_2, \ldots, x_n$, такие, что*

$$Dx_k = \lambda_k Bx_k, \quad k \in 1\colon n, \tag{4.1}$$

$$\langle Bx_i, x_j\rangle = \delta_{ij}, \quad i, j \in 1\colon n. \tag{4.2}$$

Здесь δ_{ij} — символ Кронекера ($\delta_{ij} = 1$ при $i = j$ и $\delta_{ij} = 0$ при $i \ne j$).

Доказательство. Обозначим $S = \{x \mid \langle Bx, x\rangle = 1\}$. Очевидно, что S — замкнутое множество. Покажем, что оно ограничено. Положим

$$\mu := \min_{\|x\|=1}\langle Bx, x\rangle > 0.$$

Имеем $\langle Bx, x\rangle \ge \mu\|x\|^2$ при всех $x \in \mathcal{R}^N$. Значит, $\|x\|^2 \le 1/\mu$ для $x \in S$. Это гарантирует ограниченность множества S.

Введем величину

$$\lambda_1 = \min_{x\in S}\langle Dx, x\rangle. \tag{4.3}$$

Поскольку $\langle Dx, x\rangle$ — непрерывная функция и множество S, по которому берется минимум, ограничено и замкнуто, то минимум в (4.3) достигается. Точку минимума обозначим x_1. Покажем, что $Dx_1 = \lambda_1 Bx_1$.

Имеем

$$\langle Dx, x\rangle \ge \lambda_1\langle Bx, x\rangle \quad \text{при всех } x \in \mathcal{R}^N. \tag{4.4}$$

Действительно, при $x = \mathbf{0}$ неравенство тривиально. Пусть $x \ne \mathbf{0}$. Тогда $\langle Bx, x\rangle > 0$. Вектор $\widehat{x} = x/\langle Bx, x\rangle^{1/2}$ удовлетворяет условию $\langle B\widehat{x}, \widehat{x}\rangle = 1$, т.е. принадлежит S. Поэтому для него $\langle D\widehat{x}, \widehat{x}\rangle \geq \lambda_1$, что равносильно (4.4).

Введем квадратичную функцию

$$Q_1(x) = {}^1\!/\!_2\langle (D - \lambda_1 B)x, x\rangle.$$

Согласно (4.4) имеем $Q_1(x) \geq 0$ при всех $x \in \mathcal{R}^N$. Кроме того, $Q_1(x_1) = 0$. Значит, x_1 — точка минимума функции Q_1 на $\mathcal{R}^N$. Функция Q_1, ограниченная снизу нулем, согласно лемме 3.1 является выпуклой на $\mathcal{R}^N$. По теореме 3.2 в точке минимума x_1 выполняется соотношение $(D - \lambda_1 B)x_1 = \mathbf{0}$, из которого следует, что $Dx_1 = \lambda_1 Bx_1$.

Теперь введем множество $S(x_1) = \{x \in S \,|\, \langle Bx_1, x\rangle = 0\}$ и положим

$$\lambda_2 = \min_{x \in S(x_1)} \langle Dx, x\rangle.$$

Обозначим через x_2 точку минимума. Покажем, что x_2 является решением следующей экстремальной задачи

$$\begin{gathered} Q_2(x) := {}^1\!/\!_2\langle (D - \lambda_2 B)x, x\rangle \to \min, \\ \langle Bx_1, x\rangle = 0. \end{gathered} \tag{4.5}$$

Действительно, для любого плана x задачи (4.5) выполняется неравенство

$$\langle Dx, x\rangle \geq \lambda_2 \langle Bx, x\rangle. \tag{4.6}$$

При $x = \mathbf{0}$ это тривиально. Возьмем ненулевой план x. Вектор $\widehat{x} = x/\langle Bx, x\rangle^{1/2}$ также будет планом. Кроме того, для него $\langle B\widehat{x}, \widehat{x}\rangle = 1$. Значит, $\widehat{x} \in S(x_1)$. Согласно определению λ_2 получаем $\langle D\widehat{x}, \widehat{x}\rangle \geq \lambda_2$, что равносильно (4.6).

Из (4.6) следует, что $Q_2(x) \geq 0$ на множестве планов. Вместе с тем вектор x_2 является планом задачи (4.5) и на нем $Q_2(x_2) = 0$. Значит, x_2 — решение задачи (4.5).

Функция Q_2, ограниченная снизу нулем на множестве планов Ω_2, согласно лемме 3.2 является выпуклой на Ω_2. По теореме 3.4 найдется число u_1, такое, что

$$(D - \lambda_2 B)x_2 = u_1 Bx_1.$$

Умножим это равенство скалярно на x_1. Получим

$$u_1 = \langle (D - \lambda_2 B)x_2, x_1\rangle.$$

Но $\langle Bx_2, x_1\rangle = \langle x_2, Bx_1\rangle = 0$ и

$$\langle Dx_2, x_1\rangle = \langle x_2, Dx_1\rangle = \lambda_1 \langle x_2, Bx_1\rangle = 0.$$

Значит, $u_1 = 0$. Приходим к равенству $(D - \lambda_2 B)x_2 = \mathbf{0}$, из которого следует, что $Dx_2 = \lambda_2 Bx_2$.

Этот процесс можно продолжить. Опишем p-й шаг. Пусть уже найдены числа $\lambda_1, \lambda_2, \ldots, \lambda_{p-1}$ и ненулевые векторы $x_1, x_2, \ldots, x_{p-1}$, такие, что

$$Dx_k = \lambda_k Bx_k, \qquad k \in 1\colon p-1,$$

$$\langle Bx_i, x_j \rangle = \delta_{ij}, \qquad i, j \in 1\colon p-1.$$

Введем множество $S(x_1, x_2, \ldots, x_{p-1}) = \{x \in S \mid \langle Bx_1, x\rangle = 0,\ \langle Bx_2, x\rangle = 0, \ldots, \langle Bx_{p-1}, x\rangle = 0\}$ и положим

$$\lambda_p = \min_{x \in S(x_1, x_2, \ldots, x_{p-1})} \langle Dx, x\rangle.$$

Обозначим через x_p точку минимума . Покажем, что x_p является решением следующей экстремальной задачи:

$$\begin{gathered} Q_p(x) := {}^1\!/_2 \langle (D - \lambda_p B)x, x\rangle \to \min, \\ \langle Bx_1, x\rangle = 0,\ \langle Bx_2, x\rangle = 0,\ \ldots,\ \langle Bx_{p-1}, x\rangle = 0. \end{gathered} \tag{4.7}$$

Действительно, для любого плана x задачи (4.7) согласно определению λ_p имеем

$$\langle Dx, x\rangle \ge \lambda_p \langle Bx, x\rangle.$$

Значит, $Q_p(x) \ge 0$ на планах x. Вместе с тем вектор x_p — план задачи (4.7) и $Q_p(x_p) = 0$. Получили, что x_p — решение задачи (4.7).

Функция Q_p, ограниченная снизу нулем на множестве планов Ω_p, согласно лемме 3.2 является выпуклой на Ω_p. По теореме 3.4 найдутся числа $u_1, \ldots, u_{p-1}$, такие, что

$$(D - \lambda_p B)x_p = \sum_{i=1}^{p-1} u_i Bx_i. \tag{4.8}$$

Умножив это равенство скалярно на x_j, $j \in 1\colon p-1$, получим

$$u_j = \langle (D - \lambda_p B)x_p, x_j\rangle.$$

Мы воспользовались формулой $\langle Bx_i, x_j\rangle = \delta_{ij}$. Теперь отметим, что $\langle Bx_p, x_j\rangle = \langle x_p, Bx_j\rangle = 0$ и

$$\langle Dx_p, x_j\rangle = \langle x_p, Dx_j\rangle = \lambda_j \langle x_p, Bx_j\rangle = 0.$$

Значит, $u_j = 0$ при всех $j \in 1\colon p-1$. Согласно (4.8) приходим к равенству $(D - \lambda_p B)x_p = \mathbf{0}$, из которого следует, что $Dx_p = \lambda_p Bx_p$.

Этот процесс продолжается, пока множество $S(x_1, x_2, \ldots, x_{p-1})$ будет непустым. При $p = n$ еще найдутся λ_n и x_n. Но при $p = n+1$ условия

$$\begin{gathered} \langle Bx, x\rangle = 1, \\ \langle Bx_1, x\rangle = 0, \ldots, \langle Bx_n, x\rangle = 0 \end{gathered} \tag{4.9}$$

становятся несовместными. Проверим это. Векторы $Bx_1, \ldots, Bx_n$ линейно независимы, поскольку из равенства $\sum_{k=1}^{n} c_k Bx_k = \mathbf{0}$ следует, что $\langle \sum_{k=1}^{n} c_k Bx_k, x_j \rangle = 0$, или $c_j = 0$ при всех $j \in 1{:}\, n$. Матрица системы линейных однородных уравнений

$$\langle Bx_1, x \rangle = 0, \ldots, \langle Bx_n, x \rangle = 0$$

квадратная и обратимая, поэтому система имеет только нулевое решение. Полная система уравнений (4.9) несовместна.

Таким образом, найдены вещественные числа $\lambda_1, \lambda_2, \ldots, \lambda_n$ и ненулевые векторы $x_1, x_2, \ldots, x_n$, удовлетворяющие соотношениям (4.1), (4.2). По построению $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Теорема доказана. ■

Числа λ_k называются *обобщенными собственными числами*, а соответствующие векторы x_k — *обобщенными собственными векторами* матрицы D. Обычные собственные числа и собственные векторы получаются при $B = E$.

Обозначим через P матрицу, составленную из столбцов $x_1, x_2, \ldots, x_n$. Тогда соотношения (4.1), (4.2) можно переписать в виде

$$P^\top DP = \Lambda, \quad P^\top BP = E, \tag{4.10}$$

где Λ — диагональная матрица с диагональными элементами $\lambda_1, \lambda_2, \ldots, \lambda_n$. Действительно, элемент с индексами (i, j) матрицы $P^\top DP$ равен

$$\langle x_i, Dx_j \rangle = \lambda_j \langle x_i, Bx_j \rangle = \lambda_j \delta_{ij}.$$

Тому же равен элемент с индексами (i, j) матрицы Λ. Формула $P^\top DP = \Lambda$ установлена. Попутно доказана и вторая формула $P^\top BP = E$.

Говорят, имея в виду соотношения (4.10), что матрица P одновременно приводит матрицы D и B к диагональному виду.

Установим еще одно экстремальное свойство обобщенных собственных чисел и обобщенных собственных векторов. По-прежнему через $S(a_1, \ldots, a_{p-1})$ будем обозначать множество векторов x, удовлетворяющих условиям $\langle Bx, x \rangle = 1, \langle Ba_1, x \rangle = 0, \ldots, \langle Ba_{p-1}, x \rangle = 0$.

ТЕОРЕМА 4.2 (Курант – Фишер). *Решением экстремальной задачи*

$$\min_{x \in S(a_1, \ldots, a_{p-1})} \langle Dx, x \rangle \ \to \max_{\{a_1, \ldots, a_{p-1}\}}$$

являются обобщенные собственные векторы $a_1 = x_1, \ldots, a_{p-1} = x_{p-1}$. *Максимальное значение целевой функции равно* λ_p.

Доказательство. Зафиксируем $a_1, \ldots, a_{p-1}$ и найдем вектор x_0 вида $x_0 = \sum_{i=1}^{p} c_i x_i$, принадлежащий $S(a_1, \ldots, a_{p-1})$. Условия $\langle Ba_1, x_0 \rangle = 0, \ldots, \langle Ba_{p-1}, x_0 \rangle = 0$ распишем в виде

$$\sum_{i=1}^{p} c_i \langle Ba_j, x_i \rangle = 0, \quad j \in 1{:}\, p-1. \tag{4.11}$$

У системы (4.11) $p-1$ уравнений, а неизвестных c_i — на единицу больше. По лемме 3.1 главы 1 столбцы матрицы коэффициентов системы (4.11) линейно зависимы. Коэффициенты линейной зависимости образуют ненулевое решение системы (4.11). Это решение определено с точностью до множителя, который можно подобрать так, чтобы $\sum_{i=1}^{p} c_i^2 = 1$. Тогда согласно (4.2)

$$\langle Bx_0, x_0\rangle = \left\langle \sum_{i=1}^{p} c_i Bx_i, \sum_{j=1}^{p} c_j x_j \right\rangle = \sum_{i=1}^{p} c_i^2 = 1.$$

Получили, что $x_0 \in S(a_1, \ldots, a_{p-1})$. Теперь имеем

$$\min_{x\in S(a_1,\ldots,a_{p-1})} \langle Dx, x\rangle \le \langle Dx_0, x_0\rangle =$$

$$= \left\langle \sum_{i=1}^{p} c_i Dx_i, \sum_{j=1}^{p} c_j x_j \right\rangle = \left\langle \sum_{i=1}^{p} c_i \lambda_i Bx_i, \sum_{j=1}^{p} c_j x_j \right\rangle =$$

$$= \sum_{i=1}^{p} c_i^2 \lambda_i \le \lambda_p.$$

Вместе с тем, по определению λ_p

$$\min_{x\in S(x_1,\ldots,x_{p-1})} \langle Dx, x\rangle = \lambda_p.$$

Теорема доказана. ■

ЗАДАЧИ

4.1. Пусть $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ — обобщенные собственные числа и x_1, x_2, $\ldots$, x_n — обобщенные собственные векторы матрицы D. Докажите, что
$$\lambda_p = \max_{x\in S(x_{p+1},\ldots,x_n)} \langle Dx, x\rangle.$$

4.2. Докажите, что решением экстремальной задачи
$$\max_{x\in S(a_{p+1},\ldots,a_n)} \langle Dx, x\rangle \;\to \min_{\{a_{p+1},\ldots,a_n\}}$$
являются обобщенные собственные векторы $a_{p+1} = x_{p+1}, \ldots, a_n = x_n$ и минимальное значение целевой функции равно λ_p.

2.5. Спектральное разложение симметричной матрицы

Пусть $N = 1{:}n$ и $B = E[N, N]$. Согласно теореме 4.1 для любой симметричной матрицы $D = D[N, N]$ существуют вещественные числа $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ и ненулевые векторы $x_1, x_2, \ldots, x_n$, такие, что

$$Dx_k = \lambda_k x_k, \quad k \in 1{:}n, \tag{5.1}$$

$$\langle x_i, x_j \rangle = \delta_{ij}, \quad i, j \in 1{:}n. \tag{5.2}$$

Числа λ_k называются *собственными числами*, а соответствующие векторы x_k — *собственными векторами* матрицы D. Система векторов $x_1, x_2, \ldots, x_n$, удовлетворяющая соотношению (5.2), называется *ортонормированной*, а матрица P, составленная из столбцов $x_1, x_2, \ldots, x_n$, — *ортогональной*.

Из (5.1) и (5.2) следует, что

$$P^\top D P = \Lambda, \tag{5.3}$$

$$P^\top P = E, \tag{5.4}$$

где $\Lambda = \operatorname{diag}\,(\lambda_1, \ldots, \lambda_n)$. Равенство (5.4) означает, что $(P^\top)^{-1} = P$. Переходя к обратным матрицам, получаем $P^\top = P^{-1}$, так что $PP^\top = E$. Равенство (5.3) приводит к формуле

$$D = P \Lambda P^\top. \tag{5.5}$$

Отметим, что k-й столбец матрицы $P\Lambda$ равен $\lambda_k x_k$. Поэтому согласно формуле (1.8) главы 1 и (5.5)

$$D = \sum_{k=1}^{n} \lambda_k x_k x_k^\top. \tag{5.6}$$

Соотношение (5.6) называется *спектральным разложением* матрицы D. Базисные матрицы $H_k = x_k x_k^\top$ этого разложения обладают следующими свойствами:

1) $\sum_{k=1}^{n} H_k = E$;

2) $H_k^2 = H_k$;

3) $H_i H_j = 0$ при $i \ne j$.

Свойство 1) доказывается так:

$$\sum_{k=1}^{n} H_k = \sum_{k=1}^{n} x_k x_k^\top = PP^\top = E.$$

Свойства 2) и 3) основаны на соотношении

$$H_i H_j = x_i(x_i^\top x_j)x_j^\top = \delta_{ij} x_i x_j^\top.$$

Собственные числа $\lambda_1, \dots, \lambda_n$, являющиеся коэффициентами в разложении $D = \sum_{k=1}^n \lambda_k H_k$, образуют *спектр* матрицы D.

ТЕОРЕМА 5.1. *Если все собственные числа симметричной матрицы D отличны от нуля, то матрица D обратима и*

$$D^{-1} = \sum_{k=1}^n \lambda_k^{-1} x_k x_k^\top. \tag{5.7}$$

Доказательство. В данном случае матрица $\Lambda = \mathrm{diag}\,(\lambda_1, \dots, \lambda_n)$ обратима и $\Lambda^{-1} = \mathrm{diag}\,(\lambda_1^{-1}, \dots, \lambda_n^{-1})$. Вернемся к равенству (5.5) и умножим обе его части справа последовательно на $P, \Lambda^{-1}, P^\top$. Получим

$$D(P\Lambda^{-1}P^\top) = E.$$

Это означает, что матрица D обратима и $D^{-1} = P\Lambda^{-1}P^\top$. Последнее равенство равносильно (5.7). ■

ТЕОРЕМА 5.2. *Пусть D — симметричная неотрицательно определенная матрица. Тогда матрица*

$$G = \sum_{k=1}^n \sqrt{\lambda_k}\, x_k x_k^\top$$

симметрична, неотрицательно определена и $GG = D$.

Доказательство. Напомним, что для наименьшего собственного числа λ_1 матрицы D справедлива формула

$$\lambda_1 = \min_{\|x\|=1} \langle Dx, x\rangle.$$

Если D — неотрицательно определенная матрица, то $\lambda_1 \geq 0$. Понятно, что и остальные собственные числа неотрицательны.

Проверим симметричность матрицы $H_k = x_k x_k^\top$. Имеем

$$\langle (x_k x_k^\top)u, v\rangle = \langle x_k, u\rangle \langle x_k, v\rangle,$$

$$\langle u, (x_k x_k^\top)v\rangle = \langle x_k, v\rangle \langle x_k, u\rangle.$$

Значит, для любых векторов u, v из $\mathcal{R}^N$ справедливо равенство $\langle H_k u, v\rangle = \langle u, H_k v\rangle$. По лемме 1.1 матрица H_k симметрична. Матрица G как линейная комбинация симметричных матриц также симметрична.

Поскольку

$$\langle (x_k x_k^\top)u, u\rangle = \langle x_k, u\rangle^2,$$

то матрица H_k неотрицательно определена. Матрица G как линейная комбинация с неотрицательными коэффициентами неотрицательно определенных матриц также является неотрицательно определенной.

Наконец, имеем

$$GG = \left(\sum_{i=1}^{n} \sqrt{\lambda_i}\, H_i\right) \left(\sum_{j=1}^{n} \sqrt{\lambda_j}\, H_j\right) = \sum_{i,j} \sqrt{\lambda_i \lambda_j}\, H_i H_j =$$

$$\sum_{i=1}^{n} \lambda_i H_i^2 = \sum_{i=1}^{n} \lambda_i H_i = D.$$

Теорема доказана. ■

Свойства матрицы G позволяют назвать ее *квадратным корнем* из матрицы D и обозначить $G = \sqrt{D}$ или $G = D^{1/2}$. Таким образом,

$$\sqrt{D} = \sum_{k=1}^{n} \sqrt{\lambda_k}\, x_k x_k^{\top}.$$

Для симметричной положительно определенной матрицы D (у которой наименьшее собственное число положительно) можно ввести матрицу

$$D^{-1/2} = (\sqrt{D})^{-1} = \sum_{k=1}^{n} \lambda_k^{-1/2} x_k x_k^{\top}.$$

Очевидно, что $D^{-1/2}$ — симметричная матрица. Проверим, что она положительно определена. Имеем

$$\langle D^{-1/2} u, u\rangle = \sum_{k=1}^{n} \lambda_k^{-1/2} \langle x_k, u\rangle^2 \geq 0.$$

Предположим, что $\langle D^{-1/2} u, u\rangle = 0$. Тогда $\langle x_k, u\rangle = 0$ при всех $k \in 1\colon n$. Перепишем последние равенства в матричном виде: $P^{\top} u = \mathbf{0}$. Умножив слева на P, получим $u = \mathbf{0}$. Таким образом, $\langle D^{-1/2} u, u\rangle > 0$ при $u \neq \mathbf{0}$. Положительная определенность матрицы $D^{-1/2}$ установлена.

Отметим, что $DD^{-1/2} = D^{1/2}$.

Следующая теорема дает эффектный способ ортогонализации системы линейно независимых векторов.

ТЕОРЕМА 5.3. *Пусть $A = A[M, N]$ — матрица с линейно независимыми столбцами. Тогда матрица $C = A(A^{\top}A)^{-1/2}$ удовлетворяет соотношению $C^{\top} C = E$.*

Д о к а з а т е л ь с т в о. Матрица $A^\top A$ является симметричной и положительно определенной (задача 1.2). Поэтому матрица $(A^\top A)^{-1/2}$ существует, она симметрична и положительно определена. Имеем

$$C^\top C = (A^\top A)^{-1/2} A^\top A (A^\top A)^{-1/2} = E.$$

Теорема доказана. ■

Из равенства $C = A(A^\top A)^{-1/2}$ следует, что столбцы $C_1, \ldots, C_n$ матрицы C являются линейными комбинациями столбцов $A_1, \ldots, A_n$ матрицы A, т.е.

$$C_i = \sum_{j=1}^{n} \alpha_{ij} A_j, \qquad i \in 1\colon n.$$

Вместе с тем равенство $C^\top C = E$ означает, что $\langle C_i, C_j \rangle = \delta_{ij}$ при $i, j \in 1\colon n$, т.е. векторы $C_1, \ldots, C_n$ ортонормированные. Таким образом, указано явное линейное преобразование линейно независимых векторов $A_1, \ldots, A_n$ в ортонормированные векторы $C_1, \ldots, C_n$.

Перепишем равенство $C = A(A^\top A)^{-1/2}$ в эквивалентном виде:

$$A = C(A^\top A)^{1/2}.$$

Эта формула показывает, что матрицу с линейно независимыми столбцами можно разложить в произведение двух матриц, одна из которых имеет ортонормированные столбцы, а другая является симметричной и положительно определенной.

Введем экспоненту от матрицы:

$$\exp D = E + \sum_{p=1}^{\infty} \frac{1}{p!} D^p. \tag{5.8}$$

Сходимость ряда из матриц понимается как поэлементная сходимость.

ТЕОРЕМА 5.4. *Если D — симметричная матрица, то*

$$\exp D = \sum_{k=1}^{n} \exp(\lambda_k) x_k x_k^\top. \tag{5.9}$$

Д о к а з а т е л ь с т в о. По индукции легко проверяется, что

$$D^p = \sum_{k=1}^{n} \lambda_k^p H_k.$$

Поэтому для частной суммы ряда (5.8) получаем представление

$$E + \sum_{p=1}^{s} \frac{1}{p!} D^p = \sum_{k=1}^{n} H_k + \sum_{p=1}^{s} \frac{1}{p!} \sum_{k=1}^{n} \lambda_k^p H_k =$$

$$= \sum_{k=1}^{n} \left(1 + \sum_{p=1}^{s} \frac{\lambda_k^p}{p!} \right) H_k. \tag{5.10}$$

Числовой коэффициент $\sum_{p=0}^{s} \lambda_k^p / p!$ при $s \to \infty$ стремится к $\exp(\lambda_k)$. Значит, предел правой части (5.10) равен правой части (5.9). В то же время предел левой части (5.10), по определению, равен $\exp D$. Теорема доказана. ■

ЗАДАЧИ

5.1. Докажите, что ранг симметричной матрицы D равен количеству ненулевых собственных чисел в последовательности $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$.

5.2. *Следом* квадратной матрицы называется сумма ее диагональных элементов. Докажите, что след симметричной матрицы равен сумме ее собственных чисел.

5.3. Пусть D — симметричная положительно определенная матрица, все собственные числа которой отличны от единицы. Докажите, что матрица $D+D^{-1}-2E$ является симметричной и положительно определенной.

5.4. Пусть D — симметричная матрица и

$$m\|x\|^2 \le \langle Dx, x \rangle \le M\|x\|^2 \text{ при всех } x \in \mathcal{R}^N,$$

где $m > 0$. Докажите, что

$$\|x\|^2/M \le \langle D^{-1}x, x \rangle \le \|x\|^2/m \text{ при всех } x \in \mathcal{R}^N.$$

5.5. Докажите, что в случае симметричной положительно определенной матрицы D справедливо представление

$$\tfrac{1}{2}\langle Dx, x \rangle + \langle c, x \rangle = \tfrac{1}{2}\|D^{1/2}x + D^{-1/2}c\|^2 - \tfrac{1}{2}\|D^{-1/2}c\|^2.$$

2.6. Проектирование точки на подпространство

Пусть $M = 1{:}m$, $N = 1{:}n$, $A = A[M, N]$ — матрица с линейно независимыми строками и $c \in \mathcal{R}^N$ — фиксированная точка. Рассмотрим экстремальную задачу

$$\begin{gathered} Q(x) := \tfrac{1}{2}\,\|x - c\|^2 \longrightarrow \min, \\ Ax = \mathbf{0}. \end{gathered} \tag{6.1}$$

Множество планов этой задачи обозначим Ω. Очевидно, что Ω не пусто.

Ниже будет показано, что задача (6.1) имеет единственное решение x_*. Для него

$$^1\!/_2\,\|x_* - c\|^2 \le {}^1\!/_2\,\|x - c\|^2 \quad \text{при всех } x \in \Omega. \tag{6.2}$$

Перепишем неравенство (6.2) в эквивалентном виде:

$$\|x_* - c\| \le \|x - c\| \quad \text{при всех } x \in \Omega.$$

Последнее неравенство характеризует x_* как точку из Ω, ближайшую к c (рис. 4). Говорят также, что x_* является *проекцией точки c на* Ω.

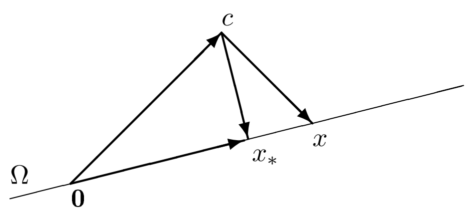

Рисунок 4.

Обратимся к целевой функции Q. Имеем

$$Q(x) = {}^1\!/_2\langle x - c, x - c\rangle = {}^1\!/_2\langle Ex, x\rangle - \langle c, x\rangle + {}^1\!/_2\,\|c\|^2.$$

Видно, что Q — квадратичная функция с матрицей $D = E$. В силу положительной определенности единичной матрицы функция Q выпукла на $\mathcal{R}^N$ и, в частности, на Ω. Наконец, значения Q ограничены снизу нулем. По теореме 3.3 функция Q имеет точку минимума на Ω. Теорема 3.4 позволяет записать критерий оптимальности для точки минимума:

$$\begin{aligned} x - c &= A^\top u, \\ Ax &= \mathbf{0}. \end{aligned} \tag{6.3}$$

Найдем решение системы (6.3). Для этого умножим левую и правую части первого равенства слева на A. Учитывая второе равенство, получаем

$$-Ac = AA^\top u.$$

По условию столбцы матрицы $A^\top$ линейно независимы, поэтому согласно теореме 4.2 главы 1 матрица $AA^\top = (A^\top)^\top A^\top$ обратима. Значит, $u = -(AA^\top)^{-1}Ac$ и

$$x = c + A^\top u = c - A^\top(AA^\top)^{-1}Ac = (E - A^\top(AA^\top)^{-1}A)c. \tag{6.4}$$

Получили единственное решение системы (6.3). Введем матрицу

$$P = E - A^\top(AA^\top)^{-1}A. \tag{6.5}$$

С ее помощью формулу (6.4) можно переписать в виде $x = Pc$.

Подведем итог:

ТЕОРЕМА 6.1. *Задача* (6.1) *имеет единственное решение* $x_* = Pc$.

Матрица P вида (6.5) называется *матрицей ортогонального проектирования* на подпространство $\Omega = \{\, x \,|\, Ax = \mathbf{0}\}$.

ТЕОРЕМА 6.2. *Матрица ортогонального проектирования P обладает следующими свойствами*:

1) $PA^\top = \mathbf{0}$, $PP = P$;

2) P — *симметричная неотрицательно определенная матрица*;

3) $\operatorname{rank} P = n - m$;

4) P *имеет* m *собственных чисел* $\lambda_1, \lambda_2, \ldots, \lambda_m$, *равных нулю, и* $n - m$ *собственных чисел* $\lambda_{m+1}, \ldots, \lambda_n$, *равных единице.*

Доказательство. Равенство $PA^\top = \mathbf{0}$ непосредственно следует из (6.5). Далее

$$PP = P(E - A^\top(AA^\top)^{-1}A) = P - (PA^\top)(AA^\top)^{-1}A = P.$$

Матрица $AA^\top$ симметрична, поэтому и обратная матрица $(AA^\top)^{-1}$ симметрична. Имеем

$$\langle A^\top(AA^\top)^{-1}Ax, y\rangle = \langle (AA^\top)^{-1}Ax, Ay\rangle = \langle Ax, (AA^\top)^{-1}Ay\rangle =$$

$$= \langle x, A^\top(AA^\top)^{-1}Ay\rangle.$$

Согласно лемме 1.1 матрица $A^\top(AA^\top)^{-1}A$ симметрична. Значит, симметрична и матрица P.

Для проверки неотрицательной определенности P воспользуемся симметричностью P и равенством $PP = P$. Получим

$$\langle Px, x\rangle = \langle PPx, x\rangle = \langle Px, Px\rangle \geq 0.$$

Свойства 1) и 2) установлены. Переходим к более глубоким свойствам 3) и 4).

Система линейных однородных уравнений $Px = \mathbf{0}$ имеет m линейно независимых решений — это столбцы $A_1^\top, \ldots, A_m^\top$ матрицы $A^\top$. Отметим, что столбцы матрицы $A^\top$ образуют базис подпространства $\mathcal{P} = \{x \,|\, Px = \mathbf{0}\}$. Действительно, для любого $x \in \mathcal{P}$ имеем согласно (6.5)

$$\mathbf{0} = Px = x - A^\top(AA^\top)^{-1}Ax = x - A^\top a,$$

т.е. $x = A^\top a$. Последнее означает, что x является линейной комбинацией столбцов матрицы $A^\top$. По основной лемме линейной алгебры в $\mathcal{P}$ не может быть более m линейно независимых векторов. Но столько есть

— это столбцы матрицы $A^\top$. По определению размерности линейного множества $\dim \mathcal{P} = m$. В силу теоремы 6.1 главы 1

$$\operatorname{rank} P = n - m.$$

От базиса $A_1^\top, \ldots, A_m^\top$ линейного множества $\mathcal{P}$ перейдем к ортонормированному базису. Для этого введем матрицу

$$X = A^\top (AA^\top)^{-1/2}.$$

По теореме 5.3 столбцы $x_1, \ldots, x_m$ матрицы X ортонормированные. Поскольку они являются линейными комбинациями столбцов $A_1^\top, \ldots, A_m^\top$ матрицы $A^\top$, то

$$Px_k = \mathbf{0}, \qquad k \in 1\colon m. \tag{6.6}$$

Теперь введем линейное множество $\mathcal{L} = \{x \mid Ax = \mathbf{0}\}$. По теореме 6.1 главы 1 $\dim \mathcal{L} = n - m$. Возьмем $n - m$ линейно независимых векторов $z_{m+1}, \ldots, z_n$ из $\mathcal{L}$. Согласно (6.5) получаем $Pz_k = z_k$, $k \in m+1\colon n$. Как и в случае множества $\mathcal{P}$, от базиса $z_{m+1}, \ldots, z_n$ перейдем к ортонормированному базису $x_{m+1}, \ldots, x_n$, так что

$$x_k = \sum_{j=m+1}^{n} \alpha_{kj} z_j, \qquad k \in m+1\colon n.$$

Но тогда

$$Px_k = \sum_{j=m+1}^{n} \alpha_{kj} P z_j = \sum_{j=m+1}^{n} \alpha_{kj} z_j = x_k,$$

т.е.

$$Px_k = x_k, \qquad k \in m+1\colon n. \tag{6.7}$$

Совокупность векторов $x_1, \ldots, x_m, x_{m+1}, \ldots, x_n$ удовлетворяет соотношению

$$\langle x_i, x_j \rangle = \delta_{ij}, \qquad i, j \in 1\colon n. \tag{6.8}$$

В проверке нуждается лишь ортогональность x_i и x_j при $i \in 1\colon m$ и $j \in m+1\colon n$. Но она легко следует из (6.6) и (6.7). Действительно,

$$\langle x_i, x_j \rangle = \langle x_i, Px_j \rangle = \langle Px_i, x_j \rangle = 0.$$

На основании (6.6) — (6.8) заключаем, что матрица P имеет собственные числа $\lambda_1 = \ldots = \lambda_m = 0$, $\lambda_{m+1} = \ldots = \lambda_n = 1$ и соответствующие собственные векторы $x_1, \ldots, x_m, x_{m+1}, \ldots, x_n$. Теорема доказана. ∎

ЗАДАЧИ

Во всех трех задачах w — единичный вектор, $\|w\| = 1$.

6.1. Пусть $\Omega = \{\, x \mid \langle w, x \rangle = 0 \}$. Докажите, что матрица ортогонального проектирования на подпространство Ω имеет вид $P = E - ww^\top$. (Такая матрица появлялась в параграфе 1.7.)

6.2. Матрица $V = E - 2ww^\top$ называется *матрицей отражения* . Найдите ее спектр.

6.3. Матрица $B = E + \gamma ww^\top$ при $\gamma > 0$ называется *матрицей растяжения пространства вдоль вектора* w. Найдите ее спектр.

2.7. Сингулярное разложение прямоугольной матрицы

В этом параграфе будет активно использоваться индексная техника.

Пусть задана ненулевая прямоугольная матрица $A = A[M, N]$ ранга r, у которой $M = 1\colon m$, $N = 1\colon n$. Наряду с матрицей A рассмотрим симметричную неотрицательно определенную матрицу $AA^\top$. Собственные числа $AA^\top$ неотрицательны, что позволяет обозначить их ϑ_k^2. Сами ϑ_k будем считать неотрицательными. Упорядочим ϑ_k^2 в порядке невозрастания

$$\vartheta_1^2 \geq \vartheta_2^2 \geq \ldots \geq \vartheta_m^2. \tag{7.1}$$

Соответствующие ортонормированные собственные векторы обозначим $\varphi_1, \varphi_2, \ldots, \varphi_m$. Из столбцов $\varphi_1, \varphi_2, \ldots, \varphi_m$ составим матрицу $\Phi = \Phi[M, M]$. Формулы (5.3), (5.4) в данном случае примут вид

$$\Phi^\top AA^\top \Phi = \Theta^2, \qquad \Phi^\top \Phi = E,$$

где $\Theta = \Theta[M, M]$ — диагональная матрица с диагональными элементами $\vartheta_1, \vartheta_2, \ldots, \vartheta_m$. По теореме 4.3 главы 1 $\operatorname{rank} AA^\top = r$, а согласно теореме 4.4 главы 1

$$\operatorname{rank} \Theta^2 = \operatorname{rank} \Phi^\top AA^\top \Phi = \operatorname{rank} AA^\top = r.$$

Таким образом, $\operatorname{rank} \Theta^2 = r$. Но ранг диагональной матрицы равен числу ненулевых диагональных элементов. Значит, только r собственных чисел ϑ_k^2 отличны от нуля. Это дает возможность уточнить соотношения (7.1):

$$\vartheta_1^2 \geq \vartheta_2^2 \geq \ldots \geq \vartheta_r^2 > 0 = \vartheta_{r+1}^2 = \ldots = \vartheta_m^2.$$

Положим $R = 1\colon r$. Согласно (5.6)

$$AA^\top = \sum_{k=1}^{m} \vartheta_k^2 \varphi_k \varphi_k^\top = \sum_{k=1}^{r} \vartheta_k^2 \varphi_k \varphi_k^\top =$$
$$= \Phi[M,R] \times \Theta^2[R,R] \times \Phi^\top[R,M]. \qquad (7.2)$$

Умножим обе части этого равенства справа на $\Phi[M,R]\times\Theta^{-1}[R,R]$. Учитывая, что $\Phi^\top[R,M] \times \Phi[M,R] = E[R,R]$, получаем

$$A[M,N] \times A^\top[N,M] \times \Phi[M,R] \times \Theta^{-1}[R,R] = \Phi[M,R] \times \Theta[R,R]. \quad (7.3)$$

Введем обозначение

$$\Psi[N,R] = A^\top[N,M] \times \Phi[M,R] \times \Theta^{-1}[R,R].$$

Тогда формулу (7.3) можно переписать в виде

$$A[M,N] \times \Psi[N,R] = \Phi[M,R] \times \Theta[R,R]. \qquad (7.4)$$

Покажем, что столбцы матрицы $\Psi[N,R]$ ортонормированные. Имеем согласно (7.2)

$$\Psi^\top[R,N] \times \Psi[N,R] = \Theta^{-1}[R,R] \times \Phi^\top[R,M] \times (A[M,N] \times A^\top[N,M])\times$$
$$\times\Phi[M,R] \times \Theta^{-1}[R,R] = \Theta^{-1}[R,R] \times (\Phi^\top[R,M] \times \Phi[M,R]) \times \Theta^2[R,R]\times$$
$$\times(\Phi^\top[R,M] \times \Phi[M,R]) \times \Theta^{-1}[R,R] = E[R,R].$$

Ортонормированность столбцов $\psi_1, \psi_2, \dots, \psi_r$ матрицы $\Psi[N,R]$ установлена.

Теперь рассмотрим симметричную неотрицательно определенную матрицу $A^\top A$. Докажем, что векторы ψ_k являются ее собственными векторами. Согласно определению $\Psi[N,R]$ и формуле (7.2) имеем

$$A^\top[N,M] \times A[M,N] \times \Psi[N,R] = A^\top[N,M] \times (A[M,N] \times A^\top[N,M])\times$$
$$\times\Phi[M,R] \times \Theta^{-1}[R,R] = A^\top[N,M] \times \Phi[M,R] \times \Theta^2[R,R] \times (\Phi^\top[R,M]\times$$
$$\times\Phi[M,R]) \times \Theta^{-1}[R,R] = A^\top[N,M] \times \Phi[M,R] \times \Theta[R,R] =$$
$$= \Psi[N,R] \times \Theta^2[R,R].$$

Перепишем это матричное равенство как равенство столбцов

$$(A^\top A)\psi_k = \vartheta_k^2 \psi_k, \qquad k \in 1\colon r. \qquad (7.5)$$

Последнее и означает, что ψ_k — собственный вектор матрицы $A^\top A$, соответствующий собственному числу ϑ_k^2.

По теореме 4.3 главы 1 $\operatorname{rank} A^\top A = r$. Отсюда следует, что матрица $A^\top A$ имеет ровно r ненулевых собственных чисел. Но они нам известны — это $\vartheta_1^2, \dots, \vartheta_r^2$. Остальные $n - r$ собственных чисел равны нулю. Найдем соответствующие собственные векторы.

Возьмем $n-r$ линейно независимых векторов, образующих базис подпространства $\mathcal{L}=\{x \mid Ax=\mathbf{0}\}$. Обозначим их $x_{r+1},\dots,x_n$. С помощью приема, указанного в теореме 5.3, перейдем к ортонормированному базису $\psi_{r+1},\dots,\psi_n$, так что

$$\psi_k=\sum_{j=r+1}^{n}\alpha_{kj}x_j, \qquad k\in r+1\colon n.$$

Поскольку $A\psi_k=\mathbf{0}$, то и

$$(A^\top A)\psi_k=\mathbf{0}, \qquad k\in r+1\colon n. \tag{7.6}$$

Совокупность векторов $\psi_1,\dots,\psi_r,\psi_{r+1},\dots,\psi_n$ удовлетворяет соотношению

$$\langle\psi_i,\psi_j\rangle=\delta_{ij}, \qquad i,j\in 1\colon n. \tag{7.7}$$

В проверке нуждается лишь ортогональность ψ_i и ψ_j при $i\in 1\colon r$ и $j\in r+1\colon n$. Но она легко следует из (7.5) и (7.6). Действительно,

$$\langle\psi_i,\psi_j\rangle=\vartheta_i^{-2}\langle(A^\top A)\psi_i,\psi_j\rangle=\vartheta_i^{-2}\langle\psi_i,(A^\top A)\psi_j\rangle=0.$$

На основании (7.5)–(7.7) заключаем, что матрица $A^\top A$ имеет собственные числа

$$\vartheta_1^2\ge\vartheta_2^2\ge\dots\ge\vartheta_r^2>0=\vartheta_{r+1}^2=\dots=\vartheta_n^2$$

и соответствующие собственные векторы $\psi_1,\psi_2,\dots,\psi_r,\psi_{r+1},\dots,\psi_n$.

Матрицу $\Psi[N,R]$ дополним столбцами $\psi_{r+1},\dots,\psi_n$ до ортогональной матрицы $\Psi=\Psi[N,N]$. Отметим, что по построению

$$A[M,N]\times\Psi[N,N\setminus R]=\mathbf{0}[M,N\setminus R]. \tag{7.8}$$

Формулы (7.4) и (7.8) можно объединить. Для этого положим $l=\min\{m,n\}$ и введем диагональную матрицу $\Theta=\Theta[M,N]$ с элементами $\Theta[i,i]=\Theta_i$ при $i\in 1\colon l$ и $\Theta[i,j]=0$ при $i\neq j$. На рис. 5 схематично представлена матрица Θ при $m\le n$ и $m>n$.

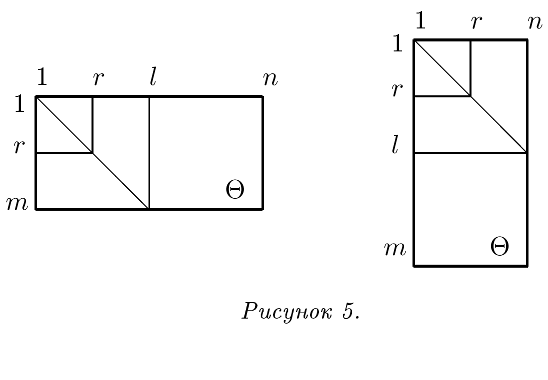

Рисунок 5.

Формулы (7.4) и (7.8) объединяются так:

$$A[M,N] \times \Psi[N,N] = \Phi[M,M] \times \Theta[M,N]. \tag{7.9}$$

Действительно,

$$\Phi[M,M] \times \Theta[M,R] = \sum_{i \in M} \Phi[M,i] \times \Theta[i,R] = \sum_{i \in R} \Phi[M,i] \times \Theta[i,R] =$$

$$= \Phi[M,R] \times \Theta[R,R] = A[M,N] \times \Psi[N,R],$$

$$\Phi[M,M] \times \Theta[M,N \setminus R] = \sum_{i \in M} \Phi[M,i] \times \Theta[i,N \setminus R] = \mathbf{0}[M,N \setminus R] =$$

$$= A[M,N] \times \Psi[N,N \setminus R].$$

Умножим обе части равенства (7.9) справа на Ψ^{-1}. Учитывая, что $\Psi^{-1} = \Psi^\top$, получаем

$$A[M,N] = \Phi[M,M] \times \Theta[M,N] \times \Psi^\top[N,N]. \tag{7.10}$$

Соотношение (7.10) называется *сингулярным разложением* матрицы A, числа $\vartheta_1 \geq \vartheta_2 \geq \ldots \geq \vartheta_r > 0 = \vartheta_{r+1} = \ldots = \vartheta_l$ — *сингулярными числами*, а ортонормированные столбцы матриц Φ и Ψ — *сингулярными базисами*.

Отметим, что определяющим является соотношение (7.9). Если обе его части транспонировать, а затем умножить слева на Ψ и справа на Φ, то придем к эквивалентному соотношению

$$A^\top[N,M] \times \Phi[M,M] = \Psi[N,N] \times \Theta^\top[N,M].$$

ЗАДАЧИ

7.1. Проверьте, что из формулы (7.10) следует равенство

$$A = \sum_{k=1}^{r} \vartheta_k \varphi_k \psi_k^\top.$$

7.2. Пусть имеется представление

$$A[M,N] = \Phi[M,M] \times \Theta[M,N] \times \Psi^\top[N,N],$$

в котором Φ и Ψ — некоторые ортогональные матрицы и Θ — диагональная матрица указанного в основном тексте вида. Докажите, что столбцы матриц Φ и Ψ образуют полные системы собственных векторов матриц $AA^\top$ и $A^\top A$ соответственно.

7.3. Докажите, что произвольную квадратную матрицу A можно представить в виде $A = PD$, где P — ортогональная матрица, а D — симметричная неотрицательно определенная матрица. (Такое представление называется *полярным разложением* матрицы A.)

7.4. Проверьте, что симметричная неотрицательно определенная матрица D из предыдущей задачи единственна.

2.8. Наилучшее приближение прямоугольной матрицы матрицами меньшего ранга

Для матрицы $A = A[M, N]$ введем евклидову норму

$$\|A\| = \left(\sum_{i\in M} \sum_{j\in N} (A[i,j])^2 \right)^{1/2}.$$

Рассмотрим квадратную матрицу $AA^\top$ и найдем сумму ее диагональных элементов:

$$\sum_{i\in M} A[i,N] \times A^\top[N,i] = \sum_{i\in M} \left(\sum_{j\in N} A[i,j] \times A^\top[j,i] \right) =$$

$$= \sum_{i\in M} \sum_{j\in N} (A[i,j])^2 = \|A\|^2.$$

Сумма диагональных элементов квадратной матрицы C называется ее *следом* и обозначается $\mathrm{tr}\ C$. Таким образом, мы получили, что

$$\|A\|^2 = \mathrm{tr}\ (AA^\top). \tag{8.1}$$

ЛЕММА 8.1. *Для матриц A_1, A_2 из $\mathcal{R}^{M,N}$ справедлива формула*

$$\|A_1 + A_2\|^2 = \|A_1\|^2 + \|A_2\|^2 + 2\mathrm{tr}\ (A_1 A_2^\top). \tag{8.2}$$

Доказательство. Согласно (8.1) имеем

$$\|A_1 + A_2\|^2 = \mathrm{tr}\ \left[(A_1 + A_2)(A_1^\top + A_2^\top) \right] =$$

$$= \|A_1\|^2 + \|A_2\|^2 + \mathrm{tr}\ (A_1 A_2^\top) + \mathrm{tr}\ (A_2 A_1^\top). \tag{8.3}$$

След матрицы $A_2 A_1^\top$ совпадает со следом транспонированной матрицы, равной $A_1 A_2^\top$, что позволяет переписать формулу (8.3) в виде (8.2). ■

Установим одно свойство следа матрицы, которое потребуется в дальнейшем.

ЛЕММА 8.2. *Пусть $A = A[M, N]$ и $B = B[N, M]$. Тогда*

$$\mathrm{tr}\ (AB) = \mathrm{tr}\ (BA).$$

Доказательство. Имеем

$$\mathrm{tr}\ (AB) = \sum_{i\in M} A[i,N] \times B[N,i] = \sum_{i\in M} \sum_{j\in N} A[i,j] \times B[j,i]. \tag{8.4}$$

$$\mathrm{tr}\ (BA) = \sum_{j\in N} B[j,M]\times A[M,j] = \sum_{j\in N}\sum_{i\in M} B[j,i]\times A[i,j]. \qquad (8.5)$$

В правых частях формул (8.4) и (8.5) стоят повторные суммы, различающиеся лишь порядком суммирования. Они равны. Значит, равны и левые части. ■

Согласно лемме 8.2, для $\|A\|^2$ наряду с (8.1) справедливо другое представление

$$\|A\|^2 = \mathrm{tr}\ (A^\top A). \qquad (8.6)$$

Множество прямоугольных матриц $A = A[M,N]$ ранга r обозначим $\mathcal{R}_r^{M,N}$. Возьмем $A\in\mathcal{R}_r^{M,N}$, зафиксируем натуральное $p<r$ и рассмотрим задачу о наилучшем приближении матрицы A матрицей ранга p. С учетом теоремы 5.1 главы 1 эту задачу можно формализовать следующим образом:

$$\varepsilon_p(X,Y) := \|A - XY^\top\|^2 \longrightarrow \min_{X\in\mathcal{R}_p^{M,P},\, Y\in\mathcal{R}_p^{N,P}}, \qquad (8.7)$$

где $|P| = p$. Для удобства будем считать, что $M = 1{:}\,m$, $N = 1{:}\,n$, $P = 1{:}\,p$.

Отметим, что в (8.7) достаточно рассматривать матрицы $Y = Y[N,P]$, удовлетворяющие условию

$$Y^\top[P,N]\times Y[N,P] = E[P,P], \qquad (8.8)$$

т.е. матрицы с ортонормированными столбцами. Действительно, пусть $Y\in\mathcal{R}_p^{N,P}$. Поскольку столбцы этой матрицы линейно независимы, то по теореме 5.3 существует $(Y^\top Y)^{-1/2}$ и матрица $\hat{Y} = Y(Y^\top Y)^{-1/2}$ удовлетворяет условию (8.8). Матрице $X\in\mathcal{R}_p^{M,P}$ сопоставим матрицу $\hat{X} = X(Y^\top Y)^{1/2}$. Получим

$$\hat{X}\hat{Y}^\top = X(Y^\top Y)^{1/2}(Y^\top Y)^{-1/2}\,Y^\top = XY^\top.$$

По теореме 4.4 главы 1

$$\mathrm{rank}\ \hat{X} = \mathrm{rank}\ X = p, \qquad \mathrm{rank}\ \hat{Y} = \mathrm{rank}\ Y = p.$$

Таким образом, плану (X,Y) задачи (8.7) мы сопоставили другой план $(\hat{X},\hat{Y})$ с матрицей $\hat{Y}$, удовлетворяющей условию (8.8), и с тем же значением целевой функции ε_p. С точки зрения минимизации ε_p, этими планами можно и ограничиться. Более того, предположение о ранге $\hat{Y}$ можно опустить, поскольку согласно теореме 3.2 главы 1 ранг матрицы, удовлетворяющей условию (8.8), автоматически равен p.

ЛЕММА 8.3. *Пусть* $Y_0 = Y_0[N,P]$ *— матрица, удовлетворяющая условию* (8.8). *Тогда*

$$\min_{X\in\mathcal{R}^{M,P}} \|A - XY_0^\top\|^2 = \|A\|^2 - \|AY_0\|^2, \qquad (8.9)$$

причем минимум достигается на единственной матрице $X_0 = AY_0$.

Подчеркнем, что минимум в левой части (8.9) берется по множеству $\mathcal{R}^{M,P}$, а не только по $\mathcal{R}_p^{M,P}$.

Доказательство. Обозначим $f(X) = \|A - XY_0^\top\|^2$. Согласно лемме 8.1 имеем

$$f(X+H) = \|A - (X+H)Y_0^\top\|^2 = \|(A - XY_0^\top) - HY_0^\top\|^2 =$$

$$\|A - XY_0^\top\|^2 + \|HY_0^\top\|^2 - 2\mathrm{tr}\ [(A - XY_0^\top)Y_0H^\top].$$

Поскольку $Y_0^\top Y_0 = E$, то

$$\|HY_0^\top\|^2 = \mathrm{tr}\ (HY_0^\top Y_0H^\top) = \|H\|^2,$$

$$\mathrm{tr}\ [(A - XY_0^\top)Y_0H^\top] = \mathrm{tr}\ [(AY_0 - X)H^\top].$$

Значит,

$$f(X+H) = f(X) - 2\mathrm{tr}\ [(AY_0 - X)H^\top] + \|H\|^2. \tag{8.10}$$

Положим $X_0 = AY_0$. При $X = X_0$ формула (8.10) принимает вид

$$f(X_0 + H) = f(X_0) + \|H\|^2.$$

Отсюда очевидным образом следует, что X_0 является единственной точкой минимума функции $f(X)$ на $\mathcal{R}^{M,P}$.

Вычислим значение $f(X_0)$. Имеем

$$f(X_0) = \|A - AY_0Y_0^\top\|^2 = \|A\|^2 + \|AY_0Y_0^\top\|^2 - 2\mathrm{tr}\ (AY_0Y_0^\top A^\top).$$

Но $\|AY_0Y_0^\top\|^2 = \mathrm{tr}\ [AY_0(Y_0^\top Y_0)Y_0^\top A^\top] = \mathrm{tr}\ (AY_0Y_0^\top A^\top)$ и $\mathrm{tr}\ (AY_0Y_0^\top A^\top) = \|AY_0\|^2$, так что

$$f(X_0) = \|A\|^2 - \|AY_0\|^2.$$

Лемма доказана. ■

На пути к решению основной задачи (8.7) нам придется рассмотреть вспомогательную задачу максимизации $\|AY\|^2$ при ограничении (8.8). Отметим, что согласно (8.6)

$$\|AY\|^2 = \mathrm{tr}\ (Y^\top A^\top AY).$$

Введем матрицу $D = A^\top A$ и обозначим через $y_1, \ldots, y_p$ столбцы матрицы Y. Тогда вспомогательную задачу можно записать так:

$$\begin{aligned} &\textstyle\sum_{k=1}^p \langle y_k, Dy_k\rangle \longrightarrow \max, \\ &\langle y_i, y_j\rangle = \delta_{ij}, \qquad i, j \in 1\colon p. \end{aligned} \tag{8.11}$$

Нам потребуются собственные числа и собственные векторы симметричной матрицы $D = A^\top A$. Так же, как в предыдущем параграфе, обозначим их через $\vartheta_1^2 \geq \vartheta_2^2 \geq \ldots \geq \vartheta_r^2 > 0 = \vartheta_{r+1}^2 = \ldots = \vartheta_n^2$ и $\psi_1, \psi_2, \ldots, \psi_r, \psi_{r+1}, \ldots, \psi_n$ соответственно.

ЛЕММА 8.4. *Решением задачи* (8.11) *являются векторы* $y_1 = \psi_1, \ldots, y_p = \psi_p$. *Наибольшее значение целевой функции равно* $\sum_{k=1}^p \vartheta_k^2$.

Если обозначить через Y_* матрицу со столбцами $\psi_1, \ldots, \psi_p$, то согласно лемме 8.4 получим

$$\max_{Y^\top Y = E} \|AY\|^2 = \|AY_*\|^2 = \sum_{k=1}^p \vartheta_k^2. \tag{8.12}$$

Доказательство. Возьмем план $(y_1, \ldots, y_p)$ задачи (8.11). Разложим вектор y_k по базису $\psi_1, \ldots, \psi_n$:

$$y_k = \sum_{j=1}^n c_{kj} \psi_j, \qquad k \in 1\colon p.$$

В силу ортонормированности базиса имеем $\langle y_k, \psi_i \rangle = c_{ki}$ и

$$1 = \langle y_k, y_k \rangle = \left\langle \sum_{i=1}^n c_{ki} \psi_i, \sum_{j=1}^n c_{kj} \psi_j \right\rangle = \sum_{i=1}^n c_{ki}^2,$$

так что $\sum_{i=1}^n c_{ki}^2 = 1$ при всех $k \in 1\colon p$. Обозначим $\tau_i = \sum_{k=1}^p c_{ki}^2$ и покажем, что $\tau_i \leq 1$ при $i \in 1\colon n$. Поскольку векторы $y_1, \ldots, y_p$ тоже ортонормированные, то

$$0 \leq \|\psi_i - \sum_{k=1}^p \langle y_k, \psi_i \rangle y_k\|^2 = \|\psi_i\|^2 - \sum_{k=1}^p \langle y_k, \psi_i \rangle^2.$$

Отсюда следует, что при всех $i \in 1\colon n$

$$\tau_i = \sum_{k=1}^p c_{ki}^2 = \sum_{k=1}^p \langle y_k, \psi_i \rangle^2 \leq \|\psi_i\|^2 = 1,$$

т.е. $\tau_i \leq 1$. Отметим также, что

$$\sum_{i=1}^n \tau_i = \sum_{i=1}^n \sum_{k=1}^p c_{ki}^2 = \sum_{k=1}^p \sum_{i=1}^n c_{ki}^2 = p.$$

Поэтому

$$\sum_{i=p+1}^n \tau_i = p - \sum_{i=1}^p \tau_i = \sum_{i=1}^p (1 - \tau_i).$$

Обратимся к целевой функции задачи (8.11). Учитывая, что $D\psi_j = \vartheta_j^2 \psi_j$, получаем

$$\sum_{k=1}^p \langle y_k, Dy_k \rangle = \sum_{k=1}^p \left\langle \sum_{j=1}^n c_{kj} \psi_j, \sum_{i=1}^n \vartheta_i^2 c_{ki} \psi_i \right\rangle =$$

$$= \sum_{k=1}^{p} \sum_{i=1}^{n} \vartheta_i^2 c_{ki} = \sum_{i=1}^{n} \vartheta_i^2 \tau_i \leq \sum_{i=1}^{p} \vartheta_i^2 \tau_i + \vartheta_{p+1}^2 \sum_{i=p+1}^{n} \tau_i =$$

$$= \sum_{i=1}^{p} \vartheta_i^2 \tau_i + \vartheta_{p+1}^2 \sum_{i=1}^{p} (1-\tau_i) \leq \sum_{i=1}^{p} \vartheta_i^2 \tau_i + \sum_{i=1}^{p} \vartheta_i^2 (1-\tau_i) = \sum_{i=1}^{p} \vartheta_i^2.$$

Таким образом, $\sum_{k=1}^{p} \langle y_k, D y_k \rangle \leq \sum_{k=1}^{p} \vartheta_k^2$ для всех планов $(y_1, \ldots, y_p)$ задачи (8.11). Вместе с тем,

$$\sum_{k=1}^{p} \langle \psi_k, D\psi_k \rangle = \sum_{k=1}^{p} \vartheta_k^2 \langle \psi_k, \psi_k \rangle = \sum_{k=1}^{p} \vartheta_k^2.$$

Лемма доказана. ■

Обратимся к задаче (8.7).

ТЕОРЕМА 8.1. *Минимум функции ε_p доставляет пара (X_*, Y_*), где $X_* = AY_*$ и Y_* — матрица, столбцами которой являются ортонормированные собственные векторы $\psi_1, \ldots, \psi_p$ матрицы $A^\top A$, соответствующие наибольшим собственным числам $\vartheta_1^2, \ldots, \vartheta_p^2$. При этом*

$$\varepsilon_p(X_*, Y_*) = \sum_{k=p+1}^{r} \vartheta_k^2.$$

Доказательство. Прежде всего покажем, что rank $X_* = p$. Согласно теореме 3.2 главы 1, rank $X_* \leq \min\{\text{rank } A, \text{rank } Y_*\} = p$, т.е. rank $X_* \leq p$. Далее $A^\top X_* = A^\top A Y_* = Y_* \Theta^2$, где $\Theta^2 = \text{diag } (\vartheta_1^2, \ldots, \vartheta_p^2)$. По той же теореме 3.2 главы 1 rank $(Y_* \Theta^2) \leq$ rank X_*. Но rank $(Y_* \Theta^2) = p$. Значит, rank $X_* \geq p$. Объединяя два неравенства, приходим к равенству rank $X_* = p$.

Теперь возьмем план (X_0, Y_0) задачи (8.7), где $X_0 \in \mathcal{R}_p^{M,P}$ и $Y_0 = Y_0[N, P]$ — матрица, удовлетворяющая условию (8.8). Согласно (8.9) и (8.12) имеем

$$\varepsilon_p(X_0, Y_0) = \|A - X_0 Y_0^\top\|^2 \geq \min_{X \in \mathcal{R}^{M,P}} \|A - X Y_0^\top\|^2 =$$

$$= \|A\|^2 - \|AY_0\|^2 \geq \|A\|^2 - \|AY_*\|^2 = \|A\|^2 - \sum_{k=1}^{p} \vartheta_k^2.$$

Вместе с тем

$$\varepsilon_p(X_*, Y_*) = \|A - AY_* Y_*^\top\|^2 = \|A\|^2 - \|AY_*\|^2 = \|A\|^2 - \sum_{k=1}^{p} \vartheta_k^2.$$

Получили, что наименьшее значение функции ε_p равно $\|A\|^2 - \sum_{k=1}^{p} \vartheta_k^2$ и достигается на паре (X_*, Y_*). Остается проверить, что

$$\|A\|^2 = \sum_{k=1}^{r} \vartheta_k^2. \tag{8.13}$$

Обозначим через Ψ ортогональную матрицу, составленную из столбцов $\psi_1, \ldots, \psi_n$. Тогда $\Psi\Psi^\top = E$ и $\Psi^\top A^\top A \Psi = \Theta^2$, где $\Theta = \mathrm{diag}\,(\vartheta_1, \ldots, \vartheta_n)$. Согласно лемме 8.2

$$\mathrm{tr}\;\Theta^2 = \mathrm{tr}\;(\Psi^\top A^\top A \Psi) = \mathrm{tr}\;(\Psi\Psi^\top A^\top A) = \mathrm{tr}\;(A^\top A) = \|A\|^2.$$

Но $\mathrm{tr}\;\Theta^2 = \sum_{k=1}^n \vartheta_k^2 = \sum_{k=1}^r \vartheta_k^2$. Соотношение (8.13) установлено. ■

Задачу (8.7) можно интерпретировать как задачу проектирования матрицы $A \in \mathcal{R}_r^{M,N}$ на множество $\mathcal{R}_p^{M,N}$. В теореме 8.1 для проекции Z_* указано представление $Z_* = X_* Y_*^\top = A Y_* Y_*^\top$. Поскольку $Y_* = \Psi[N, P]$, то

$$Z_* = A[M, N] \times \Psi[N, P] \times \Psi^\top[P, N]. \tag{8.14}$$

Воспользуемся формулой (7.4), где $\Phi[M, R]$ — матрица, столбцами которой являются собственные векторы $\varphi_1, \ldots, \varphi_r$ матрицы $AA^\top$, соответствующие собственным числам $\vartheta_1^2, \ldots, \vartheta_r^2$. Получим

$$A[M, N] \times \Psi[N, P] = \Phi[M, R] \times \Theta[R, P] = \sum_{i \in R} \Phi[M, i] \times \Theta[i, P] =$$

$$= \sum_{i \in P} \Phi[M, i] \times \Theta[i, P] = \Phi[M, P] \times \Theta[P, P].$$

Объединяя это с (8.14), приходим к окончательному представлению

$$Z_* = \Phi[M, P] \times \Theta[P, P] \times \Psi^\top[P, N] = \sum_{k=1}^p \vartheta_k \varphi_k \psi_k^\top. \tag{8.15}$$

Формула (8.15) показывает, как связана проекция Z_* с сингулярным разложением матрицы A.

Результат теоремы 8.1 можно записать в виде приближенной формулы

$$A = X_* Y_*^\top + H_*. \tag{8.16}$$

Здесь $X_* Y_*^\top$ — наилучшее приближение для матрицы A, а H_* — матрица погрешностей, у которой $\|H_*\| = \left(\sum_{k=p+1}^r \vartheta_k^2\right)^{1/2}$. Формула (8.16) практически полезна тогда, когда общее число элементов матриц X_* и Y_* значительно меньше числа элементов матрицы A, а норма матрицы погрешностей H_* достаточно мала. В этом случае приближение $X_* Y_*^\top$ обеспечивает *сжатие информации*, содержащейся в матрице A.

ЛИТЕРАТУРА К ГЛАВЕ 2

- Парлетт Б. Симметричная проблема собственных значений. Численные методы. М., 1983. 384 с.
- Лоусон Ч., Хенсон Р. Численное решение задач метода наименьших квадратов. М., 1986. 232 с.
- Гаврурин М.К., Малоземов В.Н. Экстремальные задачи с линейными ограничениями. Л., 1984. 176 с.
- Broyden C.G. Basic Matrices. An Introduction to Matrix Theory and Practice. Macmillan, London, 1975. 211 p.

РЕШЕНИЯ ЗАДАЧ

Глава 1

1.1. При $i \in M$ имеем

$$A[i,N] \times x[N] = \sum_{j \in N} A[i,j] \times \left(\sum_{k=1}^{p} \alpha_k x_k[j] \right) =$$

$$= \sum_{k=1}^{p} \alpha_k \sum_{j \in N} A[i,j] \times x_k[j] = \sum_{k=1}^{p} \alpha_k A[i,N] \times x_k[N].$$

1.2. Проверим, например, второе равенство:

$$C[i,M] \times A[M,j] = \sum_{l \in M} C[i,l] \times \left(\sum_{k=1}^{s} \beta_k A_k[l,j] \right) =$$

$$= \sum_{k=1}^{s} \beta_k \sum_{l \in M} C[i,l] \times A_k[l,j] = \sum_{k=1}^{s} \beta_k C[i,M] \times A_k[M,j].$$

1.3. Согласно определению транспонированной матрицы получаем

$$C^\top[j,i] = C[i,j] = A[i,N] \times B[N,j] = \sum_{k \in N} A[i,k] \times B[k,j] =$$

$$= \sum_{k \in N} B^\top[j,k] \times A^\top[k,i] = B^\top[j,N] \times A^\top[N,i].$$

1.4. Обозначим $N_2 = N \backslash N_1$. Тогда

$$E[N_1,N] \times x[N] = E[N_1,N_1] \times x[N_1] + E[N_1,N_2] \times x[N_2] =$$

$$= E[N_1,N_1] \times x[N_1] = x[N_1].$$

Аналогично проверяется второе равенство.

2.1. Доказательство легко следует из определения линейной зависимости векторов.

2.2. Согласно теореме 2.1, $p \leq n$. Вместе с тем, по условию орты $\{e_j\}$ можно представить в виде линейной комбинации векторов $x_1, x_2, \ldots, x_p$. Поскольку орты линейно независимы и их число равно n, то по основной лемме линейной алгебры $n \leq p$. Объединяя два неравенства, приходим к равенству $p = n$.

2.3. Для удобства считаем, что $N = 1\colon n$. К векторам $x_1, \ldots, x_p$ будем последовательно добавлять орты $e_1, e_2, \ldots, e_n$, пропуская лишь те, которые линейно выражаются через предыдущие векторы строящейся системы. В результате получим систему линейно независимых векторов $x_1, \ldots, x_p, x_{p+1}, \ldots, x_m$, таких, что все орты являются их линейными комбинациями. Но тогда любой вектор $x \in \mathcal{R}^N$ можно представить в виде линейной комбинации векторов $x_1, \ldots, x_m$. Учитывая предыдущую задачу, заключаем, что $m = n$.

3.1. Столбец с индексом i диагональной матрицы A равен $A[i,i]e_i$. Остается воспользоваться общим фактом: если векторы $\{x_i\}$ линейно независимы, то линейно независимы также векторы $\{\lambda_i x_i\}$ при ненулевых λ_i.

4.1. В равенстве $A^{-1}A = E$ перейдем к транспонированным матрицам. Получим $A^\top (A^{-1})^\top = E$. По определению обратной матрицы $(A^\top)^{-1} = (A^{-1})^\top$.

4.2. Поскольку $AB(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = E$, то по определению обратной матрицы $(AB)^{-1} = B^{-1}A^{-1}$.

4.3. Согласно теоремам 4.1 и 3.2

$$|N| = \operatorname{rank} C \leq \min\{\operatorname{rank} A, \operatorname{rank} B\}.$$

Отсюда следует, что $\operatorname{rank} A = \operatorname{rank} B = |N|$. Остается еще раз сослаться на теорему 4.1.

4.4. Допустим, что верхняя треугольная матрица $A = A[1 : n, 1 : n]$ обратима, однако $A[i,i] = 0$ при некотором $i \in 1 : n$. В этом случае подматрица $A[i : n, 1 : i]$ – нулевая. У подматрицы $A[1 : i-1, 1 : i]$ числа строк меньше числа столбцов, поэтому ее столбцы линейно зависимы. Но тогда и столбцы подматрицы $A[1 : n, 1 : i]$ линейно зависимы, что противоречит обратимости матрицы A. Доказано, что у обратимой верхней треугольной матрицы все диагональные элементы отличны от нуля.

Наоборот, пусть все диагональные элементы верхней треугольной матрицы A отличны от нуля. Покажем, что однородная система $Ax = \mathbf{0}$ имеет только нулевое решение. Возьмем любое решение x_0 системы $Ax = \mathbf{0}$. Для него

$$\sum_{j=1}^{n} A[i,j] \times x_0[j] = 0, \quad i \in 1 : n.$$

При $i = n$ имеем $A[n,n] \times x_0[n] = 0$. Отсюда следует, что $x_0[n] = 0$. В равенстве

$$\sum_{j=1}^{n-1} A[i,j] \times x_0[j] = 0$$

положим $i = n-1$. Получим $A[n-1, n-1] \times x_0[n-1] = 0$. Отсюда следует, что $x_0[n-1] = 0$. Продолжив этот процесс, придем к тому, что все компоненты вектора x_0 равны нулю.

Установлено, что однородная система $Ax = \mathbf{0}$ имеет только нулевое решение. Это гарантирует обратимость матрицы A.

4.5. Достаточно доказать, что столбцы матрицы $A[M_*, N_*]$ линейно независимы. Допустим противное. Тогда найдутся числа γ_j, $j \in N_*$, не все равные нулю, такие, что

$$\sum_{j \in N_*} \gamma_j A[i, j] = 0, \qquad i \in M_*. \tag{1}$$

Зафиксируем $i_0 \in M \backslash M_*$ и разложим небазисную строку $A[i_0, N]$ матрицы A по базисным:

$$A[i_0, N] = \sum_{i \in M_*} \beta_i A[i, N].$$

Получим

$$\sum_{j \in N_*} \gamma_j A[i_0, j] = \sum_{j \in N_*} \gamma_j \sum_{i \in M_*} \beta_i A[i, j] = \sum_{i \in M_*} \beta_i \sum_{j \in N_*} \gamma_j A[i, j] = 0.$$

Значит, равенство (1) выполняется при всех $i \in M$. Но это противоречит линейной независимости столбцов $A[M, j]$, $j \in N_*$, матрицы A.

4.6. Имеем

$$(A + B)(A^{-1} - \beta A^{-1} B A^{-1}) = E + BA^{-1} - \beta BA^{-1} - \beta BA^{-1} BA^{-1}. \tag{2}$$

Но $BA^{-1}BA^{-1} = xy^\top A^{-1} xy^\top A^{-1} = \langle y, A^{-1}x \rangle xy^\top A^{-1} = \alpha BA^{-1}$. Поэтому правая часть формулы (2) равна $E + (1 - \beta - \alpha\beta) BA^{-1} = E$. Остается воспользоваться определением обратной матрицы.

4.7. Нужно проверить справедливость четырех равенств

$$B(B^{-1} + B^{-1}CF^{-1}DB^{-1}) - CF^{-1}DB^{-1} = E,$$

$$-BB^{-1}CF^{-1} + CF^{-1} = \mathbf{0},$$

$$D(B^{-1} + B^{-1}CF^{-1}DB^{-1}) - HF^{-1}DB^{-1} = \mathbf{0},$$

$$-DB^{-1}CF^{-1} + HF^{-1} = E.$$

Первые два равенства тривиальны. Третье и четвертое следуют из того, что $(H - DB^{-1}C)F^{-1} = E$.

7.1. Пусть $\sum_{j=2}^{k}\alpha_j P_1 w_j = \mathbf{0}$. Обозначим $w = \sum_{j=2}^{k}\alpha_j w_j$. Тогда $P_1 w = \mathbf{0}$. Учитывая определение матрицы P_1, получаем $w - \langle w_1, w\rangle w_1 = \mathbf{0}$. Перепишем это равенство в виде $\sum_{j=1}^{k}\alpha_j w_j = \mathbf{0}$, где $\alpha_1 = -\langle w_1, w\rangle$. В силу линейной независимости векторов $w_1, w_2, \ldots, w_k$ все коэффициенты $\alpha_1, \alpha_2, \ldots, \alpha_k$ равны нулю. В частности, равны нулю и коэффициенты $\alpha_2, \ldots, \alpha_k$.

7.2. Воспользуемся алгоритмом вычисления ранга матрицы A, полагая в п. 4 $j_q = q$. В этом случае $f_q = F_{q-1}[M, q]$, $q = 1, 2, \ldots, n$. Проверим, что все векторы f_q ненулевые.

Докажем более сильное утверждение: у матрицы F_{q-1} первые $q-1$ столбцов нулевые, а остальные — линейно независимые. При $q = 2$ это следует из формулы $F_1 = P_1 A$ и предыдущей задачи. Аналогичные соображения используются и в индукционном переходе от $q-1$ к q.

Остается учесть, что $y_q^\top[j] = 0$ при $j \in 1\colon q-1$ (нулевые столбцы переходят в нулевые) и $y_q^\top[q] = \|f_q\|$.

7.3. Предположим, что $X_1 Y_1^\top = X_2 Y_2^\top$, где X_1, X_2 — матрицы с ортонормированными столбцами и $Y_1^\top, Y_2^\top$ — верхние треугольные матрицы с положительными диагональными элементами. Покажем, что $X_1 = X_2$ и $Y_1^\top = Y_2^\top$.

Обозначим $N = 1\colon n$ и запишем равенство первых столбцов матриц $X_1 Y_1^\top$ и $X_2 Y_2^\top$:

$$X_1[N,1] \times Y_1^\top[1,1] = X_2[N,1] \times Y_2^\top[1,1].$$

Переходя к равенству норм и учитывая, что $\|X_1[N,1]\| = \|X_2[N,1]\| = 1$, получаем $|Y_1^\top[1,1]| = |Y_2^\top[1,1]|$. Поскольку диагональные элементы матриц $Y_1^\top$ и $Y_2^\top$ положительны, то $Y_1^\top[1,1] = Y_2^\top[1,1]$ и, как следствие, $X_1[N,1] = X_2[N,1]$.

Допустим, что уже доказано равенство первых $k-1$ столбцов матриц X_1 и X_2, $Y_1^\top$ и $Y_2^\top$. Запишем равенство k-х столбцов матриц $X_1 Y_1^\top$ и $X_2 Y_2^\top$:

$$\sum_{j=1}^{k-1} X_1[N,j] \times Y_1^\top[j,k] + X_1[N,k] \times Y_1^\top[k,k] =$$

$$= \sum_{j=1}^{k-1} X_2[N,j] \times Y_2^\top[j,k] + X_2[N,k] \times Y_2^\top[k,k].$$

Умножим обе части этого равенства скалярно на $X_1[N,i] = X_2[N,i]$, $i \in 1\colon k-1$. В силу ортогональности получим $Y_1^\top[i,k] = Y_2^\top[i,k]$ при всех $i \in 1\colon k-1$. Но тогда $X_1[N,k] \times Y_1^\top[k,k] = X_2[N,k] \times$

$Y_2^\top[k,k]$. Отсюда, как и раньше, следует, что $Y_1^\top[k,k] = Y_2^\top[k,k]$ и $X_1[N,k] = X_2[N,k]$. Равенство k-х столбцов матриц X_1 и X_2, $Y_1^\top$ и $Y_2^\top$ установлено.

Глава 2

1.1. Если $AB = BA$, то

$$\langle ABx, y\rangle = \langle Bx, Ay\rangle = \langle x, BAy\rangle = \langle x, ABy\rangle.$$

Согласно лемме 1.1 матрица AB симметрична.

Наоборот, пусть матрица AB симметрична. Воспользуемся результатом задачи 1.3 главы 1. Получим

$$AB = (AB)^\top = B^\top A^\top = BA.$$

1.2. Симметричность матрицы $A^\top A$ следует, например, из соотношений

$$\langle A^\top Ax, y\rangle = \langle Ax, Ay\rangle = \langle x, A^\top Ay\rangle.$$

Далее

$$\langle A^\top Ax, x\rangle = \langle Ax, Ax\rangle = \|Ax\|^2 \geq 0.$$

Если $\langle A^\top Ax, x\rangle = 0$, то $Ax = \mathbf{0}$. В силу линейной независимости столбцов матрицы A равенство $Ax = \mathbf{0}$ возможно только при $x = \mathbf{0}$. Таким образом, $\langle A^\top Ax, x\rangle > 0$ при $x \neq \mathbf{0}$.

1.3. При $i \neq j$ имеем

$$0 < \langle D(e_i - e_j), e_i - e_j\rangle = D[i,i] - 2D[i,j] + D[j,j],$$

что равносильно требуемому.

1.4. Симметричность матрицы $D_0 = D[N_0, N_0]$ очевидна. Проверим ее положительную определенность. Возьмем ненулевой вектор $x_0 = x[N_0]$ и дополним его нулями до $x = x[N]$. Поскольку $x \neq \mathbf{0}$ и $D = D[N,N]$ — положительно определенная матрица, то

$$0 < \langle Dx, x\rangle = (D[N_0, N] \times x[N]) \times x[N_0] =$$

$$= (D[N_0, N_0] \times x[N_0]) \times x[N_0] = \langle D_0 x_0, x_0\rangle.$$

1.5. Доказательство проведем индукцией по n. При $n = 1$ утверждение очевидно, поскольку положительное число $D[1,1]$ можно представить в виде $D[1,1] = \sqrt{D[1,1]}\sqrt{D[1,1]}$. Сделаем индукционный переход от $n-1$ к n.

По условию матрица $D = D[1:n, 1:n]$ симметрична и положительно определена. Такими же свойствами обладает и матрица $D_1 = D[1:n-1, 1:n-1]$ (см. задачу 1.4). Имеем

$$D = \begin{bmatrix} D_1 & b \\ b^\top & \delta \end{bmatrix}.$$

По индукционному предположению $D_1 = L_1 L_1^\top$, где L_1 – нижняя треугольная матрица с положительными диагональными элементами. Отметим, что матрица L_1 обратима (задача 4.4 главы 1). Введем нижнюю треугольную матрицу L вида

$$L = \begin{bmatrix} L_1 & \mathbf{0} \\ p^\top & \lambda \end{bmatrix}.$$

Вектор p и число λ выберем из условия $LL^\top = D$. Поскольку

$$LL^\top = \begin{bmatrix} L_1 & \mathbf{0} \\ p^\top & \lambda \end{bmatrix} \begin{bmatrix} L_1^\top & p \\ \mathbf{0}^\top & \lambda \end{bmatrix} = \begin{bmatrix} D_1 & L_1 p \\ p^\top L_1^\top & p^\top p + \lambda^2 \end{bmatrix},$$

то достаточно обеспечить выполнение равенств

$$L_1 p = b, \qquad p^\top p + \lambda^2 = \delta.$$

Имеем $p = L_1^{-1} b$, $\lambda^2 = \delta - p^\top p$. Остается проверить, что $\delta - p^\top p > 0$. Воспользуемся формулой

$$p^\top p = b^\top (L_1^{-1})^\top L_1^{-1} b = b^\top (L_1 L_1^\top)^{-1} b = b^\top D_1^{-1} b.$$

Вектор $x_1 = [b^\top D_1^{-1}, -1]^\top$ отличен от нуля, поэтому $\langle Dx_1, x_1 \rangle > 0$. Учитывая, что

$$Dx_1 = \begin{bmatrix} D_1 & b \\ b^\top & \delta \end{bmatrix} \begin{bmatrix} D_1^{-1} b \\ -1 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ b^\top D_1^{-1} b - \delta \end{bmatrix},$$

получаем $0 < \langle Dx_1, x_1 \rangle = \delta - b^\top D_1^{-1} b = \delta - p^\top p$, т.е. $\delta - p^\top p > 0$.

2.1. По определению евклидовой нормы

$$Q(x) = 1/2\|Ax - b\|^2 = 1/2\langle Ax - b, Ax - b \rangle =$$
$$= 1/2\langle A^\top A x, x \rangle - \langle A^\top b, x \rangle + 1/2\|b\|^2.$$

Поскольку $A^\top A$ — симметричная неотрицательно определенная матрица, то квадратичная функция Q выпукла на $\mathcal{R}^N$.

2.2. В квадратичной функции $Q(x)$ сделаем подстановку $x = x_0 + By$. Согласно лемме 2.1 получим

$$Q(x_0 + By) = Q(x_0) + \langle Dx_0 + c, By \rangle + 1/2\langle DBy, By \rangle =$$
$$= 1/2\langle B^\top DBy, y \rangle + \langle B^\top (Dx_0 + c), y \rangle + Q(x_0).$$

2.3. Положим $D = {}^1\!/\!_2(G + G^\top)$. Очевидно, что D — симметричная матрица. Кроме того,

$$ {}^1\!/\!_2\langle Dx, x\rangle = {}^1\!/\!_4\langle (G + G^\top)x, x\rangle = {}^1\!/\!_2\langle Gx, x\rangle = Q(x). $$

3.1. Введем квадратичную функцию $Q(x) = {}^1\!/\!_2\langle Dx, x\rangle$. Она неотрицательна, выпукла и обращается в ноль при $x = x_0$. Поскольку x_0 является точкой минимума функции Q, то по теореме 3.2 $Dx_0 = \mathbf{0}$.

3.2. Согласно лемме 1.2 матрица D обратима, поэтому уравнение $Dy = x$ имеет единственное решение $y_* = D^{-1}x$. По теореме 3.2 функция $Q(y) = \langle Dy, y\rangle - 2\langle x, y\rangle$ достигает в точке y_* минимального значения. Найдем это минимальное значение. Имеем

$$ Q(y_*) = \langle Dy_*, y_*\rangle - 2\langle x, y_*\rangle = -\langle x, D^{-1}x\rangle. $$

3.3. По условию $Dx_* + c = \mathbf{0}$. Следовательно

$$ Q(x) - Q(x_*) = {}^1\!/\!_2\langle D(x - x_*), x - x_*\rangle = {}^1\!/\!_2\langle Dx + c, x - x_*\rangle. $$

4.1. Возьмем вектор $\hat{x}$ из $S(x_{p+1}, \ldots, x_n)$ и разложим его по обобщенным собственным векторам матрицы D:

$$ \hat{x} = \sum_{j=1}^{n} c_j x_j. $$

Умножим обе части этого равенства скалярно на Bx_i, $i \in p+1\colon n$. Учитывая, что $\langle Bx_i, \hat{x}\rangle = 0$ при $i \in p+1\colon n$, получаем $c_i = 0$. Значит,

$$ \hat{x} = \sum_{j=1}^{p} c_j x_j. $$

Из условия $\langle B\hat{x}, \hat{x}\rangle = 1$ следует, что $\sum_{j=1}^{p} c_j^2 = 1$.

Обратимся к функции $\langle Dx, x\rangle$ и оценим ее значение сверху при $x = \hat{x}$:

$$ \langle D\hat{x}, \hat{x}\rangle = \sum_{j=1}^{p} \lambda_j c_j^2 \le \lambda_p. $$

Вместе с тем, $\langle Dx_p, x_p\rangle = \lambda_p$. Получили, что наибольшее значение функции $\langle Dx, x\rangle$ на $S(x_{p+1}, \ldots, x_n)$ равно λ_p и достигается в точке x_p.

4.2. Зафиксируем $a_{p+1}, \ldots, a_n$. Так же, как при доказательстве теоремы Куранта – Фишера, построим вектор $x_0 = \sum_{j=p}^{n} c_j x_j$, принадлежащий $S(a_{p+1}, \ldots, a_n)$. В частности, будет выполняться равенство

$\sum_{j=p}^n c_j^2 = 1$. С помощью x_0 оценим значение целевой функции снизу:

$$\max_{x \in S(a_{p+1},\dots,a_n)} \langle Dx, x\rangle \geq \langle Dx_0, x_0\rangle = \sum_{j=p}^n \lambda_j c_j^2 \geq \lambda_p.$$

Вместе с тем, согласно предыдущей задаче

$$\max_{x \in S(x_{p+1},\dots,x_n)} \langle Dx, x\rangle = \lambda_p.$$

Получили, что наименьшее значение целевой функции равно λ_p и достигается при $a_{p+1} = x_{p+1}, \dots, a_n = x_n$.

5.1. Воспользуемся соотношениями (5.3), (5.4). Согласно теореме 4.4 главы 1, $\operatorname{rank} D = \operatorname{rank} \Lambda$. Остается учесть, что ранг диагональной матрицы равен числу ненулевых диагональных элементов (задача 3.1 главы 1).

5.2. Легко вычислить след одноранговой матрицы $x_k x_k^\top$:

$$\sum_{i=1}^n x_k[i] \times x_k^\top[i] = \sum_{i=1}^n (x_k[i])^2 = 1.$$

Согласно (5.6) след симметричной матрицы D равен $\sum_{k=1}^n \lambda_k$.

5.3. Учитывая (5.6) и (5.7), получаем

$$C := D + D^{-1} - 2E = \sum_{k=1}^n (\sqrt{\lambda_k} - \frac{1}{\sqrt{\lambda_k}})^2 x_k x_k^\top.$$

Отсюда следует, что

$$\langle Cx, x\rangle = \sum_{k=1}^n (\sqrt{\lambda_k} - \frac{1}{\sqrt{\lambda_k}})^2 \langle x_k, x\rangle^2 \geq 0.$$

По условию все собственные числа λ_k матрицы D положительны и отличны от единицы. Но тогда $(\sqrt{\lambda_k} - 1/\sqrt{\lambda_k})^2 > 0$ при всех $k \in 1\colon n$. Теперь так же, как в основном тексте, показывается, что $\langle Cx, x\rangle = 0$ только при $x = \mathbf{0}$.

5.4. Пусть $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ — собственные числа матрицы D. Напомним, что

$$\lambda_1 = \min_{\|x\|=1} \langle Dx, x\rangle, \qquad \lambda_n = \max_{\|x\|=1} \langle Dx, x\rangle.$$

Второе соотношение формально не доказывалось, но его легко получить, если воспользоваться разложением единичного вектора

x по базису, состоящему из собственных векторов матрицы D. По условию задачи $0 < m \leq \lambda_1 \leq \ldots \leq \lambda_n \leq M$. Но тогда $M^{-1} \leq \lambda_n^{-1} \leq \ldots \leq \lambda_1^{-1} \leq m^{-1}$. Остается учесть, что λ_n^{-1} — наименьшее собственное число, а λ_1^{-1} — наибольшее собственное число матрицы D^{-1}.

5.5. В силу свойств матриц $D^{1/2}$ и $D^{-1/2}$ имеем

$$1/2\|D^{1/2}x + D^{-1/2}c\|^2 = 1/2\langle D^{1/2}x + D^{-1/2}c, D^{1/2}x + D^{-1/2}c\rangle =$$

$$= 1/2\langle Dx, x\rangle + \langle c, x\rangle + 1/2\|D^{-1/2}c\|^2,$$

что равносильно требуемому.

6.1. В данном случае матрица A состоит из одной строки $w^\top$, поэтому $(AA^\top)^{-1} = 1$. Получаем

$$P = E - A^\top(AA^\top)^{-1}A = E - ww^\top.$$

6.2. Отметим, что $Vw = -w$. Это значит, что матрица V имеет собственное число $\lambda_1 = -1$, которому соответствует собственный вектор w. Остальные собственные числа $\lambda_2, \ldots, \lambda_n$ равны 1. Им соответствуют собственные векторы $x_2, \ldots, x_n$, образующие ортонормированный базис подпространства $\mathcal{L} = \{x \mid \langle w, x\rangle = 0\}$ (см. основной текст). Поскольку

$$\langle w, x_k\rangle = \langle w, Vx_k\rangle = \langle Vw, x_k\rangle = -\langle w, x_k\rangle,$$

то $\langle w, x_k\rangle = 0$ при $k \in 2{:}n$. Таким образом, полная система собственных векторов $w, x_2, \ldots, x_n$ является ортонормированной.

6.3. Матрица B имеет собственные числа

$$\lambda_1 = \cdots = \lambda_{n-1} = 1, \quad \lambda_n = 1 + \gamma.$$

Это доказывается так же, как в предыдущей задаче.

7.1. В силу свойств диагональной матрицы Θ и формулы (7.4) имеем

$$A = \Phi[M, R] \times \Theta[R, R] \times \Psi^\top[R, N].$$

Столбец с индексом k матрицы $\Phi[M, R] \times \Theta[R, R]$ равен $\vartheta_k \varphi_k$. Поэтому согласно формуле (1.8) главы 1

$$A = \sum_{k=1}^{r} \vartheta_k \varphi_k \psi_k^\top.$$

7.2. Для произведения $AA^\top\Phi$ получаем формулу

$$AA^\top\Phi = \Phi\Theta\Theta^\top.$$

Матрица $\Theta\Theta^\top$ квадратная и диагональная. На диагонали у нее стоят числа $\vartheta_1^2, \ldots, \vartheta_m^2$ при $m \le n$ и $\vartheta_1^2, \ldots, \vartheta_n^2, 0, \ldots, 0$ при $m > n$. Они и будут собственными числами матрицы $AA^\top$, соответствующими собственным векторам $\varphi_1, \ldots, \varphi_m$.

В свою очередь, формула $A^\top A\Psi = \Psi\Theta^\top\Theta$ показывает, что матрица $A^\top A$ имеет собственные числа $\vartheta_1^2, \ldots, \vartheta_m^2, 0, \ldots, 0$ при $m \le n$ и $\vartheta_1^2, \ldots, \vartheta_n^2$ при $m > n$ и им соответствуют собственные векторы $\psi_1, \ldots, \psi_n$.

7.3. Для квадратной матрицы A запишем сингулярное разложение $A = \Phi\Theta\Psi^\top$. Поскольку $\Psi^\top\Psi = E$, то

$$A = \Phi(\Psi^\top\Psi)\Theta\Psi^\top = (\Phi\Psi^\top)(\Psi\Theta\Psi^\top).$$

Положив $P = \Phi\Psi^\top$, $D = \Psi\Theta\Psi^\top$, получим требуемое разложение $A = PD$, в котором P — ортогональная, а D — симметричная неотрицательно определенная матрица.

7.4. Согласно предыдущей задаче

$$A^\top A = D^\top P^\top PD = D^\top D = D^2.$$

Таким образом, D есть квадратный корень из $A^\top A$. Нужно установить единственность квадратного корня — симметричной неотрицательно определенной матрицы D, такой, что $DD = A^\top A$. О существовании квадратного корня нам известно.

Возьмем матрицу D с указанными свойствами и обозначим ее собственные числа и собственные векторы через λ_1, …, λ_n и x_1, …, x_n соответственно. Покажем, что из равенства $A^\top A\psi = \vartheta^2\psi$ следует равенство $D\psi = \vartheta\psi$, где $\vartheta \ge 0$. Разложим вектор ψ по базису $x_1, \ldots, x_n$: $\psi = \sum_{k=1}^n \alpha_k x_k$. Имеем

$$A^\top A\psi = \vartheta^2\psi = \sum_{k=1}^n \vartheta^2\alpha_k x_k,$$

$$A^\top A\psi = D^2\psi = \sum_{k=1}^n \lambda_k^2\alpha_k x_k.$$

Значит, $\alpha_k(\vartheta^2 - \lambda_k^2) = 0$ при всех $k \in 1\colon n$. Если $\alpha_k \ne 0$, то $\vartheta^2 = \lambda_k^2$, или, в силу неотрицательности, $\vartheta = \lambda_k$. Учитывая это, получаем

$$D\psi = \sum_{\{k \mid \alpha_k \ne 0\}} \alpha_k\lambda_k x_k = \vartheta \sum_{\{k \mid \alpha_k \ne 0\}} \alpha_k x_k = \vartheta\sum_{k=1}^n \alpha_k x_k = \vartheta\psi.$$

Обозначим через $\vartheta_1^2, \dots, \vartheta_n^2$ и $\psi_1, \dots, \psi_n$ собственные числа и собственные векторы матрицы $A^\top A$. По доказанному

$$D\psi_k = \vartheta_k \psi_k, \qquad k \in 1:n.$$

По этой информации матрица D восстанавливается однозначно.