

你永远不知道一个强迫症能干出什么事情

倪兴程¹

2024 年 10 月 18 日

¹Email: 19975022383@163.com

Todo list

回头改证明，同时注意数域问题	3
线性方程组解链接	3
链接方程组秩与维数的公式	3
数域问题	11
链接线性方程组的关系	17
Schmidt 正交化链接	26
可逆矩阵行列式链接	32
可逆矩阵行列式链接	33
有空证明	35
行列式等于特征值的积，行列式大于 0 矩阵可逆	39
转置秩不变	39
对称幂等阵	41

目录

第一章 矩阵	1
1.1 矩阵空间	1
1.1.1 矩阵的运算	1
1.1.2 矩阵的行列式	3
1.2 矩阵的向量空间	3
1.3 线性方程组	3
1.4 矩阵的等价关系	5
1.4.1 相抵	5
1.4.2 相似	6
1.4.3 合同	6
1.5 相抵的应用	9
1.5.1 广义逆	9
1.5.2 Moore-Penrose 广义逆	12
1.5.3 线性方程组的解	16
1.6 相似的应用	18
1.6.1 特征值与特征向量	18
1.6.2 矩阵的对角化	22
1.6.3 Hermitian 矩阵的对角化	23
1.7 合同的应用——二次型	27
1.7.1 二次型的规范形	28
1.7.2 正定二次型与正定矩阵	30
1.8 特殊矩阵	35
1.8.1 幂等阵	35
1.9 矩阵的分解	36
1.9.1 SVD 分解	36
第二章 线性模型	38

Chapter 1

矩阵

1.1 矩阵空间

Definition 1.1. 由 $s \cdot m$ 个数排成 s 行、 m 列的一张表称为一个 $s \times m$ 矩阵 (*matrix*), 通常用大写英文字母表示, 其中的每一个数称为这个矩阵的一个元素, 第 i 行与第 j 列交叉位置的元素称为矩阵的 (i, j) 元, 记作 $A(i; j)$ 。一个 $s \times m$ 矩阵可以简单地记作 $A_{s \times m}$ 。如果矩阵 A 的 (i, j) 元是 a_{ij} , 那么可以记作 $A = (a_{ij})$ 。如果一个矩阵的行数和列数相同, 则称它为方阵, n 行 n 列的方阵也成为 n 阶矩阵。对于两个矩阵 A 和 B , 如果它们的行数都等于 s 且列数都等于 m , 同时还有 $A(i; j) = B(i; j), i = 1, 2, \dots, s, j = 1, 2, \dots, m$, 那么称 A 和 B 相等, 记作 $A = B$ 。

1.1.1 矩阵的运算

加减法与数量乘法

Definition 1.2. 将数域 K 上所有 $s \times m$ 矩阵组成的集合记作 $M_{s \times m}(K)$, 当 $s = m$ 时, $M_{s \times s}(K)$ 可以简记作 $M_s(K)$ 。在 $M_{s \times m}(K)$ 中定义如下运算:

1. 加法:

$$\forall A = (a_{ij}), B = (b_{ij}) \in M_{s \times m}(K), A + B = (a_{ij} + b_{ij})$$

2. 纯量乘法:

$$\forall k \in K, \forall A = (a_{ij}), kA = (ka_{ij})$$

那么 $M_{s \times m}(K)$ 构成一个线性空间。

Proof. 首先证明如上定义的加法和纯量乘法对 $M_{s \times m}(K)$ 是封闭的。由数域中加法和乘法的封闭性, $a_{ij} + b_{ij} \in K, ka_{ij} \in K, i = 1, 2, \dots, s, j = 1, 2, \dots, m$, 所以如上定义的加法与纯量乘法对 $M_{s \times m}(K)$ 是封闭的。

接下来证明如上定义的加法和纯量乘法满足线性空间中的 8 条运算法则:

1. 因为数域内的数满足加法交换律与加法结合律, 所以 $M_{s \times m}(K)$ 上的加法满足线性空间运算法则 (1)(2);
2. 取一个元素全为 0 的 $s \times m$ 矩阵, 将其记作 $\mathbf{0}$, 显然对 $\forall A \in M_{s \times m}(K)$, 有 $A + \mathbf{0} = A$, 因此 $M_{s \times m}(K)$ 中存在零元且它就是元素全为 0 的 $s \times m$ 矩阵, 称其为零矩阵 (zero matrix), 就记作 $\mathbf{0}$ 。因此, $M_{s \times m}(K)$ 上的加法满足线性空间运算法则 (3);
3. 对 $\forall A \in M_{s \times m}(K)$, 取 $-A = (-a_{ij})$, 则有 $A + (-A) = (a_{ij} - a_{ij}) = \mathbf{0}$ 。由 A 的任意性, $M_{s \times m}(K)$ 中的每个元素都具有负元, 将 $\forall A \in M_{s \times m}(K)$ 的负元就记作 $-A$ 。因此, $M_{s \times m}(K)$ 上的加法满足线性空间运算法则 (4);
4. 因为数域内的数满足乘法结合律和乘法分配律, 同时它们乘 1 的积是自身, 所以 $M_{s \times n}$ 上的纯量乘法满足线性空间运算法则 (5)(6)(7)(8)。

证明完毕。 □

Definition 1.3. 定义 $M_{s \times m}(K)$ 上矩阵的减法如下: 设 $A, B \in M_{s \times m}(K)$, 则:

$$A - B \stackrel{\text{def}}{=} A + (-B)$$

乘法

Definition 1.4. 设 $A = (a_{ij})_{s \times n}$, $B = (b_{ij})_{n \times m}$, 令 $C = (c_{ij})_{s \times m}$, 其中:

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}, \quad i = 1, 2, \dots, s, \quad j = 1, 2, \dots, m$$

则矩阵 C 称作矩阵 A 与 B 的乘积, 记作 $C = AB$ 。

初等变换

Definition 1.5. 称以下变换为矩阵的初等行变换 (elementary row operation):

1. 把一行的倍数加到另一行上;
2. 互换两行的位置;
3. 用一个非零数乘某一行。

称以下变换为矩阵的初等列变换 (elementary column operation):

1. 把一列的倍数加到另一列上;
2. 互换两列的位置;
3. 用一个非零数乘某一列。

1.1.2 矩阵的行列式

1.2 矩阵的向量空间

Definition 1.6. 设 $A = (\alpha_1, \alpha_2, \dots, \alpha_n) \in M_{m \times n}(K)$, 将:

$$\left\{ \sum_{i=1}^n k_i \alpha_i : k_i \in K \right\} \stackrel{\text{def}}{=} \mathcal{M}(A)$$

Theorem 1.1. 设 $A \in M_{m \times n}(K)$, 则:

$$\mathcal{M}(A) = \mathcal{M}(AA^T)$$

Proof. 由定义, 显然 $\mathcal{M}(AA^T) \subset \mathcal{M}(A)$ 。对于任意的 $x \perp \mathcal{M}(AA^T)$, 有 $x^T AA^T = \mathbf{0}$, 于是 $\|A^T x\|^2 = x^T AA^T x = 0$, 即 $A^T x = \mathbf{0}$, 于是 $x \perp \mathcal{M}(A)$ 。 □

回头改证明, 同时注意数域问题

Theorem 1.2. 设 $A \in M_{m \times n}(\mathbb{C})$, 则有:

$$\text{rank}(AA^H) = \text{rank}(A^H A) = \text{rank}(A)$$

Proof. 只需证明方程 $A^H Ax = \mathbf{0}$ 与 $Ax = \mathbf{0}$ 同解。注意到 $Ax = \mathbf{0}$ 则必然有 $A^H Ax = \mathbf{0}$, 而若 $A^H Ax = \mathbf{0}$, 则必有 $x^H A^H Ax = \|Ax\|^2 = 0$, 所以 $Ax = \mathbf{0}$ 。 □

线性方程组解链接

$$n - \text{rank}(A^H A) = n - \text{rank}(A)$$

链接方程组秩与维数的公式

所以:

$$\text{rank}(A^H A) = \text{rank}(A)$$

同理可得:

$$\text{rank}(AA^H) = \text{rank}(A^H) = \text{rank}(A)$$

于是有:

$$\text{rank}(AA^H) = \text{rank}(A^H A) = \text{rank}(A)$$

□

1.3 线性方程组

Definition 1.7. 设 x_1, x_2, \dots, x_n 为 n 个未知数, 若一个方程具有如下形式:

$$a_1 x_1 + a_2 x_2 + \dots + a_n x_n = b$$

其中, a_1, a_2, \dots, a_n 为系数 (coefficient), b 为常数项 (constant term), 则称该方程为线性方程 (linear equation)。由 m 个形如上式的方程组成的方程组:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{cases}$$

被称为 n 元线性方程组 (*system of linear equations, SLE*)。由矩阵乘法的定义, 该方程组也可以写作矩阵形式:

$$Ax = b$$

其中,

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Definition 1.8. 给定线性方程组 $Ax = b$, 称如下矩阵

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}.$$

为该线性方程组的增广矩阵 (*augmented matrix*), 记为 $[A|b]$ 。

Definition 1.9. 一个矩阵被称为行阶梯形矩阵 (*row echelon form, REF*), 如果它满足以下条件:

1. 所有零行 (全为零的行) 位于非零行的下方;
2. 若某一行非零, 则该行的首个非零元素 (称为主元 (*pivot*)) 位于该行之前所有行的主元右侧。

一个矩阵被称为简化行阶梯形矩阵 (*reduced row echelon form, RREF*), 如果满足以下条件:

1. 它是阶梯形矩阵;
2. 每个非零行的主元都是 1;
3. 每个主元所在列的其他元素均为 0。

Theorem 1.3. 任意一个矩阵都可以经过一系列初等行变换化成行阶梯形矩阵, 进而可以经过一系列初等行变换化成简化行阶梯形矩阵。

Definition 1.10. 设增广矩阵化简后变为阶梯形矩阵, 称每一行主元所在列所对应的未知数为主变量 (*pivot variable*), 同时称非主元所在列对应的未知数为自由未知量 (*free variable*)。

Theorem 1.4. 数域 K 上线性方程组 $Ax = b$ 有解的充分必要条件为 $\text{rank}(A) = \text{rank}([A|b])$ 。

Proof.

□

1.4 矩阵的等价关系

1.4.1 相抵

Definition 1.11. $A, B \in M_{s \times m}(K)$, 如果满足下述条件中的任意一个:

1. A 能够通过初等行变换和初等列变换变成 B ;
2. 存在数域 K 上的 s 阶初等矩阵 P_1, P_2, \dots, P_t 与 m 阶初等矩阵 Q_1, Q_2, \dots, Q_n 使得:

$$P_t \cdots P_2 P_1 A Q_1 Q_2 \cdots Q_n = B$$

3. 存在数域 K 上的 s 阶可逆矩阵 P 与 m 阶可逆矩阵 Q 使得:

$$PAQ = B$$

则称 A 与 B 相抵 (equivalent)。

上述三个条件显然是等价的。

Theorem 1.5. 相抵是 $M_{s \times m}(K)$ 上的一个等价关系。在相抵关系下, 矩阵 A 的等价类称为 A 的相抵类。

Proof. 证明是显然的。 □

Theorem 1.6. 设 $A \in M_{s \times m}(K)$, 且 $\text{rank}(A) = r$ 。如果 $r > 0$, 那么 A 相抵于如下形式的矩阵:

$$\begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

称该矩阵为 A 的相抵标准形。如果 $r = 0$, 则 A 相抵于零矩阵, 此时称零矩阵为 A 的相抵标准形。

Proof. 一个矩阵通过初等行变换一定可以变成一个简化行阶梯型矩阵, 再由初等列变换即可得到上述矩阵。 □

Theorem 1.7 (相抵的完全不变量). $A, B \in M_{s \times m}(K)$, A 与 B 相抵当且仅当它们的秩相同。

Proof. (1) 必要性: 初等行变换和初等列变换不改变矩阵的秩。

(2) 充分性: 若 A, B 的秩相同, 则它们的相抵标准形相同。因为相抵是一个等价关系, 由等价关系的对称性与传递性即可得到 A 与 B 相抵。 □

1.4.2 相似

Definition 1.12. $A, B \in M_n(K)$ 。如果存在可逆矩阵 $P \in M_n(K)$, 使得:

$$P^{-1}AP = B$$

则称 A 与 B 相似 (similar)。

Theorem 1.8. 相似是 $M_n(K)$ 上的一个等价关系。在相似关系下, 矩阵 A 的等价类称为 A 的相似类。

Proof. 证明是显然的。 □

Property 1.4.1 (相似的不变量). 相似的矩阵具有相同的行列式值、秩、迹、特征多项式、特征值 (包括重数相同)。

Proof. 设 $A, B \in M_n(K)$ 且 A 与 B 相似, 于是存在可逆矩阵 $P \in M_n(K)$ 使得 $P^{-1}AP = B$ 。

$$(1) |A| = |P^{-1}AP| = |P^{-1}| |B| |P| = |P^{-1}| |P| |B| = |B|。$$

(2) 初等行变换与初等列变换不改变矩阵的秩。

$$(3) \text{ 由 (2) 可得 } \operatorname{tr}(A) = \operatorname{tr}(P^{-1}BP) = \operatorname{tr}(BPP^{-1}) = \operatorname{tr}(B)。$$

(4)(5) 参考定理 1.23。 □

1.4.3 合同

Definition 1.13. $A, B \in M_n(K)$ 。如果存在可逆矩阵 $C \in M_n(K)$, 使得:

$$C^TAC = B$$

则称 A 与 B 合同 (congruent), 记作 $A \cong B$ 。如果对称矩阵 A 合同于一个对角矩阵, 那么称这个对角矩阵为 A 的一个合同标准形。

Theorem 1.9. 合同是 $M_n(K)$ 上的一个等价关系。在合同关系下, 矩阵 A 的等价类称为 A 的合同类。

Proof. 证明是显然的。 □

Definition 1.14. 对 n 阶矩阵的行作初等行变换, 再对该矩阵的同样标号的列作相同的初等列变换, 这种变换被称为成对初等行、列变换。

Lemma 1.1. $A, B \in M_n(K)$, 则 A 合同于 B 当且仅当 A 经过一系列成对初等行、列变换可以变成 B , 此时对 I 作其中的初等列变换即可得到可逆矩阵 C , 使得 $C^TAC = B$ 。

Proof. 由可逆矩阵的初等矩阵分解, 可得:

$$A \cong B \Leftrightarrow \text{存在数域 } K \text{ 上的可逆矩阵 } C, \text{ 使得 } C^TAC = B$$

$$\Leftrightarrow \text{存在数域 } K \text{ 上的初等矩阵 } P_1, P_2, \dots, P_t \text{ 使得}$$

$$C = P_1P_2 \cdots P_t$$

$$P_t^T \cdots P_2^T P_1^T A P_1 P_2 \cdots P_t = B$$

□

Theorem 1.10. 数域 K 上的任一对称矩阵都合同于一个对角矩阵。

Proof. 对数域 K 上对称矩阵的阶数 n 作数学归纳法。

当 $n = 1$ 时, 因为矩阵合同于自身, 同时一阶矩阵都是对角矩阵, 所以结论成立。

假设 $n - 1$ 阶对称矩阵都合同于对角矩阵, 考虑 n 阶矩阵 $A = (a_{ij})$ 。

情形一: $a_{11} \neq 0$

把 A 写成分块矩阵的形式, 然后对 A 作初等行变换与初等列变换可得:

$$\begin{pmatrix} a_{11} & A_1 \\ A_1^T & A_2 \end{pmatrix} \rightarrow \begin{pmatrix} a_{11} & A_1 \\ \mathbf{0} & A_2 - a_{11}^{-1} A_1^T A_1 \end{pmatrix} \rightarrow \begin{pmatrix} a_{11} & \mathbf{0} \\ \mathbf{0} & A_2 - a_{11}^{-1} A_1^T A_1 \end{pmatrix}$$

于是有:

$$\begin{pmatrix} 1 & \mathbf{0} \\ -a_{11}^{-1} A_1^T & I_{n-1} \end{pmatrix} \begin{pmatrix} a_{11} & A_1 \\ A_1^T & A_2 \end{pmatrix} \begin{pmatrix} 1 & -a_{11}^{-1} A_1 \\ \mathbf{0} & I_{n-1} \end{pmatrix} = \begin{pmatrix} a_{11} & \mathbf{0} \\ \mathbf{0} & A_2 - a_{11}^{-1} A_1^T A_1 \end{pmatrix}$$

因为 A 是一个对称矩阵, 所以 A_2 是一个对称矩阵, 于是:

$$(A_2 - a_{11}^{-1} A_1^T A_1)^T = A_2^T - a_{11}^{-1} A_1^T (A_1^T)^T = A_2 - a_{11}^{-1} A_1^T A_1$$

所以 $A_2 - a_{11}^{-1} A_1^T A_1$ 是 $n - 1$ 阶对称矩阵。由归纳假设可知存在可逆矩阵 $C \in M_{n-1}(K)$ 使得 $C^T (A_2 - a_{11}^{-1} A_1^T A_1) C = D$, 其中 D 是一个对角矩阵, 即:

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & C^T \end{pmatrix} \begin{pmatrix} a_{11} & \mathbf{0} \\ \mathbf{0} & A_2 - a_{11}^{-1} A_1^T A_1 \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & C \end{pmatrix} = \begin{pmatrix} a_{11} & \mathbf{0} \\ \mathbf{0} & D \end{pmatrix}$$

于是有:

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & C^T \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ -a_{11}^{-1} A_1 & I_{n-1} \end{pmatrix} \begin{pmatrix} a_{11} & A_1 \\ A_1^T & A_2 \end{pmatrix} \begin{pmatrix} 1 & -a_{11}^{-1} A_1 \\ \mathbf{0} & I_{n-1} \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & C \end{pmatrix} = \begin{pmatrix} a_{11} & \mathbf{0} \\ \mathbf{0} & D \end{pmatrix}$$

因为:

$$\left[\begin{pmatrix} 1 & -a_{11}^{-1} A_1 \\ \mathbf{0} & I_{n-1} \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & C \end{pmatrix} \right]^T = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & C \end{pmatrix}^T \begin{pmatrix} 1 & -a_{11}^{-1} A_1 \\ \mathbf{0} & I_{n-1} \end{pmatrix}^T = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & C^T \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ -a_{11}^{-1} A_1 & I_{n-1} \end{pmatrix}$$

并且:

$$\begin{pmatrix} 1 & -a_{11}^{-1} A_1 \\ \mathbf{0} & I_{n-1} \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & C \end{pmatrix}$$

是一个可逆矩阵, 所以 A 合同于对角矩阵:

$$\begin{pmatrix} a_{11} & \mathbf{0} \\ \mathbf{0} & D \end{pmatrix}$$

情形二: $a_{11} = 0$, 存在 $i \neq 1$ 使得 $a_{ii} \neq 0$

把 A 的第 $1, i$ 行呼唤, 再把所得矩阵的第 $1, i$ 列呼唤, 得到的矩阵 B 的 $(1, 1)$ 元即为 $a_{ii} \neq 0$ 。根据情形一的讨论, B 合同于一个对角矩阵。因为 B 是由 A 作成对初等行、列变换得到的, 由引理 1.1 可得 $A \cong B$ 。由合同的传递性, A 也合同于一个对角矩阵。

情形三: $a_{ii} = 0, \forall i = 1, 2, \dots, n$, 存在 $a_{ij} \neq 0, i \neq j$

把 A 的第 j 行加到第 i 行上, 再把所得矩阵的第 j 列加到第 i 列上, 得到的矩阵 E 的 (i, i) 元即为 $2a_{ij} \neq 0$ 。由情形二的讨论, E 合同于一个对角矩阵。因为 E 是由 A 作成对初等行、列变换得到的, 由引理 1.1 可得 $A \cong E$ 。由合同的传递性, A 也合同于一个对角矩阵。

情形四: $A = \mathbf{0}$

因为 $\mathbf{0}$ 是一个对角矩阵, 所以结论显然成立。 \square

Theorem 1.11. 设对角矩阵 B 是对称矩阵 A 的合同标准形, 则 B 对角线上不为 0 的元素的个数等于 A 的秩。

Proof. 因为 $A \cong B$, 所以存在可逆矩阵 C 使得 $C^T A C = B$, 于是 $\text{rank}(A) = \text{rank}(B)$ 。 \square

实对称矩阵的合同规范形

Theorem 1.12. 对于任意的对称矩阵 $A \in M_n(\mathbb{R})$, A 都合同于对角矩阵

$\text{diag}\{1, 1, \dots, 1, -1, -1, \dots, -1, 0, 0, \dots, 0\}$, 系数为 1 的平方项个数称为 A 的正惯性指数 (positive inertia index), 系数为 -1 的平方项个数称为 A 的负惯性指数 (negative inertia index), 这个对角矩阵称为 A 的合同规范形。

Proof. 任取矩阵 $A \in M_n(\mathbb{R})$, 由定理 1.10 可得 A 合同一个对角矩阵 B 。对 B 作成对初等行、列变换可将 B 对角线上的元素重新排列, 使得正值在前, 负值在中间, 零值在最后, 如此得到对角矩阵 C , C 可写作:

$$C = \begin{pmatrix} c_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_p & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & -c_{p+1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & -c_{p+2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & -c_r & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}$$

其中 $c_1, c_2, \dots, c_r > 0$ 。再对 C 作成对初等行、列变换, 即先对第 i 行除 $\sqrt{c_i}$, 再对第 i 列

除 $\sqrt{c_i}$, $i = 1, 2, \dots, n$, 即可得到对角矩阵 D :

$$D = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & -1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & -1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}$$

由引理 1.1 可得, $D \cong C$, $C \cong B$, 又因为 $A \cong B$, 由合同的传递性与对称性即可得 $A \cong D$ 。由 A 的任意性结论得证。 \square

复对称矩阵的合同规范形

Theorem 1.13. 对于任意的 $A \in M_n(\mathbb{C})$, A 都合同于对角矩阵 $\text{diag}\{1, 1, \dots, 1, 0, 0, \dots, 0\}$, 这个对角矩阵称为 A 的合同规范形。

Proof. 任取矩阵 $A \in M_n(\mathbb{C})$, 由定理 1.10 可得 $A \cong B = \text{diag}\{b_1, b_2, \dots, b_r, 0, 0, \dots, 0\}$, 其中 r 是矩阵 B 的秩, $b_1, b_2, \dots, b_r \neq 0$ 。设 $b_j = r_j \cos \theta_j + i r_j \sin \theta_j$, $\theta_j \in [0, 2\pi)$, $j = 1, 2, \dots, r$ 。因为:

$$\left[\sqrt{r_j} \left(\cos \frac{\theta_j}{2} + i \sin \frac{\theta_j}{2} \right) \right]^2 = b_j$$

将 $\sqrt{r_j} \left(\cos \frac{\theta_j}{2} + i \sin \frac{\theta_j}{2} \right)$ 记作 $\sqrt{b_j}$, 作成对初等行、列变换, 即先对第 j 行除 $\sqrt{b_j}$, 再对第 j 列除 $\sqrt{b_j}$, 则可得到矩阵 $C = \text{diag}\{1, 1, \dots, 1, 0, 0, \dots, 0\}$, 其中 1 的个数为 r 。由引理 1.1 可得, $B \cong C$ 。因为 $A \cong B$, 由合同的传递性, $A \cong C$ 。由 A 的任意性, 结论成立。 \square

1.5 相抵的应用

1.5.1 广义逆

Definition 1.15. 设 $A \in M_{m \times n}(K)$, 一切满足方程组:

$$AXA = A$$

的矩阵 X 都被称为是 A 的广义逆 (*generalized inverse*), 记为 A^- 。

Theorem 1.14. 设非零矩阵 $A \in M_{m \times n}(K)$, $\text{rank}(A) = r$ 且:

$$A = P \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q$$

其中 P, Q 分别为数域 K 上的 m 阶可逆矩阵和 n 阶可逆矩阵, 则矩阵方程:

$$AXA = A$$

一定有解, 且其通解可表示为:

$$X = Q^{-1} \begin{pmatrix} I_r & B \\ C & D \end{pmatrix} P^{-1}$$

其中 B, C, D 分别为数域 K 上任意的 $r \times (m-r)$, $(n-r) \times r$, $(n-r) \times (m-r)$ 矩阵。

Proof. 若 X 是上述矩阵方程的一个解, 则:

$$\begin{aligned} P \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} QXP \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q &= P \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q \\ \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} QXP \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} &= \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \end{aligned}$$

将 QXP 写作如下分块矩阵的形式:

$$QXP = \begin{pmatrix} H & B \\ C & D \end{pmatrix}$$

其中 H, B, C, D 分别为数域 K 上任意的 $r \times r$, $r \times (m-r)$, $(n-r) \times r$, $(n-r) \times (m-r)$ 矩阵。于是:

$$\begin{aligned} \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} H & B \\ C & D \end{pmatrix} \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} &= \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \\ \begin{pmatrix} H & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} &= \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \end{aligned}$$

所以 $H = I_r$, 因此:

$$X = Q^{-1} \begin{pmatrix} I_r & B \\ C & D \end{pmatrix} P^{-1}$$

□

Property 1.5.1. 设 $A \in M_{m \times n}(K)$, $B \in M_{m \times q}(K)$, $C \in M_{p \times n}(K)$, 则广义逆 A^- 具有如下性质:

1. A^- 唯一的充分必要条件为 A 可逆, 此时 $A^- = A^{-1}$;
2. $\text{rank}(A^-) \geq \text{rank}(A) = \text{rank}(AA^-) = \text{rank}(A^-A)$;

数域问题

3. 若 $\mathcal{M}(B) \subset \mathcal{M}(A), \mathcal{M}(C) \subset \mathcal{M}(A^T)$, 则 $C^T A^- B$ 与 A^- 的选择无关;

4. $A(A^T A)^- A^T$ 与 $(A^T A)^-$ 的选择无关;

5. $A(A^T A)^- A^T A = A, A^T A(A^T A)^- A^T = A^T$;

6. 若 A 对称, 则 $[(A)^-]^T = (A)^-$ 。

Proof. (1) 充分性: 若 A 可逆, 则 $r = n$, 由 A^- 的通解公式, 显然此时 A^- 唯一。

必要性: 若 A^- 唯一, 则 $r = n$, 显然此时 A 可逆。

(2) 由 A^- 的通解公式, $\text{rank}(A^-) \geq r = \text{rank}(A)$ 。因为:

$$\begin{aligned} AA^- &= P \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q Q^{-1} \begin{pmatrix} I_r & B \\ C & D \end{pmatrix} P^{-1} = P \begin{pmatrix} I_r & B \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^{-1} \\ A^- A &= Q^{-1} \begin{pmatrix} I_r & B \\ C & D \end{pmatrix} P^{-1} P \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q = Q^{-1} \begin{pmatrix} I_r & \mathbf{0} \\ C & \mathbf{0} \end{pmatrix} Q \end{aligned}$$

显然, $\text{rank}(AA^-) = \text{rank}(A^- A) = \text{rank}(A) = r$ 。

(3) 由已知条件, 存在矩阵 D_1, D_2 使得 $B = AD_1, C = A^T D_2$, 于是:

$$C^T A^- B = D_2^T AA^- AD_1 = D_2^T AD_1$$

(4) 由定理 1.1 可知 $\mathcal{M}(A^T) = \mathcal{M}(A^T A)$, 于是存在矩阵 B 使得 $A^T = A^T A B$, 所以有:

$$A(A^T A)^- A^T = B^T A^T A(A^T A)^- A^T A B = B^T A^T A B$$

与 $(A^T A)^-$ 无关。

(5) 设 $B = A(A^T A)^- A^T A - A$, 则:

$$\begin{aligned} B^T B &= \{A^T A[(A^T A)^-]^T A^T - A^T\} [A(A^T A)^- A^T A - A] \\ &= A^T A[(A^T A)^-]^T A^T A(A^T A)^- A^T A - A^T A[(A^T A)^-]^T A^T A \\ &\quad - A^T A(A^T A)^- A^T A + A^T A \\ &= A^T A[(A^T A)^-]^T A^T A - A^T A[(A^T A)^-]^T A^T A - A^T A + A^T A = \mathbf{0} \end{aligned}$$

所以 $B = \mathbf{0}$ (考虑 $B^T B$ 主对角线上的元素), 于是 $A(A^T A)^- A^T A = A$ 。

设 $C = A^T A(A^T A)^- A^T - A^T$, 则:

$$\begin{aligned} CC^T &= [A^T A(A^T A)^- A^T - A^T] \{A[(A^T A)^-]^T A^T A - A\} \\ &= A^T A(A^T A)^- A^T A[(A^T A)^-]^T A^T A - A^T A(A^T A)^- A^T A \\ &\quad - A^T A[(A^T A)^-]^T A^T A + A^T A \\ &= A^T A[(A^T A)^-]^T A^T A - A^T A - A^T A[(A^T A)^-]^T A^T A + A^T A = \mathbf{0} \end{aligned}$$

所以 $C = \mathbf{0}$, 于是 $A^T A(A^T A)^- A^T = A^T$ 。

(6) 此时有:

$$AXA = A \Leftrightarrow A^T X^T A^T = A^T \Leftrightarrow AX^T A = A$$

□

1.5.2 Moore-Penrose 广义逆

Definition 1.16. 设 $A \in M_{m \times n}(\mathbb{C})$ 。若 $X \in M_{n \times m}(\mathbb{C})$ 满足：

$$\begin{cases} AXA = A \\ XAX = X \\ (AX)^H = AX \\ (XA)^H = XA \end{cases}$$

则称 X 为 A 的 *Moore-Penrose* 广义逆，记作 A^+ ，上述方程组被称为 A 的 *Penrose* 方程组。

满秩分解导出的广义逆

Theorem 1.15. 设 $A \in M_{m \times n}(\mathbb{C})$ ，则 A 的 *Penrose* 方程组一定有唯一解。对 A 进行满秩分解，设 $A = BC$ ，其中 B, C 分别为列满秩矩阵与行满秩矩阵，则 A 的 *Penrose* 方程组的解可表示为：

$$X = C^H(CC^H)^{-1}(B^HB)^{-1}B^H$$

Proof. 由定理 1.2 可知 $(B^HB)^{-1}, (CC^H)^{-1}$ 存在，将上述 X 代入 A 的 *Penrose* 方程组可得：

$$\begin{aligned} XAX &= C^H(CC^H)^{-1}(B^HB)^{-1}B^H B C C^H(CC^H)^{-1}(B^HB)^{-1}B^H \\ &= C^H(CC^H)^{-1}(B^HB)^{-1}B^H = X \\ AXA &= B C C^H(CC^H)^{-1}(B^HB)^{-1}B^H B C = B C = A \\ (AX)^H &= X^H A^H = B[(B^HB)^{-1}]^H[(CC^H)^{-1}]^H C C^H B^H \\ &= B[(B^HB)^{-1}]^H[(CC^H)^{-1}]^H C C^H B^H \\ &= B[(B^HB)^H]^{-1}[(CC^H)^H]^{-1} C C^H B^H \\ &= B(B^HB)^{-1}(CC^H)^{-1} C C^H B^H \\ &= B(B^HB)^{-1}B^H \\ &= B(CC^H)(CC^H)^{-1}(B^HB)^{-1}B^H = AX \\ (XA)^H &= A^H X^H = C^H B^H B[(B^HB)^{-1}]^H[(CC^H)^{-1}]^H C \\ &= C^H(CC^H)^{-1}C = C^H(CC^H)^{-1}(B^HB)^{-1}(B^HB)C = XA \end{aligned}$$

于是 X 与 A 的 *Penrose* 方程组相容，所以 X 是解。 □

奇异值分解导出的广义逆

Theorem 1.16. 设 $A \in M_{m \times n}(\mathbb{C})$ ，则有：

$$A^+ = Q \begin{pmatrix} A^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H$$

其中 P, Q, A 为 A 的奇异值分解中相关矩阵。

Proof. 将之代入到 A 的 Penrose 方程组中可得:

$$\begin{aligned}
 AQ \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H A &= P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H \\
 &= P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H = A \\
 Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H A Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H &= Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H \\
 &= Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H \\
 A Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H &= Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H A = I
 \end{aligned}$$

因为 I 是 Hermitian 矩阵, 于是 $Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H$ 与 A 的 Penrose 方程组相容, 所以它是解。 \square

Moore-Penrose 广义逆的性质

Property 1.5.2. 设 $A \in M_{m \times n}(\mathbb{C})$, 则 A 的 Moore-Penrose 广义逆 A^+ 具有如下性质:

1. A^+ 是唯一的;
2. $(A^+)^+ = A$;
3. $(A^+)^H = (A^H)^+$;
4. $\text{rank}(A^+) = \text{rank}(A)$;
5. 若 A 是一个 Hermitian 矩阵, 则:

$$A^+ = P \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H$$

其中 Λ 为 A 的非零特征值构成的对角矩阵, P 是一个正交矩阵;

6. 若 α 是一个非零向量, 则 $\alpha^+ = \frac{\alpha^H}{\|\alpha\|^2}$;
7. $I - A^+A \geq \mathbf{0}$;
8. $(A^H A)^+ = A^+(A^H)^+$;
9. $A^+ = (A^H A)^+ A^H = A^H (A A^H)^+$ 。

Proof. (1) 设 X_1, X_2 都是 A 的 Penrose 方程组的解, 则:

$$\begin{aligned}
 X_1 &= X_1 A X_1 = X_1 (A X_2 A) X_1 = X_1 (A X_2) (A X_1) \\
 &= X_1 (A X_2)^H (A X_1)^H = X_1 (A X_1 A X_2)^H = X_1 X_2^H (A X_1 A)^H \\
 &= X_1 X_2^H A^H = X_1 (A X_2)^H = X_1 A X_2 = X_1 (A X_2 A) X_2 \\
 &= (X_1 A) (X_2 A) X_2 = (X_1 A)^H (X_2 A)^H X_2 = (X_2 A X_1 A)^H X_2 \\
 &= (X_2 A)^H X_2 = X_2 A X_2 = X_2
 \end{aligned}$$

所以 Penrose 方程组的解是唯一的。

(2) 由 Penrose 方程的对称性可直接得到。

(3) 由 A^+ 的奇异值分解表示 (定理 1.16) 可得:

$$\begin{aligned}
 (A^+)^H &= \left[Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H \right]^H = P \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}^H Q^H \\
 &= P \begin{pmatrix} (\Lambda^{-1})^H & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H = P \begin{pmatrix} (\Lambda^H)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H
 \end{aligned}$$

将其代入 A^H 的 Penrose 方程组可得:

$$\begin{aligned}
 A^H (A^+)^H A^H &= Q \begin{pmatrix} \Lambda^H & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H P \begin{pmatrix} (\Lambda^H)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H Q \begin{pmatrix} \Lambda^H & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H \\
 &= Q \begin{pmatrix} \Lambda^H & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H = A^H \\
 (A^+)^H A^H (A^+)^H &= P \begin{pmatrix} (\Lambda^H)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H \begin{pmatrix} \Lambda^H & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H P \begin{pmatrix} (\Lambda^H)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H \\
 &= P \begin{pmatrix} (\Lambda^H)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H = (A^+)^H \\
 [A^H (A^+)^H]^H &= [(A^+)^H A^H]^H = A^+ A = I
 \end{aligned}$$

因为 I 是 Hermitian 矩阵, 于是 $(A^+)^H$ 与 A^H 的 Penrose 方程组相容, 所以 $(A^+)^H = (A^H)^+$ 。

(4) 由 A^+ 的奇异值分解表示 (定理 1.16) 显然可得 $\text{rank}(A^+) = \text{rank}(A)$, 而 $\text{rank}(A) = \text{rank}(A)$, 所以有 $\text{rank}(A^+) = \text{rank}(A)$ 。

(5) 因为 A 是一个 Hermitian 矩阵, 由性质 1.6.2(3) 可知存在正交矩阵 P 使得:

$$A = P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H$$

将 $P \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H$ 代入 A 的 Penrose 方程组可得:

$$\begin{aligned} AP \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H A &= P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H P \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H \\ &= P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H = A \end{aligned}$$

$$\begin{aligned} P \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H AP \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H &= P \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H P \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H \\ &= P \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H \end{aligned}$$

$$\left[AP \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H \right]^H = \left[P \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H A \right]^H = P \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H$$

因为 $P \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H$ 是 Hermitian 矩阵, 于是 $P \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H$ 与 A 的 Penrose 方程组相容,

所以 $P \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H = A^+$ 。

(6) 将 $\frac{\alpha^H}{\|\alpha\|^2}$ 代入 α 的 Penrose 方程组可得:

$$\begin{aligned} \alpha \frac{\alpha^H}{\|\alpha\|^2} \alpha &= \alpha \\ \frac{\alpha^H}{\|\alpha\|^2} \alpha \frac{\alpha^H}{\|\alpha\|^2} &= \frac{\alpha^H}{\|\alpha\|^2} \\ \left(\alpha \frac{\alpha^H}{\|\alpha\|^2} \right)^H &= \left(\frac{\alpha^H}{\|\alpha\|^2} \alpha \right)^H = 1 \end{aligned}$$

显然 $\frac{\alpha^H}{\|\alpha\|^2} = \alpha^+$ 。

(7) 由 A^+ 的奇异值分解表示 (定理 1.16) 可得:

$$\begin{aligned} I - A^+ A &= I - Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H = I - Q \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H \\ &= I - \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix} \cong \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \end{aligned}$$

由定理 1.43(3) 的第三条可知 $I - A^+ A \geq \mathbf{0}$ 。

(8) 由 (3) 可得:

$$\begin{aligned} A^+ (A^H)^+ &= A^+ (A^+)^H = Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H P \begin{pmatrix} (\Lambda^H)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H Q^H \\ &= Q \begin{pmatrix} \Lambda^{-1} (\Lambda^H)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H = \begin{pmatrix} \Lambda^{-1} (\Lambda^H)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \end{aligned}$$

由 A 的奇异值分解 (定理 1.46) 可得:

$$\begin{aligned} A^H A &= \left[P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H \right]^H P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H \\ &= Q \begin{pmatrix} \Lambda^H & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H = \begin{pmatrix} \Lambda^H \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \end{aligned}$$

将 $A^+(A^H)^+$ 代入 $A^H A$ 的 Penrose 方程组中即可验证得到 $(A^H A)^+ = A^+(A^H)^+$ 。

(9) 由 (8)、(3) 和 A^+ 的奇异值分解表示 (定理 1.16) 可得:

$$\begin{aligned} (A^H A)^+ A^H &= A^+(A^H)^+ A^H = A^+(A^+)^H A^H \\ &= Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H P \begin{pmatrix} (\Lambda^H)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H Q \begin{pmatrix} \Lambda^H & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H \\ &= Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H = A^+ \\ A^H (A A^H)^+ &= A^H (A^H)^+ A^+ = A^H (A^+)^H A^+ \\ &= Q \begin{pmatrix} \Lambda^H & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H P \begin{pmatrix} (\Lambda^H)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H \\ &= Q \begin{pmatrix} \Lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P^H = A^+ \end{aligned} \quad \square$$

1.5.3 线性方程组的解

Theorem 1.17. 数域 K 上 n 元非齐次线性方程组 $Ax = \beta$ 有解的充分必要条件为对 A 的任一广义逆 A^- 都有:

$$\beta = AA^-\beta$$

Proof. (1) **必要性:** 若 $Ax = \beta$ 有解, 取其一个解 α , 于是对 A 的任一广义逆有:

$$\beta = A\alpha = AA^-A\alpha = AA^-\beta$$

(2) **充分性:** 若此时对 A 的任一广义逆 A^- 有 $\beta = AA^-\beta$, 则方程组可化为:

$$Ax = AA^-\beta$$

容易看出 $A^-\beta$ 就是 $Ax = \beta$ 的一个解。 □

齐次方程组解的结构

Theorem 1.18. 若数域 K 上 n 元齐次线性方程组 $Ax = \mathbf{0}$ 有解, 则它的通解为:

$$x = (I_n - A^-A)y$$

其中 A^- 是 A 的任意一个给定的广义逆, y 取遍 K^n 中的列向量。

Proof. 任取 $y \in K^n$, 有:

$$A(I_n - A^-A)y = Ay - AA^-Ay = Ay - Ay = \mathbf{0}$$

所以对任意的 $y \in K^n$, $(I_n - A^-A)y$ 都是 $Ax = \mathbf{0}$ 的解。

若 η 是 $Ax = \mathbf{0}$ 的一个解, 则:

$$(I_n - A^-A)\eta = \eta - A^-A\eta = \eta - A^-\mathbf{0} = \eta$$

所以 $Ax = \mathbf{0}$ 的任意一个解 x 都可以表示为 $(I_n - A^-A)x$ 的形式。

综上, $Ax = \mathbf{0}$ 的通解为 $x = (I_n - A^-A)y$ 。 □

非齐次方程组解的结构

Theorem 1.19 (结构 1). 若数域 K 上 n 元非齐次线性方程组 $Ax = \beta$ 有解, 则它的通解为:

$$x = A^-\beta + (I_n - A^-A)y$$

其中 A^- 是 A 的任意一个给定的广义逆, y 取遍 K^n 中的列向量。

Proof. 由定理 1.17 的充分性可知对于给定的这一 A^- , $A^-\beta$ 为 $Ax = \beta$ 的一个特解, 而由定理 1.18 可知齐次线性方程组 $Ax = \mathbf{0}$ 的通解为 $(I_n - A^-A)y$, 由 可得 $Ax = \beta$ 的通解为 □

链接线性方程组的关系

Theorem 1.20 (结构 2). 若数域 K 上 n 元非齐次线性方程组 $Ax = \beta$ 有解, 则它的通解为:

$$x = A^-\beta$$

A^- 取遍 A 的所有广义逆。

Proof. 由定理 1.17 的充分性可知对于任意的 A^- , $A^-\beta$ 都是 $Ax = \beta$ 的解。

对于 $Ax = \beta$ 的任意一个解 y , 由定理 1.19 可知存在 A 的一个广义逆 G 和 K^n 上的一个列向量 z , 使得:

$$y = G\beta + (I_n - GA)z$$

因为 $\beta \neq \mathbf{0}$, 所以 $\beta^H\beta \neq 0$, 于是存在数域 K 上的矩阵 $B = z(\beta^H\beta)^{-1}\beta^H$ 使得 $B\beta = z$, 于是:

$$y = G\beta + (I_n - GA)B\beta = [G + (I_n - GA)B]\beta$$

因为:

$$\begin{aligned} A[G + (I_n - GA)B]A &= AGA + A(I_n - GA)BA \\ &= A + ABA - AGABA \\ &= A + ABA - ABA = A \end{aligned}$$

所以 $G + (I_n - GA)B$ 是 A 的一个广义逆, 即 $Ax = \beta$ 的任一解可以表示为 $A^-\beta$ 。 □

Theorem 1.21. 在数域 K 上相容线性方程组 $Ax = \beta$ 的解集中, $x_0 = A^+\beta$ 为长度最小者。

Proof. 由定理 1.19 可知, $Ax = \beta$ 的通解可以表示为:

$$x = A^+\beta + (I - A^+A)y$$

于是:

$$\begin{aligned} \|x\| &= [A^+\beta + (I - A^+A)y]^H [A^+\beta + (I - A^+A)y] \\ &= \|x_0\|^2 + \beta^H (A^+)^H (I - A^+A)y \\ &\quad + y^H (I - A^+A)^H A^+\beta + y^H (I - A^+A)^H (I - A^+A)y \\ &= \|x_0\|^2 + 2\beta^H (A^+)^H (I - A^+A)y + \|(I - A^+A)y\|^2 \end{aligned}$$

由性质 1.5.2(9) 可得:

$$\begin{aligned} (A^+)^H (I - A^+A) &= (A^+)^H - (A^+)^H A^+A = (A^H)^+ - (A^H)^+ A^+A \\ &= (A^H)^+ - [A(A^H)]^+ A = \mathbf{0} \end{aligned}$$

于是有 $2\beta^H (A^+)^H (I - A^+A)y = 0$ 。因为 $\|(I - A^+A)y\| \geq 0$, 等号成立当且仅当 $(I - A^+A)y = \mathbf{0}$, 所以 $x = A^+\beta = x_0$ 时长度最小。 \square

1.6 相似的应用

1.6.1 特征值与特征向量

Definition 1.17. $A \in M_n(K)$ 。如果 K^n 中存在非零列向量 α , 使得:

$$A\alpha = \lambda\alpha, \lambda \in K$$

则称 λ 是 A 的一个特征值 (eigenvalue), α 是 A 属于特征值 λ 的一个特征向量 (eigenvector)。

求解特征值与特征向量

Definition 1.18. $A \in M_n(K)$, 称 $|\lambda I - A|$ 为 A 的特征多项式 (characteristic polynomial)。

Theorem 1.22. $A \in M_n(K)$, 则:

1. λ 是 A 的一个特征值当且仅当 λ 是 A 的特征多项式在数域 K 中的一个根;
2. α 是 A 属于特征值 λ 的一个特征向量当且仅当 α 是齐次线性方程组 $(\lambda I - A)x = \mathbf{0}$ 的一个非零解。

Proof. 显然:

$$\begin{aligned}
 & \lambda \text{ 是 } A \text{ 的一个特征值, } \alpha \text{ 是 } A \text{ 属于 } \lambda \text{ 的一个特征向量} \\
 \Leftrightarrow & A\alpha = \lambda\alpha, \alpha \neq \mathbf{0}, \lambda \in K \\
 \Leftrightarrow & (\lambda I - A)\alpha = \mathbf{0}, \alpha \neq \mathbf{0}, \lambda \in K \\
 \Leftrightarrow & \alpha \text{ 是齐次线性方程组 } (\lambda I - A)x = \mathbf{0} \text{ 的一个非零解, } \lambda \in K \\
 \Leftrightarrow & |\lambda I - A| = 0, \alpha \text{ 是齐次线性方程组 } (\lambda I - A)x = \mathbf{0} \text{ 的一个非零解, } \lambda \in K \\
 \Leftrightarrow & \lambda \text{ 是多项式 } |\lambda I - A| \text{ 在 } K \text{ 中的一个根,} \\
 & \alpha \text{ 是齐次线性方程组 } (\lambda I - A)x = \mathbf{0} \text{ 的一个非零解, } \lambda \in K \quad \square
 \end{aligned}$$

特征向量的性质

Property 1.6.1. $A \in M_n(K)$, 其特征向量具有如下性质:

1. 设 λ 是 A 的一个特征值, 则 A 属于 λ 的所有特征向量构成 K^n 的一个子空间。因此, 把齐次线性方程组 $(\lambda I - A)x = \mathbf{0}$ 的解空间称为 A 属于 λ 的**特征子空间 (eigenspace)**;
2. A 的属于不同特征值的特征向量是线性无关的。

Proof. (1) 任取 $k_1, k_2 \in K$ 和 A 属于特征值 λ 的两个特征向量 α, β , 则

$$A(k_1\alpha + k_2\beta) = k_1A\alpha + k_2A\beta = k_1\lambda\alpha + k_2\lambda\beta = \lambda(k_1\alpha + k_2\beta)$$

于是 $k_1\alpha + k_2\beta$ 也是 A 属于特征值 λ 的特征向量。由??可知 A 属于 λ 的所有特征向量构成 K^n 的一个子空间。

(2) 我们来证明: 设 $\lambda_1, \lambda_2, \dots, \lambda_m$ 是 $A \in M_n(K)$ 的不同的特征值, $a_{j1}, a_{j2}, \dots, a_{jr_j}$ 是 A 属于 λ_j 的线性无关的特征向量, $j = 1, 2, \dots, m$, 则向量组:

$$a_{11}, a_{12}, \dots, a_{1r_1}, a_{21}, a_{22}, \dots, a_{2r_2}, a_{m1}, a_{m2}, \dots, a_{mr_m}$$

线性无关。

1. 证明对 $n = 2$ 成立: 对于 λ_1 和 λ_2 的线性无关的特征向量 $a_{11}, a_{12}, \dots, a_{1r_1}$ 和 $a_{21}, a_{22}, \dots, a_{2r_2}$, 设:

$$k_1a_{11} + k_2a_{12} + \dots + k_{r_1}a_{1r_1} + l_1a_{21} + l_2a_{22} + \dots + l_{r_2}a_{2r_2} = \mathbf{0}$$

两边同乘 A 可得:

$$\begin{aligned}
 & k_1Aa_{11} + k_2Aa_{12} + \dots + k_{r_1}Aa_{1r_1} + l_1Aa_{21} + l_2Aa_{22} + \dots + l_{r_2}Aa_{2r_2} = \mathbf{0} \\
 & k_1\lambda_1a_{11} + k_2\lambda_1a_{12} + \dots + k_{r_1}\lambda_1a_{1r_1} + l_1\lambda_2a_{21} + l_2\lambda_2a_{22} + \dots + l_{r_2}\lambda_2a_{2r_2} = \mathbf{0}
 \end{aligned}$$

因为 $\lambda_1 \neq \lambda_2$, 所以 λ_1, λ_2 不全为 0。设 $\lambda_2 \neq 0$, 在上上个式子两端乘以 λ_2 (若 $\lambda_2 = 0$, 则同乘 λ_1) 得:

$$k_1 \lambda_2 a_{11} + k_2 \lambda_2 a_{12} + \cdots + k_{r_1} \lambda_2 a_{1r_1} + l_1 \lambda_2 a_{21} + l_2 \lambda_2 a_{22} + \cdots + l_{r_2} \lambda_2 a_{2r_2} = \mathbf{0}$$

于是:

$$k_1(\lambda_1 - \lambda_2)a_{11} + k_2(\lambda_1 - \lambda_2)a_{12} + \cdots + k_{r_1}(\lambda_1 - \lambda_2)a_{1r_1} = \mathbf{0}$$

因为 $\lambda_1 \neq \lambda_2$, 所以:

$$k_1 a_{11} + k_2 a_{12} + \cdots + k_{r_1} a_{1r_1} = \mathbf{0}$$

因为 $a_{11}, a_{12}, \dots, a_{1r_1}$ 线性无关, 所以 $k_1 = k_2 = \cdots = k_{r_1} = 0$, 从而:

$$l_1 a_{21} + l_2 a_{22} + \cdots + l_{r_2} a_{2r_2} = \mathbf{0}$$

因为 $a_{21}, a_{22}, \dots, a_{2r_2}$ 线性无关, 所以 $l_1 = l_2 = \cdots = l_{r_2} = 0$ 。

综上, 向量组 $a_{11}, a_{12}, \dots, a_{1r_1}, a_{21}, a_{22}, \dots, a_{2r_2}$ 线性无关。

2. 归纳假设: 假设对 n 个不同的特征值都有上述结论 (即 n 个不同特征值的线性无关的特征向量构成的向量组线性无关), 下面来证明对 $n+1$ 个不同的特征值也成立。

设:

$$k_{11}a_{11} + k_{12}a_{12} + \cdots + k_{1r_1}a_{1r_1} + \cdots + k_{nr_n}a_{nr_n} + l_1a_{(n+1)1} + l_2a_{(n+1)2} + \cdots + l_{r_{n+1}}a_{(n+1)r_{n+1}} = \mathbf{0}$$

两边同乘 A 可得:

$$\begin{aligned} & k_{11}Aa_{11} + k_{12}Aa_{12} + \cdots + k_{1r_1}Aa_{1r_1} + \cdots + k_{nr_n}Aa_{nr_n} \\ & + l_1Aa_{(n+1)1} + l_2Aa_{(n+1)2} + \cdots + l_{r_{n+1}}Aa_{(n+1)r_{n+1}} = \mathbf{0} \\ & k_{11}\lambda_1a_{11} + k_{12}\lambda_1a_{12} + \cdots + k_{1r_1}\lambda_1a_{1r_1} + \cdots + k_{nr_n}\lambda_na_{nr_n} \\ & + l_1\lambda_{n+1}a_{(n+1)1} + l_2\lambda_{n+1}a_{(n+1)2} + \cdots + l_{r_{n+1}}\lambda_{n+1}a_{(n+1)r_{n+1}} = \mathbf{0} \end{aligned}$$

2.1. $\lambda_{n+1} \neq 0$: 若 $\lambda_{n+1} \neq 0$, 则在上上上式两边同乘 λ_{n+1} 可得:

$$\begin{aligned} & k_{11}\lambda_{n+1}a_{11} + k_{12}\lambda_{n+1}a_{12} + \cdots + k_{1r_1}\lambda_{n+1}a_{1r_1} + \cdots + k_{nr_n}\lambda_{n+1}a_{nr_n} \\ & + l_1\lambda_{n+1}a_{(n+1)1} + l_2\lambda_{n+1}a_{(n+1)2} + \cdots + l_{r_{n+1}}\lambda_{n+1}a_{(n+1)r_{n+1}} = \mathbf{0} \end{aligned}$$

于是有:

$$k_{11}(\lambda_{n+1} - \lambda_1)a_{11} + k_{12}(\lambda_{n+1} - \lambda_1)a_{12} + \cdots + k_{1r_1}(\lambda_{n+1} - \lambda_1)a_{1r_1} + \cdots + k_{nr_n}(\lambda_{n+1} - \lambda_n)a_{nr_n} = \mathbf{0}$$

由归纳假定 $a_{11}, a_{12}, \dots, a_{1r_1}, \dots, a_{nr_n}$ 线性无关, 所以

$$k_{11}(\lambda_{n+1} - \lambda_1) = k_{12}(\lambda_{n+1} - \lambda_1) = \cdots = k_{1r_1}(\lambda_{n+1} - \lambda_1) = \cdots = k_{nr_n}(\lambda_{n+1} - \lambda_n) = 0$$

因为 $\lambda_i, i = 1, 2, \dots, n$ 之间互不相同, 所以 $\lambda_{n+1} - \lambda_1, \lambda_{n+1} - \lambda_2, \dots, \lambda_{n+1} - \lambda_n$ 不为 0, 于是 $k_{11} = k_{12} = \cdots = k_{1r_1} = \cdots = k_{nr_n} = 0$, 所以:

$$l_1a_{(n+1)1} + l_2a_{(n+1)2} + \cdots + l_{r_{n+1}}a_{(n+1)r_{n+1}} = \mathbf{0}$$

因为 $a_{(n+1)1}, a_{(n+1)2}, \dots, a_{(n+1)r_{n+1}}$ 线性无关, 所以有 $l_1 = l_2 = \dots = l_{r_{n+1}} = 0$ 。

综上 $a_{11}, a_{12}, \dots, a_{1r_1}, \dots, a_{nr_n}, a_{(n+1)1}, a_{(n+1)2}, \dots, a_{(n+1)r_{n+1}}$ 线性无关。

2.2. $\lambda_{n+1} = 0$: 若 $\lambda_{n+1} = 0$, 则此时有:

$$\begin{aligned} & k_{11}\lambda_1 a_{11} + k_{12}\lambda_1 a_{12} + \dots + k_{1r_1}\lambda_1 a_{1r_1} + \dots + k_{nr_n}\lambda_n a_{nr_n} \\ & + l_1\lambda_{n+1} a_{(n+1)1} + l_2\lambda_{n+1} a_{(n+1)2} + \dots + l_{r_{n+1}}\lambda_{n+1} a_{(n+1)r_{n+1}} \\ & = k_{11}\lambda_1 a_{11} + k_{12}\lambda_1 a_{12} + \dots + k_{1r_1}\lambda_1 a_{1r_1} + \dots + k_{nr_n}\lambda_n a_{nr_n} = \mathbf{0} \end{aligned}$$

由归纳假定 $a_{11}, a_{12}, \dots, a_{1r_1}, \dots, a_{nr_n}$ 线性无关, 所以 $k_{11}\lambda_1 = k_{12}\lambda_1 = \dots = k_{1r_1}\lambda_1 = \dots = k_{nr_n}\lambda_n = 0$ 。因为 $\lambda_1, \lambda_2, \dots, \lambda_n$ 都不是 0 ($\lambda_i, i = 1, 2, \dots, n+1$ 互不相同, 已经有 $\lambda_{n+1} = 0$ 了), 所以 $k_{11} = k_{12} = \dots = k_{1r_1} = \dots = k_{nr_n} = 0$, 于是有:

$$l_1 a_{(n+1)1} + l_2 a_{(n+1)2} + \dots + l_{r_{n+1}} a_{(n+1)r_{n+1}} = \mathbf{0}$$

因为 $a_{(n+1)1}, a_{(n+1)2}, \dots, a_{(n+1)r_{n+1}}$ 线性无关, 所以有 $l_1 = l_2 = \dots = l_{r_{n+1}} = 0$ 。

综上, $a_{11}, a_{12}, \dots, a_{1r_1}, \dots, a_{nr_n}, a_{(n+1)1}, a_{(n+1)2}, \dots, a_{(n+1)r_{n+1}}$ 线性无关。

假设存在属于不同特征值的特征向量 $\alpha_1, \alpha_2, \dots, \alpha_m$ 线性相关, 则有:

$$k_1 \alpha_1 + k_2 \alpha_2 + \dots + k_m \alpha_m = \mathbf{0}$$

其中 k_1, k_2, \dots, k_m 不全为 0。注意到 $\alpha_i, i = 1, 2, \dots, m$ 可由其对应特征值的特征子空间中的一组基线性表出, 于是有:

$$\alpha_i = \sum_{n=1}^{r_i} l_n \beta_{in}$$

其中 $\beta_{in}, n = 1, 2, \dots, r_i$ 为 α_i 对应特征值的特征子空间的一组基, 所以:

$$\sum_{i=1}^m k_i \sum_{n=1}^{r_i} l_n \beta_{in} = \sum_{i=1}^m \sum_{n=1}^{r_i} k_i l_n \beta_{in} = \mathbf{0}$$

而 $\beta_{in}, i = 1, 2, \dots, m, n = 1, 2, \dots, r_i$ 是线性无关的, 所以:

$$k_i l_n = 0, \forall i = 1, 2, \dots, m, n = 1, 2, \dots, r_i$$

因为 k_1, k_2, \dots, k_m 不全为 0, 所以存在一组 l_n 全为 0, 于是 α_i 中存在零向量, 而特征向量不是零向量, 矛盾。□

Theorem 1.23. 相似的矩阵有相同的特征多项式, 进而有相同的特征值 (包括重数相同)。

Proof. 设 $A, B \in M_n(K)$ 且 A 与 B 相似, 于是存在可逆矩阵 $P \in M_n(K)$ 使得 $P^{-1}AP = B$, 就有:

$$\begin{aligned} |\lambda I - B| &= |\lambda I - P^{-1}AP| = |P^{-1}\lambda IP - P^{-1}AP| \\ &= |P^{-1}(\lambda I - A)P| = |P^{-1}| |\lambda I - A| |P| = |\lambda I - A| \end{aligned} \quad \square$$

几何重数与代数重数

Definition 1.19. $A \in M_n(K)$, λ 是 A 的一个特征值。把 A 属于 λ 的特征子空间的维数叫作 λ 的几何重数 (*geometric multiplicity*), 把 λ 作为 A 的特征多项式的根的重数叫作 λ 的代数重数 (*algebraic multiplicity*)。

Theorem 1.24. $A \in M_n(K)$, λ_1 是 A 的一个特征值, 则 λ_1 的几何重数不超过它的代数重数。

Proof. 设 A 属于特征值 λ_1 的特征子空间 W_1 的维数为 r 。在 W_1 中取一个基 $\alpha_1, \alpha_2, \dots, \alpha_r$, 把它扩充为 K^n 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_r, \beta_1, \beta_2, \dots, \beta_{n-r}$ 。令:

$$P = (\alpha_1, \alpha_2, \dots, \alpha_r, \beta_1, \beta_2, \dots, \beta_{n-r})$$

则 P 是数域 K 上的 n 阶可逆矩阵, 并且有:

$$\begin{aligned} P^{-1}AP &= P^{-1}(A\alpha_1, A\alpha_2, \dots, A\alpha_r, A\beta_1, A\beta_2, \dots, A\beta_{n-r}) \\ &= P^{-1}(\lambda_1\alpha_1, \lambda_1\alpha_2, \dots, \lambda_1\alpha_r, A\beta_1, A\beta_2, \dots, A\beta_{n-r}) \\ &= (\lambda_1\varepsilon_1, \lambda_1\varepsilon_2, \dots, \lambda_1\varepsilon_r, P^{-1}A\beta_1, P^{-1}A\beta_2, \dots, P^{-1}A\beta_{n-r}) \\ &= \begin{pmatrix} \lambda_1 I_r & B \\ \mathbf{0} & C \end{pmatrix} \end{aligned}$$

由定理 1.23 可得:

$$\begin{aligned} |\lambda I - A| &= \begin{vmatrix} \lambda I_r - \lambda_1 I_r & -B \\ \mathbf{0} & \lambda I_{n-r} - C \end{vmatrix} \\ &= |\lambda I_r - \lambda_1 I_r| |\lambda I_{n-r} - C| \\ &= (\lambda - \lambda_1)^r |\lambda I_{n-r} - C| \end{aligned}$$

即 λ_1 的几何重数小于或等于 r , 也即 λ_1 的几何重数小于或等于它的代数重数。□

1.6.2 矩阵的对角化

Definition 1.20. 如果 n 阶矩阵 A 能够相似于一个对角矩阵, 那么称 A 可对角化 (*diagonalizable*)。

研究矩阵是否可对角化是为了计算矩阵的幂, 因为对角矩阵的幂是很好计算的。

Theorem 1.25 (矩阵可对角化的第一个充分必要条件). $A \in M_n(K)$ 可对角化的充分必要条件是: A 有 n 个线性无关的特征向量 $\alpha_1, \alpha_2, \dots, \alpha_n$, 此时令 $P = (\alpha_1, \alpha_2, \dots, \alpha_n)$, 则:

$$P^{-1}AP = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$$

其中 λ_i 是 α_i 所属的特征值, $i = 1, 2, \dots, n$ 。上述对角矩阵称为 A 的相似标准形, 除了主对角线上元素的排列次序外, A 的相似标准形是唯一的。

Proof. 显然:

$$\begin{aligned}
 & A \text{ 与对角矩阵 } D = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} \text{ 相似, 其中 } \lambda_i \in K, i = 1, 2, \dots, n \\
 \Leftrightarrow & \text{ 如果存在可逆矩阵 } P \in M_n(K), \text{ 使得 } P^{-1}AP = D \\
 & \text{ 即 } AP = PD \\
 & \text{ 即 } A(\alpha_1, \alpha_2, \dots, \alpha_n) = (\alpha_1, \alpha_2, \dots, \alpha_n)D \\
 & \text{ 即 } (A\alpha_1, A\alpha_2, \dots, A\alpha_n) = (\lambda_1\alpha_1, \lambda_2\alpha_2, \dots, \lambda_n\alpha_n) \\
 \Leftrightarrow & K^n \text{ 中有 } n \text{ 个线性无关的列向量 } \alpha_1, \alpha_2, \dots, \alpha_n \text{ 使得 } A\alpha_i = \lambda_i\alpha_i, i = 1, 2, \dots, n \quad \square
 \end{aligned}$$

Theorem 1.26 (矩阵可对角化的第二个充分必要条件). $A \in M_n(K)$ 可对角化的充分必要条件是: A 的属于不同特征值的特征子空间的维数之和等于 n 。

Proof. (1) 充分性: 由性质 1.6.1(2) 和定理 1.25 的充分性可直接得出。

(2) 必要性: 设 A 的所有不同的特征值是 $\lambda_1, \lambda_2, \dots, \lambda_m$, 它们的几何重数分别为 r_1, r_2, \dots, r_m 。若此时 A 的属于不同特征值的特征子空间的维数之和不等于 n , 由定理 1.24 可知此时 $r_1 + r_2 + \dots + r_m < n$, 那么 A 没有 n 个线性无关的特征向量, 由定理 1.25 的必要性可得 A 不可以对角化。 \square

Corollary 1.1. $A \in M_n(K)$ 如果有 n 个不同的特征值, 那么 A 可对角化。

Theorem 1.27 (矩阵可对角化的第三个充分必要条件). $A \in M_n(K)$ 可对角化的充分必要条件是: A 的特征多项式的全部复根都属于 K , 且 A 的每个特征值的几何重数等于它的代数重数。

Proof. (1) 充分性: 由定理 1.26 的充分性可直接得到。

(2) 必要性: 因为 A 可对角化, 由可对角化的定义可知 A 相似于:

$$\text{diag}(\lambda_1, \dots, \lambda_1, \dots, \lambda_m, \dots, \lambda_m) \in M_n(K)$$

其中 $\lambda_1, \lambda_2, \dots, \lambda_m$ 是 A 的全部不同的特征值, 每个特征值重复的次数为对应特征子空间的维数, λ_i 对应特征子空间的维数记为 $r_i, i = 1, 2, \dots, m$ 。因为相似的矩阵具有相同的特征多项式, 所以:

$$|\lambda I - A| = (\lambda - \lambda_1)^{r_1} (\lambda - \lambda_2)^{r_2} \dots (\lambda - \lambda_m)^{r_m}$$

于是 A 的特征多项式的根为 $\lambda_1, \lambda_2, \dots, \lambda_m$ 。因为 $\text{diag}(\lambda_1, \dots, \lambda_1, \dots, \lambda_m, \dots, \lambda_m) \in M_n(K)$, 所以 $\lambda_1, \lambda_2, \dots, \lambda_m \in K$, 于是 A 的特征多项式的全部根都属于 K 且每一个特征值的代数重数等于它的几何重数。 \square

1.6.3 Hermitian 矩阵的对角化

Definition 1.21. 若对于 $A, B \in M_n(\mathbb{C})$, 存在一个 n 阶正交矩阵 Q , 使得 $Q^{-1}AQ = B$, 则称 A 正交相似于 B 。

Theorem 1.28. 正交相似是 $M_n(\mathbb{C})$ 上的一个等价关系。

Theorem 1.29. $A \in M_n(\mathbb{C})$ 。若 A 正交相似与一个对角矩阵 D ，则 A 一定是 *Hermitian* 矩阵。

Proof. 因为 A 正交相似于 D ，所以存在正交矩阵 Q 使得 $Q^{-1}AQ = D$ ，即 $A = QDQ^{-1}$ ，于是有：

$$A^H = (QDQ^{-1})^H = (Q^{-1})^H D^H Q^H = (Q^H)^H D Q^{-1} = QDQ^{-1} = A$$

所以 A 是一个 *Hermitian* 矩阵。 □

Corollary 1.2. 正交相似一定相似，相似不一定正交相似。

Proof. 设非 *Hermitian* 矩阵 A 相似于一个对角矩阵 D ，若 A 正交相似于 D ，则 A 得是一个 *Hermitian* 矩阵，而 A 不是一个 *Hermitian* 矩阵。 □

Property 1.6.2. 设 *Hermitian* 矩阵 $A, B \in M_n(\mathbb{C})$ ，则：

1. A 的特征多项式的每一个根都是实数，从而都是 A 的特征值；
2. A 属于不同特征值的特征向量是正交的；
3. A 一定正交相似于由它的特征值构成的对角矩阵；
4. A 与 B 正交相似的充分必要条件为 A 与 B 相似。

Proof. (1) 设 λ 是 A 的特征多项式的任意一个根，将 A 看作是复数域 \mathbb{C} 上的矩阵，取 A 属于特征值 λ 的一个特征向量 α ，考虑 \mathbb{C}^n 中的内积，有：

$$\begin{aligned} (A\alpha, \alpha) &= (\lambda\alpha, \alpha) = \lambda(\alpha, \alpha) \\ (\alpha, A\alpha) &= (\alpha, \lambda\alpha) = \bar{\lambda}(\alpha, \alpha) \\ (A\alpha, \alpha) &= (A\alpha)^H \alpha = \alpha^H A^H \alpha = \alpha^H A \alpha = (\alpha, A\alpha) \end{aligned}$$

所以 $\lambda(\alpha, \alpha) = \bar{\lambda}(\alpha, \alpha)$ 。因为 α 是特征向量，所以 $\alpha \neq \mathbf{0}$ ，于是 $\lambda = \bar{\lambda}$ ，因此 λ 是一个实数。由 λ 的任意性，结论成立。

(2) 设 λ_1, λ_2 是 A 的不同的特征值（由 (1) 得它们都是实数）， α_1, α_2 分别是 A 属于 λ_1, λ_2 的一个特征向量，考虑 \mathbb{C}^n 上的标准内积：

$$\begin{aligned} \lambda_1(\alpha_1, \alpha_2) &= (\lambda_1\alpha_1, \alpha_2) = (A\alpha_1, \alpha_2) = A(\alpha_1, \alpha_2) = (\alpha_1, A^H\alpha_2) \\ &= (\alpha_1, A\alpha_2) = (\alpha_1, \lambda_2\alpha_2) = \bar{\lambda}_2(\alpha_1, \alpha_2) = \lambda_2(\alpha_1, \alpha_2) \end{aligned}$$

于是有 $(\lambda_1 - \lambda_2)(\alpha_1, \alpha_2) = 0$ 。因为 $\lambda_1 \neq \lambda_2$ ，所以 $(\alpha_1, \alpha_2) = 0$ 。

(3) 对 n 作数学归纳法。

当 $n = 1$ 时， $(1)^{-1}A(1) = A$ ，结论成立。

假设对于 $n-1$ 阶的实对称矩阵都成立, 考虑 n 阶实对称矩阵 A 。

由 (2) 可知 A 必有特征值, 取 A 的一个特征值 λ_1 和 A 属于 λ_1 的一个特征向量 η_1 , 满足 $\|\eta_1\| = 1$ 。把 η_1 扩充为 \mathbb{C}^n 的一个基并进行 Schmidt 正交化和单位化, 可得到 \mathbb{C}^n 的一个标准正交基 $\eta_1, \eta_2, \dots, \eta_n$ 。令:

$$Q_1 = (\eta_1, \eta_2, \dots, \eta_n)$$

显然 Q_1 是一个正交矩阵, 于是有 $Q_1^{-1}Q_1 = (Q_1^{-1}\eta_1, Q_1^{-1}\eta_2, \dots, Q_1^{-1}\eta_n) = (e_1, e_2, \dots, e_n)$ 。注意到:

$$Q_1^{-1}AQ_1 = Q_1^{-1}(A\eta_1, A\eta_2, \dots, A\eta_n) = (Q_1^{-1}A\eta_1, Q_1^{-1}A\eta_2, \dots, Q_1^{-1}A\eta_n) = \begin{pmatrix} \lambda_1 & \alpha \\ \mathbf{0} & B \end{pmatrix}$$

因为 $(Q_1^{-1}AQ_1)^H = Q_1^H A^H (Q_1^{-1})^H = Q_1^{-1}A(Q_1^H)^H = Q_1^{-1}AQ_1$, 所以 $Q_1^{-1}AQ_1$ 是一个对称阵, 于是 $\alpha = \mathbf{0}$, B 是一个 $n-1$ 阶 Hermitian 阵。由归纳假设, 存在 $n-1$ 阶正交矩阵 Q_2 使得 $Q_2^{-1}BQ_2 = \text{diag}\{\lambda_2, \lambda_3, \dots, \lambda_n\}$ 。令:

$$Q = Q_1 \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_2 \end{pmatrix}$$

则:

$$Q^H Q = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_2^H \end{pmatrix} Q_1^H Q_1 \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_2 \end{pmatrix} = I$$

即 Q 是一个正交矩阵。同时:

$$\begin{aligned} Q^{-1}AQ &= \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_2^H \end{pmatrix} Q_1^H A Q_1 \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_2 \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_2^H \end{pmatrix} \begin{pmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & B \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_2 \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & Q_2^H B Q_2 \end{pmatrix} = \begin{pmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & \text{diag}\{\lambda_2, \lambda_3, \dots, \lambda_n\} \end{pmatrix} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} \end{aligned}$$

所以 A 正交相似于对角矩阵 $\text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 。由 $AQ = Q \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 可以得到 $\lambda_1, \lambda_2, \dots, \lambda_n$ 是 A 的特征值。

综上, 结论成立。

(4) 必要性: 正交相似也是相似。

充分性: 因为 A 与 B 相似, 由定理 1.23 可知 A 与 B 有相同的特征值 (包括重数) $\lambda_1, \lambda_2, \dots, \lambda_n$ 。由 (3) 可得 A 与 B 都正交相似于 $\text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ (考虑 λ_i 的顺序的话只需要更改 Q 中列向量的顺序)。因为正交相似具有对称性与传递性, 所以 A 正交相似于 B 。□

求解正交矩阵 Q

Theorem 1.30. 对于 Hermitian 阵 $A \in M_n(\mathbb{C})$, 求正交矩阵 Q 使得 $Q^{-1}AQ$ 为对角阵的步骤如下:

1. 求出 A 的所有特征值 $\lambda_1, \lambda_2, \dots, \lambda_m$;
2. 对于每一个特征值 λ_i , 求得其特征子空间的一组基 $\alpha_{1i}, \alpha_{2i}, \dots, \alpha_{r_i i}$, 并对它们进行 *Schmidt* 正交化与单位化, 得到 $\eta_{1i}, \eta_{2i}, \dots, \eta_{r_i i}$;
3. 令 $Q = (\eta_{11}, \eta_{21}, \dots, \eta_{r_1 1}, \dots, \eta_{r_m m})$, Q 即为所求。

Schmidt 正
交化链接

Proof. 由可知 η_{ij} , $i = 1, 2, \dots, r_j$, $j = 1, 2, \dots, m$ 是 A 属于 λ_j 的特征值。根据性质 1.6.2(2) 可知:

$$\begin{aligned}
 Q^{-1}AQ &= Q^H(A\eta_{11}, A\eta_{21}, \dots, A\eta_{r_1 1}, \dots, A\eta_{r_m m}) \\
 &= \begin{pmatrix} \eta_{11}^H \\ \eta_{21}^H \\ \vdots \\ \eta_{r_1 1}^H \\ \vdots \\ \eta_{r_m m}^H \end{pmatrix} (\lambda_1 \eta_{11}, \lambda_1 \eta_{21}, \dots, \lambda_1 \eta_{r_1 1}, \dots, \lambda_m \eta_{r_m m}) \\
 &= \text{diag}\{\lambda_1 \eta_{11}^H \eta_{11}, \lambda_1 \eta_{21}^H \eta_{21}, \dots, \lambda_1 \eta_{r_1 1}^H \eta_{r_1 1}, \dots, \lambda_m \eta_{r_m m}^H \eta_{r_m m}\} \\
 &= \text{diag}\{\lambda_1, \dots, \lambda_1, \dots, \lambda_m, \dots, \lambda_m\}
 \end{aligned}$$

□

实对称矩阵特征值的极值性质

Theorem 1.31. 设 $A \in M_n(\mathbb{R})$, A 的特征值从大到小记作 $\lambda_1, \lambda_2, \dots, \lambda_n$, $\varphi_1, \varphi_2, \dots, \varphi_n$ 为对应的标准正交化特征向量, 则:

$$\max_{x \neq 0} \frac{x^T A x}{x^T x} = \lambda_1 = \varphi_1^T A \varphi_1 \quad \min_{x \neq 0} \frac{x^T A x}{x^T x} = \lambda_n = \varphi_n^T A \varphi_n$$

Proof. 由性质 1.6.2(3) 可知存在一个正交矩阵 Q 使得 $Q^{-1}AQ = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} = \Lambda$ 。对任意的 $x \in \mathbb{R}^n$, 因为 Q 为正交矩阵, Q 可逆, 所以关于 y 的非齐次线性方程组 $Qy = x$ 有唯一解, 于是对于这个存在且唯一的 y , 有:

$$\begin{aligned}
 \frac{x^T A x}{x^T x} &= \frac{y^T Q^T A Q y}{y^T Q^T Q y} = \frac{y^T \Lambda y}{y^T y} = \frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2} \leq \lambda_1 \frac{\sum_{i=1}^n y_i^2}{\sum_{i=1}^n y_i^2} = \lambda_1 \\
 \frac{x^T A x}{x^T x} &= \frac{y^T Q^T A Q y}{y^T Q^T Q y} = \frac{y^T \Lambda y}{y^T y} = \frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2} \geq \lambda_n \frac{\sum_{i=1}^n y_i^2}{\sum_{i=1}^n y_i^2} = \lambda_n
 \end{aligned}$$

当 y 为 $(1, 0, 0, \dots, 0)^T$ 时第一式取等号, 当 y 为 $(0, 0, \dots, 0, 1)^T$ 时第二式取等号, 此时 x 分别为 φ_1 和 φ_n 。 □

1.7 合同的应用——二次型

Definition 1.22. 数域 K 上的一个 n 元二次型 (quadratic form) 是系数在 K 中的 n 个变量的二元齐次多项式, 它的一般形式为:

$$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$$

其中 $a_{ij} = a_{ji}$, $1 \leq i, j \leq n$ 。矩阵:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix}$$

被称为二次型 $f(x_1, x_2, \dots, x_n)$ 的矩阵, 它是一个对称矩阵, 主对角元依次是 $x_1^2, x_2^2, \dots, x_n^2$ 的系数, (i, j) 元是 $x_i x_j$ 系数的一半, 其中 $i \neq j$ 。令:

$$x = (x_1, x_2, \dots, x_n)^T$$

则二次型 $f(x_1, x_2, \dots, x_n)$ 可写作 $x^T A x$ 。

Definition 1.23. 令 $x = (x_1, x_2, \dots, x_n)^T$, $y = (y_1, y_2, \dots, y_n)^T$, 可逆矩阵 $C \in M_n(K)$, 则关系式 $x = Cy$ 称为变量 x_1, x_2, \dots, x_n 到变量 y_1, y_2, \dots, y_n 的一个非退化线性变换 (invertible linear transformation)。如果 C 是一个正交矩阵, 则称变量变换 $x = Cy$ 为一个正交变换 (orthogonal transformation)。

Definition 1.24. 对于数域 K 上的两个 n 元二次型 $x^T A x$ 与 $y^T A y$, 如果存在一个非退化线性变换 $x = Cy$, 把 $x^T A x$ 变成 $y^T B y$, 那么称二次型 $x^T A x$ 与 $y^T B y$ 等价, 记作 $x^T A x \cong y^T B y$ 。如果二次型 $x^T A x$ 等价于一个只含平方项的二次型, 那么称这个只含平方项的二次型是 $x^T A x$ 的一个标准形。

Theorem 1.32. 数域 K 上两个 n 元二次型 $x^T A x$ 与 $y^T B y$ 等价当且仅当 n 阶对称矩阵 A 与 B 合同, 于是二次型的等价也是一个等价关系。

Proof. (1) 充分性: 因为 $A \cong B$, 所以存在可逆矩阵 C 使得 $C^T A C = B$ 。作非退化线性变换 $x = Cy$, 可得到 $(Cy)^T A (Cy) = y^T C^T A C y = y^T B y$, 所以 $x^T A x \cong y^T B y$ 。

(2) 必要性: 因为 $x^T A x \cong y^T B y$, 所以存在非退化线性变换 $x = Cy$, C 是一个可逆矩阵, 把 $x^T A x$ 变为 $y^T B y$, 即 $(Cy)^T A (Cy) = y^T C^T A C y = y^T B y$, 所以 $C^T A C = B$, 即 $A \cong B$ 。

因为合同是一个等价关系, 显然可得二次型的等价也是一个等价关系。 \square

Theorem 1.33. 数域 K 上任一 n 元二次型都等价于一个只含平方项的二次型。

Proof. 当二次型的矩阵是对角矩阵时该二次型只含平方项, 由定理 1.10 与定理 1.32 可立即得出结论. \square

Theorem 1.34. 设 n 元二次型 $x^T Ax$ 的矩阵 A 合同于对角矩阵 $D = \text{diag}\{d_1, d_2, \dots, d_n\}$, 即存在可逆矩阵 C 使得 $C^T AC = D$. 令 $x = Cy$, 则可以得到 $x^T Ax$ 的一个标准形:

$$d_1 y_1^2 + d_2 y_2^2 + \cdots + d_n y_n^2$$

Proof. 将 $x = Cy$ 代入可得:

$$x^T Ax = (Cy)^T A(Cy) = y^T C^T ACy = y^T Dy = \sum_{i=1}^n d_i y_i^2 \quad \square$$

Theorem 1.35. 数域 K 上 n 元二次型 $x^T Ax$ 的任一标准形中, 系数不为 0 的平方项个数等于它的矩阵 A 的秩。

Proof. 设 n 元二次型 $x^T Ax$ 经过非退化线性变换 $x = Cy$ 化成标准形 $d_1 y_1^2 + d_2 y_2^2 + \cdots + d_r y_r^2$, 其中 d_1, d_2, \dots, d_r 都不为 0, 则:

$$C^T AC = \text{diag}\{d_1, d_2, \dots, d_r, 0, \dots, 0\}$$

于是 $\text{diag}\{d_1, d_2, \dots, d_r, 0, \dots, 0\}$ 是 A 的一个合同标准形。由定理 1.11 可得 $\text{rank}(A) = r$. \square

Definition 1.25. 称二次型 $x^T Ax$ 的矩阵 A 的秩为二次型 $x^T Ax$ 的秩。

1.7.1 二次型的规范形

实二次型的规范形

Definition 1.26. 实数域上的二次型称为实二次型。由定理 1.12 可知 n 元实二次型 $x^T Ax$ 的矩阵 A 合同于一个对角矩阵 $\text{diag}\{1, 1, \dots, 1, -1, -1, \dots, -1, 0, 0, \dots, 0\}$, 再由定理 1.32 可知经过一个适当的非退化线性变换可以将 $x^T Ax$ 化作:

$$z_1^2 + z_2^2 + \cdots + z_p^2 - z_{p+1}^2 - z_{p+2}^2 - \cdots - z_r^2$$

称此形式为二次型 $x^T Ax$ 的规范形, 其特征为: 只含平方项且平方项系数为 1, -1, 0, 系数为 1 的平方项在最前面, 系数为 -1 的平方项在中间, 系数为 0 的平方项在最后。实二次型 $x^T Ax$ 的规范形被两个自然数 p 和 r 决定。

Theorem 1.36 (Sylvester's Law of Inertia). n 元实二次型 $x^T Ax$ 的规范形是唯一的。

Proof. 设 n 元实二次型 $x^T Ax$ 的秩为 r , 假设 $x^T Ax$ 分别经过非退化线性变换 $x = Cy$ 和 $x = Bz$ 变成两个规范形:

$$\begin{aligned} x^T Ax &= y_1^2 + y_2^2 + \cdots + y_p^2 - y_{p+1}^2 - y_{p+2}^2 - \cdots - y_r^2 \\ x^T Ax &= z_1^2 + z_2^2 + \cdots + z_q^2 - z_{q+1}^2 - z_{q+2}^2 - \cdots - z_r^2 \end{aligned}$$

要证规范形唯一，即证 $p = q$ 。

由 $x = Cy$ 和 $x = Bz$ 可知，经过非退化线性变换 $z = (B^{-1}C)y$ 后有：

$$z_1^2 + z_2^2 + \cdots + z_q^2 - z_{q+1}^2 - z_{q+2}^2 - \cdots - z_r^2 = y_1^2 + y_2^2 + \cdots + y_p^2 - y_{p+1}^2 - y_{p+2}^2 - \cdots - y_r^2$$

记 $D = B^{-1}C = (d_{ij})$ 。假设 $p > q$ ，我们想找到变量 y_1, y_2, \dots, y_n 的一组取值，使得上式右端大于 0，而左端小于或等于 0，从而产生矛盾。令：

$$y = (y_1, y_2, \dots, y_p, 0, 0, \dots, 0)^T$$

其中 y_1, y_2, \dots, y_p 是待定的实数，使得变量 z_1, z_2, \dots, z_q 的值全为 0。因为 $z = Dy$ ，所以：

$$\begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1p} \\ d_{21} & d_{22} & \cdots & d_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ d_{q1} & d_{q2} & \cdots & d_{qp} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_q \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

因为 $p > q$ ，所以上述齐次线性方程组有非零解，即存在非零向量 $y = (y_1, y_2, \dots, y_p, 0, 0, \dots, 0)^T$ 使得 $z_1 = z_2 = \cdots = z_q = 0$ 。此时有：

$$\begin{aligned} z_1^2 + z_2^2 + \cdots + z_q^2 - z_{q+1}^2 - z_{q+2}^2 - \cdots - z_r^2 &\leq 0 \\ y_1^2 + y_2^2 + \cdots + y_p^2 - y_{p+1}^2 - y_{p+2}^2 - \cdots - y_r^2 &> 0 \end{aligned}$$

矛盾。因此 $p \leq q$ 。同理可得 $q \leq p$ ，于是 $p = q$ ，规范形唯一。 \square

Definition 1.27. 在实二次型 $x^T Ax$ 的规范形中，系数为 1 的平方项个数 p 称为 $x^T Ax$ 的正惯性指数，系数为 -1 的平方项个数 $r - p$ 称为 $x^T Ax$ 的负惯性指数，正惯性指数减去负惯性指数所得的差 $2p - r$ 称为 $x^T Ax$ 称为 $x^T Ax$ 的符号差 (signature)。

Theorem 1.37. 两个 n 元实二次型等价

\Leftrightarrow 它们的规范形相同

\Leftrightarrow 它们的秩相等，并且正惯性指数也相等。

Proof. 第一条由定理 1.36 以及二次型等价的传递性、对称性可直接得到（必要性的证明中需要考虑规范形的定义，然后使用定理 1.36），第二条是显然的。 \square

显然矩阵 A 的正惯性指数与负惯性指数就等于二次型 $x^T Ax$ 的正惯性指数与负惯性指数，也等于 A 的合同标准形主对角线上大于 0 的元素的个数与小于 0 的个数。

Theorem 1.38. 两个 n 阶实对称矩阵合同 \Leftrightarrow 它们的秩相等，并且正惯性指数也相等。

Proof. 由定理 1.32 可得矩阵合同等价于各自对应的二次型等价，再由定理 1.37 可得两个二次型的秩与正惯性指数都相等。因为矩阵的秩与正惯性指数等于对应的二次型的秩与正惯性指数，所以结论成立。 \square

复二次型的规范形

Definition 1.28. 复数域上的二次型称为复二次型。由定理 1.13 可知 n 元复二次型 $x^T Ax$ 的矩阵 A 合同于一个对角矩阵 $\text{diag}\{1, 1, \dots, 1, 0, 0, \dots, 0\}$, 再由定理 1.32 可知经过一个适当的非退化线性变换可以将 $x^T Ax$ 化作:

$$z_1^2 + z_2^2 + \dots + z_r^2$$

称此形式为二次型 $x^T Ax$ 的规范形, 其特征为: 只含平方项且平方项系数为 1, 0, 系数为 1 的平方项在前面, 系数为 0 的平方项在后面。

Theorem 1.39. 复二次型 $x^T Ax$ 的规范形是唯一的。

Proof. 复二次型 $x^T Ax$ 的规范形完全由它的秩 r 所决定。 \square

Theorem 1.40. 两个 n 元复二次型等价

\Leftrightarrow 它们的规范形相同

\Leftrightarrow 它们的秩相等。

Proof. 第一条由定理 1.39 以及二次型的传递性、对称性可直接得到 (必要性的证明中需要考虑规范形的定义, 然后使用定理 1.39), 第二条是显然的。 \square

1.7.2 正定二次型与正定矩阵

Definition 1.29. 如果对 \mathbb{R}^n 中任意非零列向量 α , 都有 $\alpha^T A \alpha > 0$, 则称 n 元实二次型 $x^T Ax$ 是正定 (positive definite) 的。

Definition 1.30. 若实二次型 $x^T Ax$ 是正定的, 则称实对称矩阵 A 是正定的, 并称 A 为正定矩阵 (positive definite matrix), 记为 $A > 0$ 。

Theorem 1.41. n 元实二次型 $x^T Ax$ 是正定的当且仅当它的正惯性指数等于 n 。

Proof. (1) 必要性: 设 $x^T Ax$ 是正定的, 作非退化线性变换 $x = Cy$ 化成规范形:

$$y_1^2 + y_2^2 + \dots + y_p^2 - y_{p+1}^2 - y_{p+2}^2 - \dots - y_r^2$$

如果 $p < n$, 则 y_n^2 的系数为 0 或 -1 , 取 $y = (0, 0, \dots, 1)^T$, 则有 $y^T C^T A C y = -y_n^2$ 为 0 或 -1 , 取 $\alpha = Cy$ 即有 $\alpha^T A \alpha$ 为 0 或 -1 , 与二次型 $x^T Ax$ 的正定性矛盾, 所以 $p = n$ 。

(2) 充分性: 设 $x^T Ax$ 的正惯性指数等于 n , 则可以作一个非退化线性变换 $x = Cy$ 将该二次型化作规范形:

$$y^T C^T A C y = y_1^2 + y_2^2 + \dots + y_n^2$$

因为矩阵 C 可逆, 所以关于 y 的齐次线性方程组 $C^{-1}x = \mathbf{0}$ 只有零解。任取非零向量 $\alpha \in \mathbb{R}^n$, 则 $C^{-1}\alpha$ 不是零向量, 令 $y = C^{-1}\alpha$, 于是 $\alpha^T (C^{-1})^T C^T A C C^{-1}\alpha > 0$, 即 $\alpha^T A \alpha > 0$ 。由 α 的任意性, $x^T Ax$ 是正定的。 \square

Theorem 1.42. 由上述定理可得到如下推论:

1. 对于 n 元实二次型 $x^T Ax$, 下述说法等价:

- $x^T Ax$ 是正定的;
- $x^T Ax$ 的规范形为 $y_1^2 + y_2^2 + \cdots + y_n^2$;
- $x^T Ax$ 的标准形中的 n 个系数都大于 0;

2. 与正定二次型等价的实二次型也是正定的;

3. 对于 n 阶实对称矩阵 A , 下述说法等价:

- A 是正定的;
- A 的正惯性指数为 n ;
- $A \cong I$;
- A 的合同标准形中主对角元都大于 0;
- A 的特征值都大于 0;
- A 的顺序主子式都大于 0。

4. 与正定矩阵合同的实对称矩阵也是正定矩阵。

5. 正定矩阵的行列式大于 0;

Proof. (1) $1 \Leftrightarrow 2$: 由上一定理, $x^T Ax$ 正定当且仅当它的正惯性指数为 n , 而 $x^T Ax$ 的正惯性指数为 n 当且仅当它的规范形为 $y_1^2 + y_2^2 + \cdots + y_n^2$ 。

$2 \Rightarrow 3$: 由标准形化规范形的步骤, 若 $x^T Ax$ 的规范形为 $y_1^2 + y_2^2 + \cdots + y_n^2$, 则其标准形中的 n 个系数必然都大于 0;

$3 \Rightarrow 2$: 当 $x^T Ax$ 的标准形中的 n 个系数都大于 0 时, 也必然可以将其化为 $y_1^2 + y_2^2 + \cdots + y_n^2$ 。

(2) 由 (4)、定理 1.32 和正定矩阵的定义可直接得到。

(3) $1 \Rightarrow 2$: 因为 A 是正定的, 所以 n 元二次型 $x^T Ax$ 是正定的, 由上一定理可得 $x^T Ax$ 的正惯性指数为 n 。因为 A 的正惯性指数等于 $x^T Ax$ 的正惯性指数, 所以 A 的正惯性指数为 n 。

$2 \Rightarrow 3$: 因为 A 的正惯性指数为 n , 由矩阵正惯性指数的定义, A 合同于 I 。

$3 \Rightarrow 4$: 因为 A 合同于 I , 由合同规范形的定义, I 是 A 的合同规范形, 由合同标准型化合同规范形的步骤, A 的合同标准型中主对角元都大于 0。

$4 \Rightarrow 5$: 由性质 1.6.2(3) 可知 $A \cong \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, 其中 $\lambda_i, i = 1, 2, \dots, n$ 是 A 的特征值。显然 $\text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 是 A 的一个合同标准型, 因为 A 的合同标准型中主对角元都大于 0, 所以 A 的特征值都大于 0。

$5 \Rightarrow 2$: 显然。

$2 \Rightarrow 1$: 由定理 1.32、上一定理和矩阵正定的定义可直接得到。

1 \Rightarrow 6: 设 n 阶实对称矩阵 A 是正定的, 则对于 $k = 1, 2, \dots, n-1$, 把 A 写成分块矩阵:

$$A = \begin{pmatrix} A_k & B_1 \\ B_1^T & B_2 \end{pmatrix}$$

其中 $|A_k|$ 是 A 的 k 阶顺序主子式。在 \mathbb{R}^k 中任取一个非零向量 δ , 因为 A 是正定矩阵, 所以:

$$\begin{pmatrix} \delta \\ \mathbf{0} \end{pmatrix}^T A \begin{pmatrix} \delta \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \delta^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} A_k & B_1 \\ B_1^T & B_2 \end{pmatrix} \begin{pmatrix} \delta \\ \mathbf{0} \end{pmatrix} = \delta^T A_k \delta > 0$$

由 δ 的任意性, A_k 是正定矩阵。由 (5), $|A_k| > 0, k = 1, 2, \dots, n-1, |A| > 0$ 。

6 \Rightarrow 1: 对实对称矩阵 A 的阶数 n 作数学归纳法。

当 $n = 1$ 时, 因为 A 的顺序主子式都大于 0, 所以 A 的唯一一个元素大于 0, 显然此时 A 是正定矩阵。

假设对于 $n-1$ 阶实对称矩阵命题为真, 考虑 n 阶实对称矩阵 $A = (a_{ij})$, 将其写作分块矩阵的形式:

$$A = \begin{pmatrix} A_{n-1} & \alpha \\ \alpha^T & a_{nn} \end{pmatrix}$$

其中 A_{n-1} 是 $n-1$ 阶实对称矩阵, 因为 A_{n-1} 的所有顺序主子式是 A 的 1 到 $n-1$ 阶顺序主子式, 它们都大于 0, 由归纳假设可得 A_{n-1} 是正定的。根据 (5) 可知 A_{n-1} 可逆。由 (3) 的第三条可知存在可逆矩阵 $C \in M_{n-1}(\mathbb{R})$ 使得 $C^T A_{n-1} C = I$ 。因为:

$$\begin{pmatrix} I & \mathbf{0} \\ -\alpha^T A_{n-1}^{-1} & 1 \end{pmatrix} \begin{pmatrix} A_{n-1} & \alpha \\ \alpha^T & a_{nn} \end{pmatrix} \begin{pmatrix} I & -A_{n-1}^{-1} \alpha \\ \mathbf{0} & 1 \end{pmatrix} = \begin{pmatrix} A_{n-1} & \mathbf{0} \\ \mathbf{0} & a_{nn} - \alpha^T A_{n-1}^{-1} \alpha \end{pmatrix}$$

注意到:

$$\begin{pmatrix} I & \mathbf{0} \\ -\alpha^T A_{n-1}^{-1} & 1 \end{pmatrix}^T = \begin{pmatrix} I & (-\alpha^T A_{n-1}^{-1})^T \\ \mathbf{0} & 1 \end{pmatrix} = \begin{pmatrix} I & -A_{n-1}^{-1} \alpha \\ \mathbf{0} & 1 \end{pmatrix}$$

且:

$$\begin{pmatrix} I & \mathbf{0} \\ -\alpha^T A_{n-1}^{-1} & 1 \end{pmatrix}$$

可逆, 所以 A 合同于矩阵:

$$\begin{pmatrix} A_{n-1} & \mathbf{0} \\ \mathbf{0} & a_{nn} - \alpha^T A_{n-1}^{-1} \alpha \end{pmatrix}$$

因为:

$$\begin{aligned} \begin{vmatrix} A_{n-1} & \mathbf{0} \\ \mathbf{0} & a_{nn} - \alpha^T A_{n-1}^{-1} \alpha \end{vmatrix} &= \begin{vmatrix} I & \mathbf{0} \\ -\alpha^T A_{n-1}^{-1} & 1 \end{vmatrix} \begin{vmatrix} A_{n-1} & \alpha \\ \alpha^T & a_{nn} \end{vmatrix} \begin{vmatrix} I & -A_{n-1}^{-1} \alpha \\ \mathbf{0} & 1 \end{vmatrix} \\ &= \begin{vmatrix} A_{n-1} & \alpha \\ \alpha^T & a_{nn} \end{vmatrix} = |A| \end{aligned}$$

可逆矩阵行列式链接

所以 $|A_{n-1}|(a_{nn} - \alpha^T A_{n-1}^{-1} \alpha) = |A| > 0$, 而 $|A_{n-1}| > 0$, 所以 $a_{nn} - \alpha^T A_{n-1}^{-1} \alpha > 0$. 因为:

$$\begin{aligned} & \begin{pmatrix} C & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}^T \begin{pmatrix} A_{n-1} & \mathbf{0} \\ \mathbf{0} & a_{nn} - \alpha^T A_{n-1}^{-1} \alpha \end{pmatrix} \begin{pmatrix} C & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} \\ &= \begin{pmatrix} C^T A_{n-1} C & \mathbf{0} \\ \mathbf{0} & a_{nn} - \alpha^T A_{n-1}^{-1} \alpha \end{pmatrix} = \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & a_{nn} - \alpha^T A_{n-1}^{-1} \alpha \end{pmatrix} \end{aligned}$$

而:

$$B = \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & a_{nn} - \alpha^T A_{n-1}^{-1} \alpha \end{pmatrix}$$

主对角线上的元素都大于 0, 由 (3) 的第四条可知 B 是一个正定矩阵。因为 $|C|1 = |C| \neq 0$, 所以:

$$\begin{pmatrix} C & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}$$

可逆矩阵行列式链接

可逆。于是:

$$\begin{pmatrix} A_{n-1} & \mathbf{0} \\ \mathbf{0} & a_{nn} - \alpha^T A_{n-1}^{-1} \alpha \end{pmatrix}$$

合同于 B 。根据合同的传递性, A 合同于正定矩阵 B 。由 (4), A 是一个正定矩阵。

(4) 设 A 是一个正定矩阵, B 是一个实对称矩阵且合同于 A 。由 (3) 的第三条可知 A 合同于 I , 根据合同的传递性, B 也合同于 I 。再由 (3) 的第三条可得 B 也是一个正定矩阵。

(5) 设 A 是一个正定矩阵, 由 (3) 的第三条可得 $A \cong I$, 即存在可逆矩阵 C , 使得 $C^T A C = I$, 于是:

$$|C^T A C| = |C^T| |A| |C| = |A| |C|^2 = 1$$

因为 $|C|^2 > 0$, 所以 $|A| > 0$ 。 □

半正定二次型与半正定矩阵

Definition 1.31. 如果对 \mathbb{R}^n 中任意非零列向量 α , 都有 $\alpha^T A \alpha \geq 0$, 则称 n 元实二次型 $x^T A x$ 是半正定 (positive semidefinite) 的。

Definition 1.32. 若实二次型 $x^T A x$ 是半正定的, 则称实对称矩阵 A 是半正定的, 并称 A 为半正定矩阵 (positive semidefinite matrix), 记为 $A \geq 0$ 。

Theorem 1.43. 由上述定理可得到如下推论:

1. 对于 n 元实二次型 $x^T A x$, $\text{rank}(A) = r$, 下述说法等价:

- $x^T A x$ 是半正定的;
- $x^T A x$ 的正惯性指数等于 r ;
- $x^T A x$ 的规范形为 $y_1^2 + y_2^2 + \cdots + y_r^2$;
- $x^T A x$ 的标准形中的 n 个系数都非负;

2. 与半正定二次型等价的实二次型也是半正定的;
3. 对于 n 阶实对称矩阵 A , $\text{rank}(A) = r$, 下述说法等价:

- A 是半正定的;
- A 的正惯性指数为 r ;
- $A \cong \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$;
- A 的合同标准形中主对角元都非负;
- A 的特征值都非负;
- A 的主子式都非负。

4. 与半正定矩阵合同的实对称矩阵也是半正定矩阵。

5. 半正定矩阵的行列式为 0;

Proof. (1) $1 \Rightarrow 3$: 作非退化线性变换 $x = Cy$ 把 $x^T Ax$ 化作规范形:

$$y_1^2 + y_2^2 + \cdots + y_p^2 - y_{p+1}^2 - y_{p+2}^2 - \cdots - y_r^2$$

若 $p < r$, 取 $\alpha = (0, 0, \dots, 0, 1, 0, 0, \dots, 0)$, 其中只有第 r 位为 1, 则 $(C\alpha)^T A(C\alpha) = \alpha C^T A C \alpha = -1$, 与 $x^T Ax$ 的非负定性矛盾, 所以 $p = r$ 。

$3 \Rightarrow 2$: 显然。

$2 \Rightarrow 4$: 显然。

$4 \Rightarrow 1$: 作非退化线性变换 $x = Cy$ 把 $x^T Ax$ 化作一个标准形 $d_1 y_1^2 + d_2 y_2^2 + \cdots + d_n y_n^2$, 其中 $d_i \geq 0, i = 1, 2, \dots, n$ 。任取 $\alpha \in \mathbb{R}^n$ 且 $\alpha \neq \mathbf{0}$ 。因为 C 可逆, 所以 $C^{-1}\alpha = \mathbf{0}$ 只有零解, 于是 $C^{-1}\alpha = (b_1, b_2, \dots, b_n) \neq \mathbf{0}$, 所以:

$$(C^{-1}\alpha)^T C^T A C C^{-1}\alpha = \sum_{i=1}^n d_i b_i^2 \geq 0$$

而:

$$(C^{-1}\alpha)^T C^T A C C^{-1}\alpha = \alpha^T (C^{-1})^T C^T A C C^{-1}\alpha = \alpha^T (C^T)^{-1} C^T A C C^{-1}\alpha = \alpha^T A \alpha$$

所以 $\alpha^T A \alpha \geq 0$ 。由 α 的任意性, $x^T Ax$ 半正定。

(2) 由 (4)、定理 1.32 和半正定矩阵的定义可直接得到。

(3) $1 \Rightarrow 2$: 因为 A 是半正定的, 所以 $x^T Ax$ 是半正定的。由 (1) 的第二条, $x^T Ax$ 的正惯性指数等于 r , 而 A 的正惯性指数等于 $x^T Ax$ 的正惯性指数, 所以 A 的正惯性指数为 r 。

$2 \Rightarrow 3$: 因为 A 的正惯性指数为 r , 由矩阵正惯性指数的定义, $A \cong \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ 。

$3 \Rightarrow 4$: 因为 $A \cong C = \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, 所以 C 是 A 的合同规范形。由合同标准形化合同规范形的步骤, A 的合同标准形中主对角元都大于 0。

4 \Rightarrow 5: 由性质 1.6.2(3) 可知 $A \cong \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, 其中 $\lambda_i, i = 1, 2, \dots, n$ 是 A 的特征值。显然 $\text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 是 A 的一个合同标准型, 因为 A 的合同标准型中主对角元都非负, 所以 A 的特征值都非负。

5 \Rightarrow 2: 因为 $\text{rank} = r$, 所以 A 的相似标准形主对角线上的元素有且只有 r 个非零, 由条件它们也非负, 于是它们为正数, 显然此时 A 的正惯性指数为 r 。

2 \Rightarrow 1: 由定理 1.32、(1) 的第二条和矩阵半正定的定义可直接得到。

1 \Rightarrow 6:

6 \Rightarrow 5:

有空证明

(4) 设 A 是一个半正定矩阵, B 是一个实对称矩阵且合同于 A 。由 (3) 的第三条可知 $A \cong C = \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, 根据合同的传递性, $B \cong C$ 。再由 (3) 的第三条可得 B 也是一个半正定矩阵。

(5) 设 A 是一个 n 阶半正定矩阵, 由 (3) 的第三条, 存在可逆矩阵 C 使得:

$$C^T A C = B = \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

而 $\text{rank}(B) = r$, 因为可逆变换不改变矩阵的秩, 所以 $\text{rank}(A) = r < n$, 于是 $|A| = 0$ 。□

负定矩阵

Definition 1.33. 如果对 \mathbb{R}^n 中任意非零列向量 α , 都有 $\alpha^T A \alpha < 0$, 则称 n 元实二次型 $x^T A x$ 是负定 (negative definite) 的。

Definition 1.34. 若实二次型 $x^T A x$ 是负定的, 则称实对称矩阵 A 是负定的, 并称 A 为负定矩阵 (negative definite matrix), 记为 $A < 0$ 。

Theorem 1.44. 对称矩阵 $A \in M_n(\mathbb{R})$ 负定的充分必要条件为: 它的奇数阶顺序主子式都小于 0, 偶数阶顺序主子式都大于 0。

Proof. 设 $|A_k|$ 为 A 的 k 阶顺序主子式, 由定理 1.42(3) 的第六条:

$$\begin{aligned} & A \text{ 是负定矩阵} \\ \Leftrightarrow & (-A) \text{ 是正定矩阵} \\ \Leftrightarrow & (-1)^k |A_k| > 0 \\ \Leftrightarrow & \begin{cases} |A_k| > 0, & k \text{ 为偶数} \\ |A_k| < 0, & k \text{ 为奇数} \end{cases} \end{aligned}$$

□

1.8 特殊矩阵

1.8.1 幂等阵

幂等阵

Property 1.8.1. 设 $A \in M_n(K)$ 是一个幂等阵, $\text{rank}(A) = r$, 则:

1. A 的特征值只能是 1 或 0;
2. $\text{tr}(A) = \text{rank}(A)$;
3. $\text{rank}(A) + \text{rank}(I_n - A) = n$;
4. 存在秩为 r 的 $B \in M_n(K)$ 使得 $A = B(B^T B)^{-1} B^T$;

Proof. (1) 设 λ 为 A 的一个特征值, φ 为对应的特征向量, 因为 A 是一个幂等阵, 所以 $A^2\varphi = A\varphi = \lambda\varphi$, 又因为:

$$A^2\varphi = AA\varphi = A\lambda\varphi = \lambda A\varphi = \lambda^2\varphi$$

所以 $(\lambda^2 - \lambda)\varphi = \mathbf{0}$ 。因为 φ 是特征向量, 所以 $\varphi \neq \mathbf{0}$, 于是 $\lambda^2 - \lambda = 0$, 即 $\lambda = 1$ 或 $\lambda = 0$ 。由 λ 的任意性, 结论成立。 □

Property 1.8.2. 设 $A \in M_n(\mathbb{C})$ 是一个 *Hermitian* 幂等阵, 则:

- 1.
2. $\text{rank}(A) = \text{tr}(A)$ 。

Proof. (1) (2) 由性质 1.6.2(3) 和 (1) 可知存在一个正交矩阵 Q 使得:

$$A = Q^{-1} \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q$$

由性质 1.4.1 可得:

$$\text{rank}(A) = \text{rank}(I_r) = r$$

根据??(3) 可得:

$$\text{tr}(A) = \text{tr} \left[Q^{-1} \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q \right] = \text{tr} \left[\begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q Q^{-1} \right] = \text{tr} \left[\begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \right] = r$$

所以有 $\text{rank}(A) = \text{tr}(A)$ 。 □

1.9 矩阵的分解

1.9.1 SVD 分解

Theorem 1.45. 设 $A \in M_{m \times n}(\mathbb{C})$, 则 $AA^H, A^H A$ 是半正定矩阵。

Proof. 设 $\lambda_i, i = 1, 2, \dots, n$ 是矩阵 $A^H A$ 的特征值, ξ_i 是对应的特征向量, 则:

$$A^H A \xi_i = \lambda_i \xi_i \rightarrow \xi_i^H A^H A \xi_i = \lambda_i \xi_i^H \xi_i \rightarrow \|A \xi_i\|^2 = \lambda_i \|\xi_i\|^2$$

由于左式非负, 所以右式非负, 而 $\|\xi_i\|^2$ 非负, 因此 λ_i 非负, 由定理 1.43(3) 的第五条可知 AA^T 是半正定矩阵。□

Theorem 1.46. 设 $A \in M_{m \times n}(\mathbb{C})$, $\text{rank}(A) = r$, 则存在两个正交矩阵 $P \in M_m(\mathbb{C})$, $Q \in M_n(\mathbb{C})$ 使得:

$$A = P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H$$

其中 $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_r\}$, $\lambda_i > 0$, λ_i^2 为 $A^H A$ 的正特征值。

Proof. 由定理 1.2 可知 $\text{rank}(A^H A) = \text{rank}(A)$ 。于是 $A^H A$ 确实有 r 个正特征值。因为 $A^H A$ 是一个 Hermitian 矩阵, 由性质 1.6.2 可知存在正交矩阵 $Q \in M_n(\mathbb{C})$ 使得:

$$Q^H A^H A Q = \begin{pmatrix} \Lambda^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

记 $B = AQ$, 则:

$$B^H B = \begin{pmatrix} \Lambda^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

这表明 B 的列向量相互正交, 且前 r 个列向量的长度分别为 $\lambda_1, \lambda_2, \dots, \lambda_r$, 后 $n - r$ 个列向量为零向量, 于是存在一个正交矩阵 $P \in M_m(\mathbb{C})$ 使得:

$$B = P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

因为 $B = AQ$, 所以:

$$A = P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^{-1} = P \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H \quad \square$$

Definition 1.35. 设 $A \in M_{m \times n}(\mathbb{C})$, $\text{rank}(A) = r$, $A^H A$ 的正特征值为 $\lambda_i, i = 1, 2, \dots, r$, 称 $\delta_i = \sqrt{\lambda_i}$ 为矩阵 A 的奇异值 (singular value)。

Chapter 2

线性模型

Definition 2.1.

$$\begin{cases} y = X\beta + \varepsilon \\ E(\varepsilon) = \mathbf{0} \\ \text{Cov}(\varepsilon) = \sigma^2 I_n \end{cases}$$

其中 y 为 $n \times 1$ 观测向量, X 为 $n \times p$ 的设计矩阵, β 为 $p \times 1$ 的未知参数向量, ε 为随机误差, σ^2 为误差方差。

Definition 2.2. 称方程 $X^T X\beta = X^T y$ 为正则方程。

Theorem 2.1. 对于定义 2.1, $\hat{\beta} = (X^T X)^- X^T y$ 是其唯一的最小二乘解。

Proof. 注意到:

$$\begin{aligned} Q(\beta) &= \|y - X\beta\|^2 = (y - X\beta)^T (y - X\beta) \\ &= y^T y - y^T X\beta - \beta^T X^T y - \beta^T X^T X\beta \\ &= y^T y - 2y^T X\beta - \beta^T X^T X\beta \\ \frac{\partial y^T X\beta}{\partial \beta} &= X^T y, \quad \frac{\partial \beta^T X^T X\beta}{\partial \beta} = 2X^T X\beta \\ \frac{\partial Q(\beta)}{\partial \beta} &= 2X^T y - 2X^T X\beta = 0 \\ X^T X\beta &= X^T y \end{aligned}$$

由定理 1.1 可知方程 $X^T X\beta = X^T y$ 是相容的, 根据定理 1.20 可知其通解为:

$$\hat{\beta} = (X^T X)^- X^T y$$

其中 $(X^T X)^-$ 是 $X^T X$ 的任意一个广义逆矩阵。

对任意的 β , 有:

$$\begin{aligned} Q(\beta) &= \|y - X\beta\|^2 = \|y - X\hat{\beta} + X\hat{\beta} - X\beta\|^2 = \|y - X\hat{\beta} + X(\hat{\beta} - \beta)\|^2 \\ &= \|y - X\hat{\beta}\|^2 + \|X(\hat{\beta} - \beta)\|^2 + 2(y - X\hat{\beta})^T X(\hat{\beta} - \beta) \end{aligned}$$

注意到正则方程即为:

$$X^T(y - X\beta) = \mathbf{0}$$

于是:

$$2(y - X\hat{\beta})^T X(\hat{\beta} - \beta) = 2[X^T(y - X\hat{\beta})]^T(\hat{\beta} - \beta) = 0$$

所以:

$$Q(\beta) = \|y - X\hat{\beta}\|^2 + \|X(\hat{\beta} - \beta)\|^2$$

上第二项总是非负的, 由范数的性质其为 0 当且仅当 $X\hat{\beta} = X\beta$, 即当且仅当 $X^T X\beta = X^T X\hat{\beta} = X^T y$, 所以使 $Q(\beta)$ 达到最小值的 β 必为正则方程的解 $\hat{\beta} = (X^T X)^{-1} X^T y$. \square

推导 2.1. 若 $\text{rank}(X) = p$, 则 X 的列向量组线性无关。考虑二次型 $y^T X^T X y$, $y^T X^T X y = 0 \Leftrightarrow \|Xy\| = 0 \Leftrightarrow Xy = \mathbf{0}$, 而 X 的列向量是线性无关的, 所以不存在非零向量的 y 使得 $Xy = \mathbf{0}$, 于是 $y^T X^T X y$ 是一个正定二次型, $X^T X$ 是一个正定矩阵。由定理 1.42(3) 的第五点和可得 $X^T X$ 可逆。此时 $\hat{\beta} = (X^T X)^{-1} X^T y$, 称 $\hat{\beta}$ 为 β 的最小二乘估计 (least squares estimate)。

行列式等于特征值的积, 行列式大于 0 矩阵可逆

Definition 2.3. 若存在 $n \times 1$ 向量 α 使得 $E(\alpha^T y) = c^T \beta$ 对一切的 β 成立, 则称 $c^T \beta$ 为可估函数 (estimable function)。

Property 2.0.1. 对于定义 2.1, $c^T \beta$ 和 $d^T \beta$ 是可估函数, $\hat{\beta}$ 是正则方程的解, 则:

1. 使 $c^T \beta$ 成为可估函数的全体向量 c 构成 $\mathcal{M}(X^T)$;
2. 若 $c_1^T \beta$ 和 $c_2^T \beta$ 都是可估函数, 则对任意常数 a_1, a_2 , $a_1 c_1^T \beta + a_2 c_2^T \beta$ 也是可估函数;
3. 线性无关的可估函数组最多有 $\text{rank}(X)$ 个可估函数;
4. $c^T \hat{\beta}$ 与 $(X^T X)^{-}$ 的选择无关;
5. $c^T \hat{\beta}$ 为 $c^T \beta$ 的无偏估计;
6. $\text{Var}(c^T \hat{\beta}) = \sigma^2 c^T (X^T X)^{-} c$, $\text{Cov}(c^T \hat{\beta}, d^T \hat{\beta}) = \sigma^2 c^T (X^T X)^{-} d$, 且与 $(X^T X)^{-}$ 的选择无关;
7. $c^T \hat{\beta}$ 是 $c^T \beta$ 唯一的 BLUE;
8. 设 $\varphi_i = c_i^T \beta$, $i = 1, 2, \dots, k$ 都是可估函数, $\alpha_1, \alpha_2, \dots, \alpha_k \in \mathbb{R}$, 则 $\varphi = \sum_{i=1}^k \alpha_i \varphi_i$ 也是可估的, 且 $\hat{\varphi} = \sum_{i=1}^k \alpha_i c_i^T \hat{\beta}$ 是 φ 的 BLU 估计。

Proof. (1) $c^T \beta$ 是可估函数 \Leftrightarrow 存在 $n \times 1$ 向量 α 使得 $E(\alpha^T y) = \alpha^T E(y) = \alpha^T X\beta = c^T \beta$ 对一切的 β 成立 $\Leftrightarrow c = X^T \alpha$ 。

(2) 由 (1) 直接可得。

(3) 由 (1) 和直接可得。

转置秩不变

(4) 因为 $c^T\beta$ 可估, 由 (1) 可知存在 $n \times 1$ 向量 α 使得 $c = X^T\alpha$, 于是:

$$c^T\hat{\beta} = \alpha^T X(X^T X)^{-1} X^T y$$

由性质 1.5.1(4) 即可得出结论。

(5) 因为 $c^T\beta$ 可估, 由 (1) 可知存在 $n \times 1$ 向量 α 使得 $c = X^T\alpha$, 根据性质 1.5.1(5) 可得:

$$E(c^T\hat{\beta}) = E[\alpha^T X(X^T X)^{-1} X^T y] = \alpha^T X(X^T X)^{-1} X^T X\beta = \alpha^T X\beta = c^T\beta$$

(6) 因为 $c^T\beta, d^T\beta$ 是可估函数, 所以存在 α, γ 使得 $c = X^T\alpha, d = X^T\gamma$ 。由 ??(3) 和性质 1.5.1(6)(5) 可知:

$$\begin{aligned} \text{Cov}(c^T\hat{\beta}, d^T\hat{\beta}) &= \text{Cov}[\alpha^T X(X^T X)^{-1} X^T y, \gamma^T X(X^T X)^{-1} X^T y] \\ &= \alpha^T X(X^T X)^{-1} X^T \text{Cov}(y) X[(X^T X)^{-1}]^T X^T \gamma \\ &= \alpha^T X(X^T X)^{-1} X^T \sigma^2 I_n X(X^T X)^{-1} X^T \gamma \\ &= \sigma^2 \alpha^T X(X^T X)^{-1} d \\ &= \sigma^2 c^T (X^T X)^{-1} d \end{aligned}$$

由性质 1.5.1(4) 及上第三行可知 $\text{Cov}(c^T\hat{\beta}, d^T\hat{\beta})$ 与 $(X^T X)^{-1}$ 的选择无关。

(7) 无偏性由 (5) 可得, 线性性由正则方程可知, 下证方差最小。设 $a^T y$ 为 $c^T\beta$ 的任一无偏估计, 由 (1) 的过程可知 $c = X^T a$ 。根据性质 1.5.2(3) 和 (6) 可得:

$$\begin{aligned} \text{Var}(a^T y) - \text{Var}(c^T\hat{\beta}) &= \sigma^2 [a^T a - c^T (X^T X)^{-1} c] \\ &= \sigma^2 [a^T - c^T (X^T X)^{-1} X^T] [a - X(X^T X)^{-1} c] \\ &= \sigma^2 \|a - X(X^T X)^{-1} c\|^2 \geq 0 \end{aligned}$$

上式第一行到第二行是由于性质 1.5.2(3):

$$\begin{aligned} &[a^T - c^T (X^T X)^{-1} X^T] [a - X(X^T X)^{-1} c] \\ &= a^T a - a^T X(X^T X)^{-1} c - c^T (X^T X)^{-1} X^T a + c^T (X^T X)^{-1} X^T X(X^T X)^{-1} c \\ &= a^T a - c^T (X^T X)^{-1} c - a^T X(X^T X)^{-1} c + a^T X(X^T X)^{-1} X^T X(X^T X)^{-1} c \\ &= a^T a - c^T (X^T X)^{-1} c - a^T X(X^T X)^{-1} c + a^T X(X^T X)^{-1} c \\ &= a^T a - c^T (X^T X)^{-1} c \end{aligned}$$

由范数的性质可知 $\text{Var}(a^T y) = \text{Var}(c^T\hat{\beta})$ 当且仅当 $a = X(X^T X)^{-1} c$, 由性质 1.5.2(3) 可知 $a = X(X^T X)^{-1} c \Leftrightarrow a^T = c^T (X^T X)^{-1} X^T \Leftrightarrow a^T y = c^T (X^T X)^{-1} X^T y = c^T\hat{\beta}$ 。

(8) 因为 $\varphi_1, \varphi_2, \dots, \varphi_k$ 都是可估函数, 所以存在 b_1, b_2, \dots, b_k 使得 $E(b_i^T y) = c_i^T\beta$, 于是:

$$E\left(\sum_{i=1}^k \alpha_i b_i^T y\right) = \sum_{i=1}^k \alpha_i E(b_i^T y) = \sum_{i=1}^k \alpha_i c_i^T \beta = \sum_{i=1}^k \alpha_i \varphi_i = \varphi$$

所以取 $\alpha = \sum_{i=1}^k \alpha_i b_i$ 即可得到 $E(\alpha^T y) = \varphi$, φ 是可估的。

由 (5) 可得 $c_i^T \hat{\beta}$ 是 $c_i^T \beta$ 的无偏估计, 所以:

$$E(\hat{\varphi}) = E\left(\sum_{i=1}^k \alpha_i c_i^T \hat{\beta}\right) = \sum_{i=1}^k \alpha_i E(c_i^T \hat{\beta}) = \sum_{i=1}^k \alpha_i c_i^T \beta = \varphi$$

即 $\hat{\varphi}$ 是一个无偏估计。

令 $c = \sum_{i=1}^k \alpha_i c_i$, 则 $\varphi = c^T \beta$ 。设 $\gamma^T y$ 是 φ 的一个无偏估计, 于是由 (7) 可得:

$$\text{Var}(\gamma^T y) - \text{Var}(c^T \hat{\beta}) = \sigma^2 \|\gamma - X(X^T X)^{-1} c\|^2$$

上式等于 0 $\Leftrightarrow \gamma^T y = c^T \hat{\beta} = \hat{\varphi}$, 即 $\hat{\varphi}$ 是唯一的 BLUE。 \square

Definition 2.4. 对于定义 2.1, 若 $c^T \beta$ 是可估函数, 称 $c^T \hat{\beta}$ 为 $c^T \beta$ 的 LS 估计, 其中 $\hat{\beta}$ 为正则方程的解。

Definition 2.5. 称 $\hat{e} = y - X\hat{\beta}$ 为残差向量。

Property 2.0.2. 对于定义 2.1, $\hat{\beta}$ 为正则方程的解, 则残差向量 \hat{e} 满足 $E(\hat{e}) = 0$, $\text{Cov}(\hat{e}) = \sigma^2(I - P_X)$ 。

Proof. 由可知:

$$\begin{aligned} E(\hat{e}) &= E(y - X\hat{\beta}) = E[I_n y - X(X^T X)^{-1} X^T y] = (I_n - P_X) E(y) \\ &= (I_n - P_X) X\beta = (X - X)\beta = 0 \end{aligned}$$

$$\text{Cov}(\hat{e}) = \text{Cov}[(I_n - P_X)y] = (I_n - P_X) \text{Cov}(y)(I_n - P_X) = \sigma^2(I_n - P_X)$$

\square

对称幂等阵