

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	5
Hermitian 转置不改变秩	5
数域问题	15
Schmidt 正交化链接	30
可逆矩阵行列式链接	36
可逆矩阵行列式链接	37
有空证明	39
存在性证明	42
K^n 子空间为闭集	42
期望的线性性	48
Jensen 不等式链接	52
期望的线性性质，Lebesgue 积分	53
链接 Lebesgue 积分性质	55
链接独立性条件	56
需要证明对小于的都存在	57
写完变量变换法链接过来	60
期望的性质	61
写完矩阵的秩做链接	61
写完变量变换法链接过来	61
期望的性质	64
需要补充证明，但这里涉及到了 Jordan 标准形，学完再来补。	70
链接独立方差等于和	73
行列式等于特征值的积，行列式大于 0 矩阵可逆	77
$A(X^T X)^{-1} A^T$ 的可逆性	81
$A(X^T X)^{-1} A^T$ 的正定性	87

目录

第一部分 代数	1
第一章 矩阵	2
1.1 矩阵空间	2
1.1.1 矩阵的运算	2
1.1.2 矩阵的行列式	4
1.1.3 矩阵的秩	4
1.2 矩阵的向量空间	5
1.3 线性方程组	5
1.3.1 初等方法	7
1.3.2 秩与子空间	7
1.3.3 解的结构	8
1.4 矩阵的等价关系	9
1.4.1 相抵	9
1.4.2 相似	10
1.4.3 合同	10
1.5 相抵的应用	13
1.5.1 广义逆	13
1.5.2 Moore-Penrose 广义逆	16
1.5.3 线性方程组的解	20
1.6 相似的应用	22
1.6.1 特征值与特征向量	22
1.6.2 矩阵的对角化	26
1.6.3 Hermitian 矩阵的对角化	27
1.7 合同的应用——二次型	31
1.7.1 二次型的规范形	32
1.7.2 正定二次型与正定矩阵	34
1.8 特殊矩阵	39

1.8.1	幂等阵	39
1.8.2	正交投影阵	41
1.9	矩阵的分解	43
1.9.1	SVD 分解	43
第二章	随机变量的数字特征	44
2.1	期望	44
2.2	方差	44
2.3	矩	45
2.3.1	原点矩	45
2.3.2	中心矩	46
2.4	协方差	46
2.5	二次型	48
2.6	矩母函数	52
2.7	累积量生成函数	53
2.8	特征函数	54
2.9	Fisher 信息量	58
第三章	正态分布与三大抽样分布	60
3.1	多元正态分布	60
3.1.1	多元正态分布的定义	60
3.1.2	多元正态分布的性质	62
3.1.3	矩阵正态分布的定义	69
3.1.4	矩阵正态分布的性质	71
3.2	χ^2 分布, t 分布和 F 分布	72
3.2.1	χ^2 分布	72
3.2.2	t 分布	74
3.2.3	F 分布	74
第四章	线性模型	76
4.1	一般线性模型	76
4.1.1	参数估计	77
4.1.2	约束最小二乘估计	80
4.1.3	实际计算	83
4.2	正态线性模型	83
4.2.1	参数估计	83
4.2.2	假设检验	86
4.2.3	置信域	88
4.3	误差协方差推广	89

4.3.1	广义最小二乘估计	89
4.3.2	最小二乘统一理论	89
4.4	似然比检验	90
附录		91
4.5	数域	91
4.6	等价关系	91
4.7	矩阵	93
4.7.1	Kronecker 乘积	93
4.7.2	迹	94
4.7.3	向量化算子	95
4.7.4	平方根阵	96

第一部分

代数

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

1.1.3 矩阵的秩

Definition 1.6. 矩阵 A 的列向量组的秩称为 A 的列秩, 行向量组的秩称为 A 的行秩

Lemma 1.1. 阶梯形矩阵 J 的行秩与等于列秩且都等于非零行数, J 的主元所在的行构成行向量组的一个极大线性无关组, 主元所在列构成列向量组的一个极大线性无关组。

Lemma 1.2. 矩阵的初等行变换不改变行秩, 初等列变换不改变列秩。

Proof. 证明三种变换前后的向量组是等价的, 由??(3) 即可得出结论。列变换的情况可由转置与行变换的结论得到。□

Lemma 1.3. 矩阵的初等行变换不改变矩阵列向量组之间的线性相关性:

1. 设矩阵 A 经过初等行变换变成矩阵 B , 则 A 的列向量组线性相关当且仅当 B 的列向量组线性相关;
2. 设矩阵 A 经过初等行变换变成矩阵 B , 若 B 的第 j_1, j_2, \dots, j_r 列构成 B 的列向量组的一个极大线性无关组, 则 A 的第 j_1, j_2, \dots, j_r 也构成 A 的一个极大线性无关组¹;
3. 初等行变换不改变列秩。

Proof. (1) 将矩阵 A, B 看作齐次线性方程组的矩阵, 则 $Ax = \mathbf{0}$ 和 $Bx = \mathbf{0}$ 同解, 于是 $Ax = \mathbf{0}$ 有非零解当且仅当 $Bx = \mathbf{0}$ 有非零解, 即 A 的列向量组线性相关当且仅当 B 的列向量组线性相关。

(2) A 的第 j_1, j_2, \dots, j_r 列经过初等行变换构成 B 的第 j_1, j_2, \dots, j_r 列, 由 (1) 可知它们线性无关。任取其它列第 l 列, 则 A 的第 j_1, j_2, \dots, j_r, l 列经过初等行变换构成 B 的第 j_1, j_2, \dots, j_r, l 列, 因为 B 的第 j_1, j_2, \dots, j_r 列构成 B 的列向量组的一个极大线性无关组, 所以 B 的第 j_1, j_2, \dots, j_r, l 列线性相关, 由 (1) 可知 A 的第 j_1, j_2, \dots, j_r, l 列也线性相关, 所以 A 的第 j_1, j_2, \dots, j_r, l 列构成 A 的一个极大线性无关组。

(3) 由 (2) 直接得到。□

Theorem 1.1. 任意矩阵的行秩都等于列秩。

Proof. 任取矩阵 A , 记 A 的阶梯形矩阵为 J 。由引理 1.2 可知则 A 的行秩等于 J 的行秩, 由引理 1.1 可知 J 的行秩等于 J 的列秩, 由引理 1.3(3) 可知 J 的列秩等于 A 的列秩, 于是 A 的行秩等于 A 的列秩。由 A 的任意性, 结论成立。□

Definition 1.7. 矩阵 A 的行秩和列秩统称为矩阵 A 的秩, 记为 $\text{rank}(A)$ 。

Corollary 1.1. 矩阵的初等变换不改变矩阵的秩。

Proof. 由引理 1.2 和定理 1.1 立即得到。□

¹与引理 1.1 联合起来提供了求矩阵列向量组的极大线性无关组的方法。

1.2 矩阵的向量空间

Definition 1.8. 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 1.2. 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 1.3. 设 $A \in M_{m \times n}(\mathbb{C})$, 则有:

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

Proof. 由性质 1.3.1(3) 可知只需证明方程 $A^H Ax = \mathbf{0}$ 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 $Ax = \mathbf{0}$ 。于是:

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

所以:

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

同理可得:

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

Hermitian
转置不改变秩

于是有:

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

□

1.3 线性方程组

Definition 1.9. 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 1.10. 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 1.11. 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 1.4. 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 1.12. 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*)。

1.3.1 初等方法

Theorem 1.5. 数域 K 上的 n 元线性方程组的解的情况只有三种可能：

1. 无解：增广矩阵化成的阶梯形方程出现 $0 = d$ 且 $d \neq 0$ ；
2. 有解：
 - (a) 唯一解：阶梯形矩阵的非零行数 r 等于未知量个数 n ；
 - (b) 无穷多解：阶梯形矩阵的非零行数 r 小于未知量个数 n ；

这导致：

1. 数域 K 上 n 元齐次线性方程组有非零解的充分必要条件为：系数矩阵经过初等行变换化成的阶梯形矩阵中非零行数 $r < n$ ；
2. 数域 K 上 n 元齐次线性方程组的方程数 m 若小于未知量数 n ，则一定有非零解。

1.3.2 秩与子空间

Theorem 1.6. 数域 K 上 n 元线性方程组 $Ax = b$ (即 $\sum_{i=1}^n \alpha_i x_i = b$, 其中 α_i 为 A 的列向量) 有解的充分必要条件为：

1. $b \in \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$ ；
2. $\text{rank}(A) = \text{rank}([A|b])$ ；

进一步可得唯一解与无穷多解的判别方法：

1. 唯一解： $\text{rank}(A) = n$ ；
2. 无穷多解： $\text{rank}(A) < n$ 。

这导致齐次线性方程组有非零解的充分必要条件为 $\text{rank}(A) < n$ 。

Proof. (1) 显然。

(2) 由??(4) 可得 $Ax = b$ 有解 $\Leftrightarrow b \in \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle \Leftrightarrow \langle \alpha_1, \alpha_2, \dots, \alpha_n, \beta \rangle = \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle \Leftrightarrow \dim \langle \alpha_1, \alpha_2, \dots, \alpha_n, \beta \rangle = \dim \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle \Leftrightarrow \text{rank}(A) = \text{rank}([A|b])$ 。

(3) 若 $\text{rank}(A) = n$ ，则阶梯形矩阵的非零行数 $r = n$ ，由定理 1.5 可得此时有唯一解。

(4) 与 (3) 类似。 □

1.3.3 解的结构

齐次线性方程组

Property 1.3.1. 数域 K 上 n 元齐次线性方程组 $Ax = \mathbf{0}$ 的解具有如下性质:

1. 若 α, β 是解, 对任意的 $c_1, c_2 \in K$, $k_1\alpha + k_2\beta$ 也是解;
2. 解空间 W 构成 K^n 的一个子空间;
3. 解空间 W 满足 $\dim(W) = n - \text{rank}(A)$ 。

Proof. (1) $A(k_1\alpha + k_2\beta) = k_1A\alpha + k_2A\beta = \mathbf{0}$ 。

(2) 由 (1) 立即可得。

(3) 设 A 的列向量组为 $\alpha_1, \alpha_2, \dots, \alpha_n$, A 的行数为 m 。定义线性映射 $\mathcal{T}: \alpha \rightarrow A\alpha$, 则 \mathcal{T} 是 K^n 到 \mathbb{K}^m 的一个线性映射。于是有:

$$\text{Ker}(\mathcal{T}) = \{\alpha \in K^n : \mathcal{T}\alpha = \mathbf{0}\} = \{\alpha \in K^n : A\alpha = \mathbf{0}\} = W$$

$$\text{Im}(\mathcal{T}) = \{A\alpha : \alpha \in K^n\} = \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$$

所以由??(4) 可得:

$$\dim(\text{Ker } \mathcal{T}) = \dim(W)$$

$$\text{rank}(A) = \text{rank}\{\alpha_1, \alpha_2, \dots, \alpha_n\} = \dim \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle = \dim(\text{Im } \mathcal{T})$$

由??(10) 即可得到:

$$\dim(K^n) = \dim(\text{Ker } \mathcal{T}) + \dim(\text{Im } \mathcal{T}) = \dim(W) + \text{rank}(A)$$

即 $n = \dim(W) + \text{rank}(A)$ 。 □

Definition 1.13. 设数域 K 上 n 元齐次线性方程组 $Ax = \mathbf{0}$ 有非零解, 称它的解空间 W 的一组基为基础解系 (*fundamental solution set*)。

非齐次线性方程组

Property 1.3.2. 数域 K 上 n 元非齐次线性方程组 $Ax = b$ 的解具有如下性质:

1. 若 α, β 是解, 则 $\alpha - \beta$ 为 $Ax = \mathbf{0}$ 的解;
2. 设 W 为 $Ax = \mathbf{0}$ 的解空间, 若 α 是 $Ax = b$ 的解, 则对任意的 $\beta \in W$, $\alpha + \beta$ 也是 $Ax = b$ 的解;
3. 设 W 为 $Ax = \mathbf{0}$ 的解空间, 则 $Ax = b$ 的解集 U 可以表示为:

$$U = \{\alpha + \beta : \beta \in W\}$$

其中 α 为 $Ax = b$ 的任意一个解;

4. $Ax = b$ 的解唯一当且仅当 $Ax = \mathbf{0}$ 的解空间为零空间。

Proof. (1) $A(\alpha - \beta) = A\alpha - A\beta = b - b = \mathbf{0}$ 。

(2) $A(\alpha + \beta) = A\alpha + A\beta = b + \mathbf{0} = b$ 。

(3) 由 (1)(2) 可得。

(4) 由 (3) 立即可得。 □

1.4 矩阵的等价关系

1.4.1 相抵

Definition 1.14. $A, B \in M_{s \times m}(K)$, 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 1.7. 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 1.8. 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 1.9 (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 1.15. $A, B \in M_n(K)$ 。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 1.10. 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。

$$(3) \text{ 由性质 4.7.2(3) 可得 } \operatorname{tr}(A) = \operatorname{tr}(P^{-1}BP) = \operatorname{tr}(BPP^{-1}) = \operatorname{tr}(B)。$$

(4)(5) 参考定理 1.25。 □

1.4.3 合同

Definition 1.16. $A, B \in M_n(K)$ 。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 1.11. 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 1.17. 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 1.4. $A, B \in M_n(K)$, 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 1.12. 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:

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

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 1.4 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 1.4 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 1.13. 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 1.14. 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 1.12 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 1.4 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 1.15. 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 1.12 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 1.4 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 1.18. 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 1.16. 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 $[(A)^-]^T = (A)^-$;

7. 若存在 α 使得 $c = A^T \alpha$, 则 $c^T (A^T A)^- A^T A = c^T$.

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:

$$C^T A^- B = D_2^T A A^- A D_1 = D_2^T A D_1$$

(4) 由定理 1.2 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:

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

(7) 由 (5) 可得:

$$c^T (A^T A)^- A^T A = \alpha^T A (A^T A)^- A^T A = \alpha^T A = c^T \quad \square$$

1.5.2 Moore-Penrose 广义逆

Definition 1.19. 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 1.17. 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:

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

Proof. 由定理 1.3 可知 $(B^H B)^{-1}, (CC^H)^{-1}$ 存在, 将上述 X 代入 A 的 Penrose 方程组可得:

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

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

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

其中 P, Q, Λ 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:

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

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。 □

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:

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

其中 Λ 为 A 的非零特征值构成的对角矩阵, Q 是一个正交矩阵;

6. 若 α 是一个非零向量, 则 $\alpha^+ = \frac{\alpha^H}{\|\alpha\|^2}$;

$$7. I - A^+A \geq 0;$$

$$8. (A^HA)^+ = A^+(A^H)^+;$$

$$9. A^+ = (A^HA)^+A^H = A^H(AA^H)^+.$$

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

$$\begin{aligned} X_1 &= X_1AX_1 = X_1(AX_2A)X_1 = X_1(AX_2)(AX_1) \\ &= X_1(AX_2)^H(AX_1)^H = X_1(AX_1AX_2)^H = X_1X_2^H(AX_1A)^H \\ &= X_1X_2^HA^H = X_1(AX_2)^H = X_1AX_2 = X_1(AX_2A)X_2 \\ &= (X_1A)(X_2A)X_2 = (X_1A)^H(X_2A)^HX_2 = (X_2AX_1A)^HX_2 \\ &= (X_2A)^HX_2 = X_2AX_2 = X_2 \end{aligned}$$

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 1.18) 可得:

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

将其代入 A^H 的 Penrose 方程组可得:

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

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 1.18) 显然可得 $\text{rank}(A^+) = \text{rank}(\Lambda)$, 而 $\text{rank}(\Lambda) = \text{rank}(A)$, 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 Q 使得:

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

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

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

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

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

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

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

(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 1.18) 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 1.45(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 1.49) 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 1.18) 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 1.19. 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 1.20. 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 $x = (I_n - A^-A)y$ 。 \square

非齐次方程组解的结构

Theorem 1.21 (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 1.19 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 1.20 可知齐次线性方程组 $Ax = \mathbf{0}$ 的通解为 $(I_n - A^-A)y$, 由性质 1.3.2(3) 可得 $Ax = \beta$ 的通解为 $x = A^-\beta + (I_n - A^-A)y$ 。 \square

Theorem 1.22 (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 1.19 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 1.21 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 $A^-\beta$ 。 \square

Theorem 1.23. 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 1.21 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 1.20. $A \in M_n(K)$ 。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 1.21. $A \in M_n(K)$, 称 $|\lambda I - A|$ 为 A 的特征多项式 (characteristic polynomial)。

Theorem 1.24. $A \in M_n(K)$, 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 1.25. 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 1.22. $A \in M_n(K)$, 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 1.26. $A \in M_n(K)$, 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 1.25 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。 \square

1.6.2 矩阵的对角化

Definition 1.23. 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 1.27 (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 1.28 (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 1.27 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 1.26 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 1.27 的必要性可得 A 不可以对角化。 \square

Corollary 1.2. $A \in M_n(K)$ 如果有 n 个不同的特征值, 那么 A 可对角化。

Theorem 1.29 (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 1.28 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 1.24. 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 1.30. 正交相似是 $M_n(\mathbb{C})$ 上的一个等价关系。

Theorem 1.31. $A \in M_n(\mathbb{C})$ 。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 1.3. 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 矩阵。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 $\lambda_2, \lambda_3, \dots, \lambda_n$ 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 1.25 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 1.32. 对于 Hermitian 矩阵 $A \in M_n(\mathbb{C})$, 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:

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

实对称矩阵特征值的极值性质

Theorem 1.33. 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 。 \square

1.7 合同的应用——二次型

Definition 1.25. 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 1.26. 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 1.27. 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 1.34. 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 1.35. 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 1.12 与定理 1.34 可立即得出结论. \square

Theorem 1.36. 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:

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

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 1.37. 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 $d_1 y_1^2 + d_2 y_2^2 + \dots + d_r y_r^2$, 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 1.13 可得 $\text{rank}(A) = r$. \square

Definition 1.28. 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 1.29. 实数域上的二次型称为实二次型。由定理 1.14 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 1.34 可知经过一个适当的非退化线性变换可以将 $x^T Ax$ 化作:

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

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, -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 1.38 (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:

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

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 1.30. 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 1.39. 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 1.38 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 1.38），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 1.40. 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 1.34 可得矩阵合同等价于各自对应的二次型等价，再由定理 1.39 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 1.31. 复数域上的二次型称为复二次型。由定理 1.15 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 1.34 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 1.41. 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 1.42. 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 1.41 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 1.41), 第二条是显然的。 \square

1.7.2 正定二次型与正定矩阵

Definition 1.32. 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 1.33. 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 1.43. 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 1.44. 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 1.34 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 1.34、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 1.34. 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 1.35. 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 1.45. 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 1.34 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 1.34、(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 1.36. 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 1.37. 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 1.46. 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 1.44(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

Definition 1.38. 若矩阵 $A \in M_n(K)$ 满足 $A^2 = A$, 则称 A 为幂等矩阵 (idempotent matrix)。

Theorem 1.47. $A^-A, AA^-, I - A^-A, I - AA^-$ 都是幂等阵。

Proof. 代入广义逆方程即可得出结论。 \square

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. A 幂等 $\Leftrightarrow \text{rank}(A) + \text{rank}(I_n - A) = n$;
4. 若 A 是 Hermitian 矩阵, 则存在秩为 r 的 $B \in M_n(K)$ 使得 $A = B(B^H B)^{-1} B^H$;
5. 若 A 是 Hermitian 矩阵, 则 $A^+ = A$ 。

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。

(2) 因为 $\text{rank}(A) = r$, 所以存在可逆矩阵 P, Q 使得:

$$A = P \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q = \begin{pmatrix} P_1 & P_2 \end{pmatrix} \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = P_1 Q_1$$

其中 P_1 为 $n \times r$ 矩阵, Q_1 为 $r \times n$ 矩阵, 于是 $A = P_1 Q_1$ 。因为 A 是一个幂等阵, 所以:

$$\begin{aligned} P \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q P \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q &= P \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q \\ \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q P \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} &= \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \\ \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} P_1 & P_2 \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} &= \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \\ \begin{pmatrix} Q_1 P_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} &= \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \end{aligned}$$

即 $Q_1 P_1 = I_r$ 。由性质 4.7.2(3) 可得:

$$\text{tr}(A) = \text{tr}(P_1 Q_1) = \text{tr}(Q_1 P_1) = \text{tr}(I_r) = r = \text{rank}(A)$$

(3) 必要性: 因为 A 是一个幂等阵, 所以 $(I_n - A)(I_n - A) = I_n - 2A + A^2 = I_n - A$, 即 $I_n - A$ 也是一个幂等阵。由性质 4.7.2(1) 和 (2) 可知:

$$n = \text{tr}(I_n) = \text{tr}(I_n - A + A) = \text{tr}(I_n - A) + \text{tr}(A) = \text{rank}(I_n - A) + \text{rank}(A)$$

充分性: 因为 $\text{rank}(A) = r$, 由性质 1.3.1(3) 可知存在齐次线性方程组 $Ax = \mathbf{0}$ 存在 $n - r$ 个线性无关的解, 它们是 A 的特征值 0 的特征向量。因为 $\text{rank}(I_n - A) = n - r$, 所以齐次线性方程组 $(I_n - A)x = \mathbf{0}$ 有 $n - r$ 个线性无关的解, 即 $Ax = x$ 有 $n - r$ 个线性无关的解, A 的特征值 1 有 $n - r$ 个线性无关的特征向量, 于是存在可逆矩阵 P 使得:

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

所以 $A^2 = A$, A 是一个幂等阵。

(4) 因为 A 是 Hermitian 幂等阵, 由性质 1.6.2(3) 与 (1)(2) 可得存在正交阵 $Q = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix}$ 使得:

$$A = Q \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^H = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} Q_1^H \\ Q_2^H \end{pmatrix} = Q_1 Q_1^H = Q_1 (Q_1^H Q_1)^{-1} Q_1^H$$

因为 Q_1 是正交矩阵 Q 的前 r 列构成的矩阵, 所以 Q_1 的列向量组线性无关, $\text{rank}(Q_1) = r$ 。于是取 $B = Q_1$ 即可。

(5) 因为 A 是一个 Hermitian 幂等阵, 由性质 1.5.2(5) 和 (1) 可得:

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

令 $X = A^+$, 将其代入关于 A^+ 的 Penrose 方程组可知相容。由性质 1.5.2(2)(1) 即可得 $A = A^+$ 。□

1.8.2 正交投影阵

Definition 1.39. 设 $x \in K^n$, E 是 K^n 的一个子空间。对 x 作分解:

$$x = y + z, y \in E, z \in E^\perp$$

则称 y 为 x 在 E 上的正交投影。若 P 为 n 阶方阵, 且对任意的 $x \in K^n$, 都有 $y = Px \in E$ 和 $x - y \in E^\perp$, 则称 P 为向 E 的正交投影矩阵 (*orthogonal projection matrix*)。

Property 1.8.2. 设 P_1, P_2 为两个正交投影阵, $A \in M_{m \times n}(K)$, P_A 是向 $\mathcal{M}(A)$ 的正交投影阵, 则:

1. $\text{rank}(P_A) = \text{rank}(A)$;
2. $P_A = A(A^T A)^- A^T$;
3. $P \in M_n(K)$ 为正交投影阵当且仅当 P 为对称幂等阵;
4. $I_m - P_A$ 是对称幂等阵;
5. $P \in M_n(K)$ 为正交投影阵当且仅当对任意的 $x \in K^n$ 有:

$$\|x - Px\| = \inf_{u \in \mathcal{M}(P)} \|x - u\|$$

6. $P = P_1 + P_2$ 为正交投影阵 $\Leftrightarrow P_1P_2 = P_2P_1 = \mathbf{0}$;

7. 当 $P_1P_2 = P_2P_1 = \mathbf{0}$ 时, $P = P_1 + P_2$ 为向 $\mathcal{M}(P_1) \oplus \mathcal{M}(P_2)$ 上的正交投影阵;

8. $P = P_1P_2$ 为正交投影阵 $\Leftrightarrow P_1P_2 = P_2P_1$;

9. 当 $P_1P_2 = P_2P_1$ 时, $P = P_1P_2$ 为向 $\mathcal{M}(P_1) \cap \mathcal{M}(P_2)$ 上的正交投影阵;

10. $P = P_1 - P_2$ 为正交投影阵 $\Leftrightarrow P_1P_2 = P_2P_1 = P_2$;

11. 当 $P_1P_2 = P_2P_1 = P_2$ 时, $P = P_1 - P_2$ 为向 $\mathcal{M}(P_1) \oplus \mathcal{M}(P_2)^\perp$ 上的正交投影阵。

Proof. (1) 因为对任意的 $x \in K^n$ 有 $P_Ax \in \mathcal{M}(A)$, 所以 $\mathcal{M}(P_A) \subseteq \mathcal{M}(A)$ 。对任意的 $y \in \mathcal{M}(A)$, 存在 $x \in K^n$ 有 $P_Ax = y$ (只要取 $x = y + z$, z 是 $\mathcal{M}(A)^\perp$ 中的一个向量即可), 于是 $\mathcal{M}(A) \subseteq \mathcal{M}(P_A)$, 所以 $\mathcal{M}(A) = \mathcal{M}(P_A)$ 。由??(4) 可得:

$$\text{rank}(P_A) = \dim[\mathcal{M}(P_A)] = \dim[\mathcal{M}(A)] = \text{rank}(A)$$

(2) 由 $\mathcal{M}(A)$ 的定义可知 $\mathcal{M}(A)$ 是 K^m 的子空间, 根据??和??(4) 可知存在 $\mathcal{M}(A)$ 的补空间, 且这个补空间的维数为 $m - \text{rank}(A)$ 。取补空间的一组正交于 A 的列向量组的基 $b_1, b_2, \dots, b_{m-\text{rank}(A)}$, 则该补空间等于由这组基构成列向量组的矩阵 B 的列空间, 即 $\mathcal{M}(B) = \mathcal{M}(A)^\perp$ 。于是对任意的 $x \in K^m$ 有:

$$x = \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = A\alpha + B\beta$$

$$P_Ax = P_AA\alpha + P_AB\beta = A\alpha, \quad \forall \alpha \in K^{\text{rank}(A)}, \beta \in K^{m-\text{rank}(A)}$$

所以:

$$\begin{cases} P_AA = A \\ P_AB = \mathbf{0} \end{cases}$$

由第二个方程可以得到 P_A^T 的每一列都与 B 的每一列正交, 于是 $\mathcal{M}(P_A^T) \subseteq \mathcal{M}(B)^\perp = \mathcal{M}(A)$, 于是存在矩阵 C 使得 $P_A^T = AC$ 。将该式代入第一个方程可得到 $C^T A^T A = A$, 即 $A^T AC = A^T$, 因为该方程是相容的, 由定理 1.22 可知 C 的通解为 $(A^T A)^- A^T$, 于是 $P_A^T = A(A^T A)^- A^T$ 。由性质 1.5.1(6) 可得 $P_A = A(A^T A)^- A^T$ 。

(3) 充分性: 由 (2) 和性质 1.8.1(4) 立即得出。

必要性: 设 P 是向 $\mathcal{M}(A)$ 的正交投影阵。对称性由 (2) 和性质 1.5.1(6) 得出。由性质 1.5.1(5) 可知:

$$P^2 = A(A^T A)^- A^T A(A^T A)^- A^T = A(A^T A)^- A^T$$

即 P 幂等, 于是 P 是一个对称幂等阵。

(4) 由 (3) 可知 P_A 是对称幂等阵, 显然 $I_m - P_A$ 也是对称幂等阵。

(5) 显然 $\mathcal{M}(B)$ 是一个凸闭集, 由??立即可得。

存在性证明

K^n 子空间
为闭集

□

1.9 矩阵的分解

1.9.1 SVD 分解

Theorem 1.48. 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 1.45(3) 的第五条可知 AA^H 是半正定矩阵。□

Theorem 1.49. 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 1.3 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:

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

□

Definition 1.40. 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

随机变量的数字特征

2.1 期望

Definition 2.1. 设 X 是一个随机变量，具有分布函数 $F(x)$ 。若 X 的积分有限，则称：

$$E(X) = \int_{-\infty}^{+\infty} x dF(x)$$

为 X 的数学期望 (*mathematical expectation*)。

2.2 方差

Property 2.2.1. 设 X, Y 是随机变量，则：

1. $\text{Var}(X) = E(X^2) - [E(X)]^2$;
2. $\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}[E(X|Y)]$;
3. $\text{Var}(X \pm Y) = \text{Var}(X) \pm \text{Cov}(X, Y) + \text{Var}(Y)$ ，若 X, Y 不相关，则有 $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$;

Proof. (1) 设 $E(X) = \mu$ ，由方差的定义：

$$\text{Var}(X) = E[(X - \mu)^2] = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$$

(2) 由 (1) 可得：

$$\begin{aligned} E[\text{Var}(X|Y)] &= E\{E(X^2|Y) - [E(X|Y)]^2\} \\ &= E[E(X^2|Y)] - E\{[E(X|Y)]^2\} \\ &= E(X^2) - E\{[E(X|Y)]^2\} \\ \text{Var}[E(X|Y)] &= E\{[E(X|Y)]^2\} - \{E[E(X|Y)]\}^2 \\ &= E\{[E(X|Y)]^2\} - [E(X)]^2 \end{aligned}$$

于是:

$$E[\text{Var}(X|Y)] + \text{Var}[E(X|Y)] = E(X^2) - [E(X)]^2 = \text{Var}(X)$$

(3) 由方差的定义可得:

$$\begin{aligned} \text{Var}(X \pm Y) &= E[X \pm Y - E(X \pm Y)]^2 \\ &= E\{[X - E(X) \pm [Y - E(Y)]]\}^2 \\ &= E\{[X - E(X)]^2 \pm 2[X - E(X)][Y - E(Y)] + [Y - E(Y)]^2\} \\ &= \text{Var}(X) \pm 2 \text{Cov}(X, Y) + \text{Var}(Y) \end{aligned}$$

□

2.3 矩

2.3.1 原点矩

Definition 2.2. 设 X 是一个随机变量, $n \in \mathbb{N}^+$ 。若数学期望:

$$\mu_n = E(X^n)$$

存在, 则称 μ_n 为 X 的 n 阶原点矩 (*raw moment*)。

Definition 2.3. 设 \mathbf{X} 是一个 n 维随机向量, $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{N}$ 。若数学期望:

$$\mu_{\alpha_1, \alpha_2, \dots, \alpha_n} = E\left(\prod_{i=1}^n \mathbf{X}_i^{\alpha_i}\right)$$

存在, 则称 $\mu_{\alpha_1, \alpha_2, \dots, \alpha_n}$ 为 \mathbf{X} 的阶数为 $(\alpha_1, \alpha_2, \dots, \alpha_n)$ 的原点矩。

Theorem 2.1. 设 X 是一个随机变量, $m \in \mathbb{N}^+$ 。若 X 的 m 阶原点矩 μ_m 存在, 则 X 具有所有不超过 m 阶的原点矩。

Proof. 取任意的 $n < m$ 且 $n \in \mathbb{N}^+$, 则显然:

$$|x^n| \leq \begin{cases} 1, & |x| \leq 1 \\ |x^m|, & |x| > 1 \end{cases}$$

于是 $|x^n| \leq 1 + |x^m|$ 。因为 X 有 m 阶中心矩, 所以:

$$\int_{-\infty}^{+\infty} |x^m| p(x) dx < +\infty$$

于是:

$$\int_{-\infty}^{+\infty} |x|^n p(x) dx \leq \int_{-\infty}^{+\infty} (|x|^m + 1) p(x) dx = \int_{-\infty}^{+\infty} |x|^m p(x) dx + 1 < +\infty$$

所以 X 具有 n 阶原点矩。由 n 的任意性, 结论成立。

□

2.3.2 中心矩

Definition 2.4. 设 X 是一个随机变量, $\mu_1 = E(X)$, $n \in \mathbb{N}^+$ 。若数学期望:

$$\nu_n = E[(X - \mu_1)^n]$$

存在, 则称 ν_n 为 X 的 n 阶中心矩 (*central moment*)。

Definition 2.5. 设 \mathbf{X} 是一个 n 维随机向量, $\mu = E(\mathbf{X})$, $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{N}$ 。若数学期望:

$$\nu_{\alpha_1, \alpha_2, \dots, \alpha_n} = E \left[\prod_{i=1}^n (\mathbf{X}_i - \mu_i)^{\alpha_i} \right]$$

存在, 则称 $\nu_{\alpha_1, \alpha_2, \dots, \alpha_n}$ 为 \mathbf{X} 的阶数为 $(\alpha_1, \alpha_2, \dots, \alpha_n)$ 的中心矩。

Theorem 2.2. 随机变量 X 的中心矩 ν_n 与原点矩 μ_n 之间存在如下关系:

$$\nu_n = \sum_{i=0}^n \binom{n}{i} \mu_i (-\mu_1)^{n-i}$$

Proof. 由中心矩的定义可得:

$$\nu_n = E[(X - \mu_1)^n] = E \left[\sum_{i=0}^n \binom{n}{i} X^i (-\mu_1)^{n-i} \right] = \sum_{i=0}^n \binom{n}{i} \mu_i (-\mu_1)^{n-i} \quad \square$$

2.4 协方差

Definition 2.6. 随机向量 \mathbf{X} 与随机向量 \mathbf{Y} 的协方差 (*covariance*) 矩阵定义为:

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = E \left[(\mathbf{X} - E(\mathbf{X})) (\mathbf{Y} - E(\mathbf{Y}))^T \right]$$

若 $\mathbf{X} = \mathbf{Y}$, 则可将 $\text{Cov}(\mathbf{X}, \mathbf{Y})$ 简写为 $\text{Cov}(\mathbf{X})$ 。

Definition 2.7. 设 X, Y 是两个随机变量, 则:

1. 若 $\text{Cov}(X, Y) > 0$, 称 X, Y 正相关 (*positively correlated*);
2. 若 $\text{Cov}(X, Y) < 0$, 称 X, Y 负相关 (*negatively correlated*);
3. 若 $\text{Cov}(X, Y) = 0$, 称 X, Y 。

Property 2.4.1. 协方差矩阵具有如下性质:

1. X 是一个 n 维随机向量, 则 $\text{tr}[\text{Cov}(\mathbf{X})] = \sum_{i=1}^n \text{Var}(\mathbf{X}_i)$;
2. X 是一个 n 维随机向量, 则 $\text{Cov}(\mathbf{X})$ 是半正定的对称矩阵;
3. 设 A 和 B 分别为 $p \times n$ 和 $q \times m$ 非随机矩阵, \mathbf{X} 和 \mathbf{Y} 分别为 n 维、 m 维随机向量, 则:

$$\text{Cov}(A\mathbf{X}, B\mathbf{Y}) = A \text{Cov}(\mathbf{X}, \mathbf{Y}) B^T$$

4. 若 \mathbf{X} 是一个常数项量, \mathbf{Y} 是一个随机向量, 则 $\text{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}$;

5. 设 $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ 为随机向量, 则:

$$\text{Cov}(\mathbf{X} + \mathbf{Y}, \mathbf{Z}) = \text{Cov}(\mathbf{X}, \mathbf{Z}) + \text{Cov}(\mathbf{Y}, \mathbf{Z})$$

$$\text{Cov}(\mathbf{X}, \mathbf{Y} + \mathbf{Z}) = \text{Cov}(\mathbf{X}, \mathbf{Y}) + \text{Cov}(\mathbf{X}, \mathbf{Z})$$

6. $\text{Cov}(\mathbf{X}) = \text{E}(\mathbf{X}\mathbf{X}^T) - [\text{E}(\mathbf{X})][\text{E}(\mathbf{X})]^T$ 。

Proof. (1) $\text{Cov}(\mathbf{X})$ 在 (i, i) 位置上的元素为:

$$\text{E}\left[\left(\mathbf{X}_i - \text{E}(\mathbf{X}_i)\right)\left(\mathbf{X}_i - \text{E}(\mathbf{X}_i)\right)^T\right] = \text{E}\left[\left(\mathbf{X}_i - \text{E}(\mathbf{X}_i)\right)^2\right] = \text{Var}(\mathbf{X}_i)$$

所以 $\text{tr}[\text{Cov}(\mathbf{X})] = \sum_{i=1}^n \text{Var}(\mathbf{X}_i)$ 。

(2) 因为:

$$\begin{aligned}\text{Cov}(\mathbf{X})_{(i,j)} &= \text{E}\left[\left(\mathbf{X}_i - \text{E}(\mathbf{X}_i)\right)\left(\mathbf{X}_j - \text{E}(\mathbf{X}_j)\right)^T\right] \\ &= \text{E}\left[\left(\mathbf{X}_j - \text{E}(\mathbf{X}_j)\right)\left(\mathbf{X}_i - \text{E}(\mathbf{X}_i)\right)^T\right] \\ &= \text{Cov}(\mathbf{X})_{(j,i)}\end{aligned}$$

所以 $\text{Cov}(\mathbf{X})$ 是一个对称矩阵。

取 n 维非随机向量 c , 设 $Y = c^T \mathbf{X}$ 则有:

$$\begin{aligned}\text{Var}(Y) &= \text{Var}(c^T \mathbf{X}) \\ &= \text{E}\left[\left(c^T \mathbf{X} - \text{E}(c^T \mathbf{X})\right)\left(c^T \mathbf{X} - \text{E}(c^T \mathbf{X})\right)\right] \\ &= \text{E}\left[\left(c^T \mathbf{X} - c^T \text{E}(\mathbf{X})\right)\left(c^T \mathbf{X} - c^T \text{E}(\mathbf{X})\right)^T\right] \\ &= \text{E}\left\{c^T \left(\mathbf{X} - \text{E}(\mathbf{X})\right) \left[c^T \left(\mathbf{X} - \text{E}(\mathbf{X})\right)\right]^T\right\} \\ &= c^T \text{E}\left[\left(\mathbf{X} - \text{E}(\mathbf{X})\right)\left(\mathbf{X} - \text{E}(\mathbf{X})\right)^T\right] c \\ &= c^T \text{Cov}(\mathbf{X}) c \geq 0\end{aligned}$$

由 c 的任意性, $\text{Cov}(\mathbf{X})$ 是半正定的。

(3) 类似于 (2) 中的推导, 有:

$$\begin{aligned}\text{Cov}(A\mathbf{X}, B\mathbf{Y}) &= \text{E}\left[\left(A\mathbf{X} - \text{E}(A\mathbf{X})\right)\left(B\mathbf{Y} - \text{E}(B\mathbf{Y})\right)^T\right] \\ &= A \text{E}\left[\left(\mathbf{X} - \text{E}(\mathbf{X})\right)\left(\mathbf{Y} - \text{E}(\mathbf{Y})\right)^T\right] B^T \\ &= A \text{Cov}(\mathbf{X}, \mathbf{Y}) B^T\end{aligned}$$

(4) 由协方差的定义直接可得;

(5) 由可得:

期望的线性
性

$$\begin{aligned}
 \text{Cov}(\mathbf{X} + \mathbf{Y}, \mathbf{Z}) &= E \left[\left(\mathbf{X} + \mathbf{Y} - E(\mathbf{X} + \mathbf{Y}) \right) \left(\mathbf{Z} - E(\mathbf{Z}) \right)^T \right] \\
 &= E \left[\left(\mathbf{X} + \mathbf{Y} - E(\mathbf{X}) - E(\mathbf{Y}) \right) \left(\mathbf{Z} - E(\mathbf{Z}) \right)^T \right] \\
 &= E \left[\left(\mathbf{X} - E(\mathbf{X}) \right) \left(\mathbf{Z} - E(\mathbf{Z}) \right)^T + \left(\mathbf{Y} - E(\mathbf{Y}) \right) \left(\mathbf{Z} - E(\mathbf{Z}) \right)^T \right] \\
 &= E \left[\left(\mathbf{X} - E(\mathbf{X}) \right) \left(\mathbf{Z} - E(\mathbf{Z}) \right)^T \right] + E \left[\left(\mathbf{Y} - E(\mathbf{Y}) \right) \left(\mathbf{Z} - E(\mathbf{Z}) \right)^T \right] \\
 &= \text{Cov}(\mathbf{X}, \mathbf{Z}) + \text{Cov}(\mathbf{Y}, \mathbf{Z})
 \end{aligned}$$

$$\begin{aligned}
 \text{Cov}(\mathbf{X}, \mathbf{Y} + \mathbf{Z}) &= E \left[\left(\mathbf{X} - E(\mathbf{X}) \right) \left(\mathbf{Y} + \mathbf{Z} - E(\mathbf{Y} + \mathbf{Z}) \right)^T \right] \\
 &= E \left[\left(\mathbf{X} - E(\mathbf{X}) \right) \left(\mathbf{Y} + \mathbf{Z} - E(\mathbf{Y}) - E(\mathbf{Z}) \right)^T \right] \\
 &= E \left[\left(\mathbf{X} - E(\mathbf{X}) \right) \left(\mathbf{Y} - E(\mathbf{Y}) \right)^T + \left(\mathbf{X} - E(\mathbf{X}) \right) \left(\mathbf{Z} - E(\mathbf{Z}) \right)^T \right] \\
 &= E \left[\left(\mathbf{X} - E(\mathbf{X}) \right) \left(\mathbf{Y} - E(\mathbf{Y}) \right)^T \right] + E \left[\left(\mathbf{X} - E(\mathbf{X}) \right) \left(\mathbf{Z} - E(\mathbf{Z}) \right)^T \right] \\
 &= \text{Cov}(\mathbf{X}, \mathbf{Y}) + \text{Cov}(\mathbf{X}, \mathbf{Z})
 \end{aligned}$$

(6) 显然:

$$\begin{aligned}
 \text{Cov}(\mathbf{X}) &= E \left[\left(\mathbf{X} - E(\mathbf{X}) \right) \left(\mathbf{X} - E(\mathbf{X}) \right)^T \right] = E \left[\left(\mathbf{X} - E(\mathbf{X}) \right) \mathbf{X}^T - \left(\mathbf{X} - E(\mathbf{X}) \right) E(\mathbf{X})^T \right] \\
 &= E \left[\left(\mathbf{X} - E(\mathbf{X}) \right) \mathbf{X}^T \right] - E[\mathbf{X} - E(\mathbf{X})] E(\mathbf{X})^T = E(\mathbf{X} \mathbf{X}^T) - [E(\mathbf{X})][E(\mathbf{X})]^T \quad \square
 \end{aligned}$$

2.5 二次型

Definition 2.8. \mathbf{X} 是一个 n 维随机向量, $A = (a_{ij})$ 为 n 阶非随机实对称阵, 则随机变量:

$$\mathbf{X}^T A \mathbf{X} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} \mathbf{X}_i \mathbf{X}_j$$

称为 \mathbf{X} 的二次型。

随机变量二次型的均值

Theorem 2.3. \mathbf{X} 是一个 n 维随机向量, $E(\mathbf{X}) = \mu$, $\text{Cov}(\mathbf{X}) = \Sigma$, 则:

$$E(\mathbf{X}^T A \mathbf{X}) = \mu^T A \mu + \text{tr}(A \Sigma)$$

Proof.

$$\begin{aligned}
E(\mathbf{X}^T A \mathbf{X}) &= E[(\mathbf{X} - \mu + \mu)^T A (\mathbf{X} - \mu + \mu)] \\
&= E[(\mathbf{X} - \mu)^T A (\mathbf{X} - \mu)] + E[(\mathbf{X} - \mu)^T A \mu] + E[\mu^T A (\mathbf{X} - \mu)] + E(\mu^T A \mu) \\
&= E\{\text{tr}[(\mathbf{X} - \mu)^T A (\mathbf{X} - \mu)]\} + \mu^T A \mu \\
&= E\{\text{tr}[A (\mathbf{X} - \mu)(\mathbf{X} - \mu)^T]\} + \mu^T A \mu \\
&= \text{tr} E[A (\mathbf{X} - \mu)(\mathbf{X} - \mu)^T] + \mu^T A \mu \\
&= \text{tr}\{A E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T]\} + \mu^T A \mu \\
&= \text{tr}(A \Sigma) + \mu^T A \mu
\end{aligned}$$

第二行到第三行利用到了 $E(\mathbf{X}) = \mu$ 以及 $(\mathbf{X} - \mu)^T A (\mathbf{X} - \mu) = \text{tr}[(\mathbf{X} - \mu)^T A (\mathbf{X} - \mu)]$, 后式成立是因为 $(\mathbf{X} - \mu)^T A (\mathbf{X} - \mu)$ 是一个标量, 标量的迹自然等于自身。第三行到第四行使用到了性质 4.7.2(3)。□

独立随机变量二次型的方差

Theorem 2.4. 设随机变量 $X_i, i = 1, 2, \dots, n$ 相互独立, $E(X_i) = \mu_i, \text{Var}(X_i) = \sigma^2, \nu_k^{(i)} = E[(X_i - \mu_i)^k], \mathbf{X} = (X_1, X_2, \dots, X_n)^T, \mu = (\mu_1, \mu_2, \dots, \mu_n)^T, A = (a_{ij})$ 为 n 阶非随机实对称阵, $a = (a_{11}, a_{22}, \dots, a_{nn})^T, b = (\nu_3^{(1)} a_{11}, \nu_3^{(2)} a_{22}, \dots, \nu_3^{(n)} a_{nn})^T$, 则:

$$\text{Var}(\mathbf{X}^T A \mathbf{X}) = \sum_{i=1}^n a_{ii}^2 \nu_4^{(i)} + \sigma^4 [2 \text{tr}(A^2) - 3a^T a] + 4\sigma^2 \mu^T A^2 \mu + 4\mu^T A b$$

Proof. 由性质 2.2.1(1) 可得:

$$\text{Var}(\mathbf{X}^T A \mathbf{X}) = E[(\mathbf{X}^T A \mathbf{X})^2] - [E(\mathbf{X}^T A \mathbf{X})]^2$$

由题设可知:

$$E(\mathbf{X}) = \mu, \text{Var}(\mathbf{X}) = \sigma^2 I$$

根据定理 2.3 可得:

$$\begin{aligned}
[E(\mathbf{X}^T A \mathbf{X})]^2 &= [\text{tr}(A \sigma^2 I) + \mu^T A \mu]^2 = [\sigma^2 \text{tr}(A) + \mu^T A \mu]^2 \\
&= \sigma^4 [\text{tr}(A)]^2 + 2\sigma^2 \text{tr}(A) \mu^T A \mu + (\mu^T A \mu)^2
\end{aligned}$$

同时:

$$\begin{aligned}
(\mathbf{X}^T A \mathbf{X})^2 &= [(\mathbf{X} - \mu + \mu)^T A (\mathbf{X} - \mu + \mu)]^2 \\
&= [(\mathbf{X} - \mu)^T A (\mathbf{X} - \mu) + 2\mu^T A (\mathbf{X} - \mu) + \mu^T A \mu]^2 \\
&= [(\mathbf{X} - \mu)^T A (\mathbf{X} - \mu)]^2 + 4[\mu^T A (\mathbf{X} - \mu)]^2 + (\mu^T A \mu)^2 \\
&\quad + 4(\mathbf{X} - \mu)^T A (\mathbf{X} - \mu) \mu^T A (\mathbf{X} - \mu) + 2(\mathbf{X} - \mu)^T A (\mathbf{X} - \mu) \mu^T A \mu \\
&\quad + 4\mu^T A (\mathbf{X} - \mu) \mu^T A \mu
\end{aligned}$$

令 $\mathbf{Y} = \mathbf{X} - \mu$, 则有 $E(\mathbf{Y}) = \mathbf{0}$, 再由定理 2.3 可得:

$$\begin{aligned} E[(\mathbf{X}^T A \mathbf{X})^2] &= E[(\mathbf{Y}^T A \mathbf{Y})^2] + 4E[(\mu^T A \mathbf{Y})^2] + (\mu^T A \mu)^2 \\ &\quad + 4E(\mathbf{Y}^T A \mathbf{Y} \mu^T A \mathbf{Y}) + 2\mu^T A \mu \sigma^2 \text{tr}(A) \end{aligned}$$

考虑:

$$\begin{aligned} E[(\mathbf{Y}^T A \mathbf{Y})^2] &= E\left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_{ij} a_{kl} \mathbf{Y}_i \mathbf{Y}_j \mathbf{Y}_k \mathbf{Y}_l\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_{ij} a_{kl} E(\mathbf{Y}_i \mathbf{Y}_j \mathbf{Y}_k \mathbf{Y}_l) \end{aligned}$$

作分类讨论:

1. i, j, k, l 互不相同, 则 $E(\mathbf{Y}_i \mathbf{Y}_j \mathbf{Y}_k \mathbf{Y}_l) = E(\mathbf{Y}_i)E(\mathbf{Y}_j)E(\mathbf{Y}_k)E(\mathbf{Y}_l) = 0$;
2. i, j, k, l 中存在某两个值相同:
 - 此时另外两个不同, 则 $E(\mathbf{Y}_i \mathbf{Y}_j \mathbf{Y}_k \mathbf{Y}_l) = 0$;
 - 此时另外两个也相同 (即 $i = j, k = l$ 或 $i = k, j = l$ 或 $i = l, j = k$), 则 $E(\mathbf{Y}_i \mathbf{Y}_j \mathbf{Y}_k \mathbf{Y}_l) = \sigma^4$ 。
3. i, j, k, l 中存在某三个值相同, 则 $E(\mathbf{Y}_i \mathbf{Y}_j \mathbf{Y}_k \mathbf{Y}_l) = 0$;
4. i, j, k, l 相同, 则 $E(\mathbf{Y}_i \mathbf{Y}_j \mathbf{Y}_k \mathbf{Y}_l) = \nu_4^{(i)}$ 。

于是:

$$\begin{aligned} E[(\mathbf{Y}^T A \mathbf{Y})^2] &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_{ij} a_{kl} E(\mathbf{Y}_i \mathbf{Y}_j \mathbf{Y}_k \mathbf{Y}_l) \\ &= \sum_{i=1}^n a_{ii}^2 \nu_4^{(i)} + \sigma^4 \left(\sum_{i \neq k} a_{ii} a_{kk} + \sum_{i \neq j} a_{ij}^2 + \sum_{i \neq j} a_{ij} a_{ji} \right) \\ &= \sum_{i=1}^n a_{ii}^2 \nu_4^{(i)} + \sigma^4 \left(\sum_{i \neq k} a_{ii} a_{kk} + 2 \sum_{i \neq j} a_{ij}^2 \right) \end{aligned}$$

因为:

$$\begin{aligned} \sum_{i \neq k} a_{ii} a_{kk} &= [\text{tr}(A)]^2 - a^T a \\ \sum_{i \neq j} a_{ij}^2 &= \text{tr}(A A^T) - a^T a = \text{tr}(A^2) - a^T a \end{aligned}$$

所以:

$$E[(\mathbf{Y}^T A \mathbf{Y})^2] = \sum_{i=1}^n a_{ii}^2 \nu_4^{(i)} + \sigma^4 \{ [\text{tr}(A)]^2 + 2 \text{tr}(A^2) - 3a^T a \}$$

由定理 2.3 和性质 4.7.2(3) 可得:

$$\begin{aligned} E[(\mu^T A Y)^2] &= E(\mu^T A Y \mu^T A Y) = E(Y^T A \mu \mu^T A Y) = \text{tr}(A \mu \mu^T A \sigma^2 I) \\ &= \sigma^2 \text{tr}(A \mu \mu^T A) = \sigma^2 \text{tr}(\mu^T A^2 \mu) = \sigma^2 \mu^T A^2 \mu \end{aligned}$$

注意到:

$$\begin{aligned} E(Y^T A Y \mu^T A Y) &= E\left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} Y_i Y_j \sum_{k=1}^n \sum_{l=1}^n a_{kl} \mu_k Y_l\right) \\ &= E\left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_{ij} a_{kl} \mu_k Y_i Y_j Y_l\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_{ij} a_{kl} \mu_k E(Y_i Y_j Y_l) \end{aligned}$$

和之前的讨论类似, 可以得到:

$$E(Y_i Y_j Y_l) = \begin{cases} \nu_3^{(i)}, & i = j = l \\ 0, & \text{其他情况} \end{cases}$$

于是有:

$$E(Y^T A Y \mu^T A Y) = \sum_{i=1}^n \sum_{k=1}^n a_{ii} \nu_3^{(i)} a_{ki} \mu_k$$

令 $b = (\nu_3^{(1)} a_{11}, \nu_3^{(2)} a_{22}, \dots, \nu_3^{(n)} a_{nn})^T$, 则:

$$E(Y^T A Y \mu^T A Y) = \sum_{i=1}^n \sum_{k=1}^n a_{ii} \nu_3^{(i)} a_{ki} \mu_k = \mu^T A b$$

将以上求得的期望值全部代入, 即可得到:

$$\begin{aligned} E[(X^T A X)^2] &= E[(Y^T A Y)^2] + 4 E[(\mu^T A Y)^2] + (\mu^T A \mu)^2 \\ &\quad + 4 E(Y^T A Y \mu^T A Y) + 2 \mu^T A \mu \sigma^2 \text{tr}(A) \\ &= \sum_{i=1}^n a_{ii}^2 \nu_4^{(i)} + \sigma^4 \{[\text{tr}(A)]^2 + 2 \text{tr}(A^2) - 3 a^T a\} \\ &\quad + 4 \sigma^2 \mu^T A^2 \mu + (\mu^T A \mu)^2 + 4 \mu^T A b + 2 \mu^T A \mu \sigma^2 \text{tr}(A) \end{aligned}$$

于是:

$$\begin{aligned} \text{Var}(X^T A X) &= E[(X^T A X)^2] - [E(X^T A X)]^2 \\ &= \sum_{i=1}^n a_{ii}^2 \nu_4^{(i)} + \sigma^4 \{[\text{tr}(A)]^2 + 2 \text{tr}(A^2) - 3 a^T a\} \\ &\quad + 4 \sigma^2 \mu^T A^2 \mu + (\mu^T A \mu)^2 + 4 \mu^T A b + 2 \mu^T A \mu \sigma^2 \text{tr}(A) \\ &\quad - \sigma^4 [\text{tr}(A)]^2 - 2 \sigma^2 \text{tr}(A) \mu^T A \mu - (\mu^T A \mu)^2 \\ &= \sum_{i=1}^n a_{ii}^2 \nu_4^{(i)} + \sigma^4 [2 \text{tr}(A^2) - 3 a^T a] + 4 \sigma^2 \mu^T A^2 \mu + 4 \mu^T A b \end{aligned} \quad \square$$

2.6 矩母函数

Definition 2.9. 设 X 是一个随机变量。称:

$$M_X(t) = E(e^{tX})$$

为 X 的矩母函数 (*moment-generating function, m.g.f.*), 其中 $t \in \mathbb{R}$ 。

Definition 2.10. 设 \mathbf{X} 是一个 n 维随机向量。称:

$$M_{\mathbf{X}}(t) = E(e^{t^T \mathbf{X}})$$

为 \mathbf{X} 的矩母函数, 其中 $t \in \mathbb{R}^n$ 。

Property 2.6.1. 设 \mathbf{X} 是一个 n 维随机向量, 则其矩母函数 $M_{\mathbf{X}}(t)$ 具有如下性质:

1. $M_{\mathbf{X}}(\mathbf{0}) = 1$;
2. $M_{\mathbf{X}}(t) \geq e^{t^T \mu}$, 其中 μ 是 \mathbf{X} 的均值向量;
3. 矩母函数与概率分布之间存在一个双射, 即 $M_{\mathbf{X}}(t) = M_{\mathbf{Y}}(t)$ 当且仅当 \mathbf{X} 与 \mathbf{Y} 具有相同的概率分布;
4. 设 m 维随机向量 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ 彼此独立, α_i 为常数, β_i 为 m 维常数向量, 则 $\mathbf{Y} = \sum_{i=1}^n (\alpha_i \mathbf{X}_i + \beta_i)$ 的特征函数为:

$$M_{\mathbf{Y}}(t) = \prod_{i=1}^n e^{t^T \beta_i} M_{\mathbf{X}_i}(\alpha_i t)$$

5. $M_X^{(n)}(0) = \mu_n$, 其中 X 是一个随机变量, μ_n 是 X 的 n 阶原点矩;
6. $M_{\mathbf{X}}(t)$ 有如下幂级数展开:

$$M_{\mathbf{X}}(t) = \sum_{(m_1, m_2, \dots, m_n) \in \mathbb{N}^n} \mu_{m_1, m_2, \dots, m_n} \prod_{i=1}^n \frac{t_i^{m_i}}{m_i!}$$

Proof. (1) $M_{\mathbf{X}}(\mathbf{0}) = E(e^0) = 1$ 。

Jensen 不等式链接

(2) 由 Jensen 不等式直接可得。

(3)

(4) 由矩母函数定义可得:

$$M_{\mathbf{Y}}(t) = E(e^{t^T \mathbf{Y}}) = E\left(\exp\left\{t^T \sum_{i=1}^n (\alpha_i \mathbf{X}_i + \beta_i)\right\}\right) = E\left(\prod_{i=1}^n e^{\alpha_i t^T \mathbf{X}_i}\right) \prod_{i=1}^n e^{t^T \beta_i}$$

因为 \mathbf{X}_i 互相独立, 所以 $\alpha_i \mathbf{X}_i$ 也相互独立, 于是有:

$$M_{\mathbf{Y}}(t) = E\left(\prod_{i=1}^n e^{\alpha_i t^T \mathbf{X}_i}\right) \prod_{i=1}^n e^{t^T \beta_i} = \prod_{i=1}^n E\left(e^{\alpha_i t^T \mathbf{X}_i}\right) \prod_{i=1}^n e^{t^T \beta_i} = \prod_{i=1}^n e^{t^T \beta_i} M_{\mathbf{X}_i}(\alpha_i t)$$

(5) 将 e^{tX} 展开为幂级数:

$$M_X(t) = E(e^{tX}) = E\left(\sum_{n=0}^{+\infty} \frac{t^n X^n}{n!}\right)$$

于是:

$$M_X^{(n)}(t) = E\left(X^n + \sum_{m=n+1}^{+\infty} \frac{t^m X^m}{m!}\right) = \mu_n + \sum_{m=1}^{+\infty} \frac{t^m}{m!} \mu_m$$

所以:

$$M_X^{(n)}(0) = E(X^n) = \mu_n$$

(6) 由可得:

$$\begin{aligned} M_{\mathbf{X}}(t) &= E(e^{t^T \mathbf{X}}) = E\left(\exp\left\{\sum_{i=1}^n t_i \mathbf{X}_i\right\}\right) = E\left[\sum_{m=0}^{+\infty} \frac{1}{m!} \left(\sum_{i=1}^n t_i \mathbf{X}_i\right)^m\right] \\ &= \sum_{m=0}^{+\infty} \frac{1}{m!} E\left[\left(\sum_{i=1}^n t_i \mathbf{X}_i\right)^m\right] = \sum_{m=0}^{+\infty} \frac{1}{m!} E\left(\sum_{\sum_{i=1}^n m_i=m} \frac{m!}{m_1!m_2!\cdots m_n!} \prod_{i=1}^n (t_i \mathbf{X}_i)^{m_i}\right) \\ &= \sum_{m=0}^{+\infty} \frac{1}{m!} \sum_{\sum_{i=1}^n m_i=m} \frac{m!}{m_1!m_2!\cdots m_n!} E\left[\prod_{i=1}^n (t_i \mathbf{X}_i)^{m_i}\right] \\ &= \sum_{m=0}^{+\infty} \sum_{\sum_{i=1}^n m_i=m} \frac{1}{m_1!m_2!\cdots m_n!} E\left(\prod_{i=1}^n \mathbf{X}_i^{m_i}\right) \prod_{i=1}^n t_i^{m_i} \\ &= \sum_{(m_1, m_2, \dots, m_n) \in \mathbb{N}^n} \mu_{m_1, m_2, \dots, m_n} \prod_{i=1}^n \frac{t_i^{m_i}}{m_i!} \end{aligned}$$

期望的线性性质,
Lebesgue
积分

□

2.7 累积量生成函数

Definition 2.11. 设 X 是一个随机变量。称 $K_X(t) = \log M_X(t)$ 为 X 的累积量生成函数 (cumulant-generating function, c.g.f.), 其中 $t \in \mathbb{R}$ 。

Definition 2.12. 设 \mathbf{X} 是一个 n 维随机向量。称 $K_{\mathbf{X}}(t) = \log M_{\mathbf{X}}(t)$ 为 \mathbf{X} 的累积量生成函数, 其中 $t \in \mathbb{R}^n$ 。

Definition 2.13. 设 \mathbf{X} 是一个 n 维随机向量。因为:

$$\begin{aligned} M_{\mathbf{X}}(t) &= \sum_{(m_1, m_2, \dots, m_n) \in \mathbb{N}^n} \frac{1}{m_1!m_2!\cdots m_n!} \prod_{i=1}^n t_i^{m_i} \mu_{m_1, m_2, \dots, m_n} \\ &= 1 + \sum_{\substack{(m_1, m_2, \dots, m_n) \in \mathbb{N}^n \\ (m_1, m_2, \dots, m_n) \neq \mathbf{0}}} \frac{1}{m_1!m_2!\cdots m_n!} \prod_{i=1}^n t_i^{m_i} \mu_{m_1, m_2, \dots, m_n} \end{aligned}$$

由对数函数的幂级数展开可得：

$$\begin{aligned} K_{\mathbf{X}}(t) &= \log \left(1 + \sum_{\substack{(m_1, m_2, \dots, m_n) \in \mathbb{N}^n \\ (m_1, m_2, \dots, m_n \neq \mathbf{0})}} \frac{1}{m_1! m_2! \dots m_n!} \prod_{i=1}^n t_i^{m_i} \mu_{m_1, m_2, \dots, m_n} \right) \\ &= \sum_{j=1}^{+\infty} (-1)^{j+1} \frac{1}{j} \left(\sum_{\substack{(m_1, m_2, \dots, m_n) \in \mathbb{N}^n \\ (m_1, m_2, \dots, m_n \neq \mathbf{0})}} \frac{1}{m_1! m_2! \dots m_n!} \prod_{i=1}^n t_i^{m_i} \mu_{m_1, m_2, \dots, m_n} \right)^j \end{aligned}$$

Property 2.7.1.

2.8 特征函数

Definition 2.14. 设 X 是一个随机变量。称：

$$\varphi_X(t) = E(e^{itX})$$

为 X 的特征函数 (*characteristic function, c.f.*), 其中 $t \in \mathbb{R}$ 。

Definition 2.15. 设 \mathbf{X} 是一个 n 维随机向量。称：

$$\varphi_{\mathbf{X}}(t) = E(e^{it^T \mathbf{X}})$$

为 \mathbf{X} 的特征函数, 其中 $t \in \mathbb{R}^n$ 。

Definition 2.16. 设 \mathbf{X} 是一个 $m \times n$ 随机矩阵。称：

$$\varphi_{\mathbf{X}}(t) = E \left[\exp \left(i \operatorname{tr}(t^T \mathbf{X}) \right) \right]$$

为 \mathbf{X} 的特征函数, 其中 $t \in M_{m \times n}(\mathbb{R})$ 。

Property 2.8.1. 设 $X, Y, X_1, X_2, \dots, X_n$ 是随机变量, $\alpha_1, \alpha_2, \dots, \alpha_n, \beta_1, \beta_2, \dots, \beta_n$ 为常数, 则：

1. X 的特征函数 $\varphi_X(t)$ 存在；
2. $|\varphi_X(t)| \leq \varphi_X(0) = 1$ ；
3. $\varphi_X(-t) = \overline{\varphi_X(t)}$ ；
4. 若 X_1, X_2, \dots, X_n 相互独立, 则 $Y = \sum_{k=1}^n (\alpha_k X_k + \beta_k)$ 的特征函数为：

$$\varphi_Y(t) = \prod_{k=1}^n e^{it\beta_k} \varphi_{X_k}(\alpha_k t)$$

5. X_1, X_2, \dots, X_n 相互独立的充分必要条件为:

$$\varphi_{X_1, \dots, X_n}(t_1, t_2, \dots, t_n) = \prod_{i=1}^n \varphi_{X_i}(t_i)$$

6. 特征函数与概率分布之间存在一个双射, 即 $\varphi_X(t) = \varphi_Y(t)$ 当且仅当 X 与 Y 具有相同的概率分布。

7. 若 $E(X^n)$ 存在, 则 $\varphi_X^{(n)}(t)$ 存在, 且对 $1 \leq k \leq n$ 有:

$$E(X^k) = i^{-k} \varphi_X^{(k)}(0)$$

特别的:

$$E(X) = -i\varphi_X'(0), \text{Var}(X) = -\varphi_X''(0) + [\varphi_X'(0)]^2$$

8. 若 $\varphi_X(t)$ 在 $t=0$ 处最高有 n 阶导数, 如果 n 为奇数, 则 X 具有所有不超过 $n-1$ 阶的原点矩; 若 n 为偶数, 则 X 具有所有不超过 n 阶的原点矩;

9. $\varphi_X(t)$ 在 \mathbb{R} 上一致连续;

10. $\varphi_X(t)$ 是半正定的, 即对任意的 $n \in \mathbb{N}^+$ 及任意的 $t = (t_1, t_2, \dots, t_n)^T \in \mathbb{R}^n$ 和任意的 $c = (c_1, c_2, \dots, c_n)^T \in \mathbb{C}^n$, 令 $A = [\varphi_X(t_i - t_j)] \in M_n(\mathbb{C})$, 则有:

$$c^T A \bar{c} = \sum_{i=1}^n \sum_{j=1}^n c_i \bar{c}_j \varphi_X(t_i - t_j) \geq 0$$

Proof. (1) 因为:

$$e^{itX} = \cos(tX) + i \sin(tX)$$

所以 $|e^{itX}| = 1$, 于是:

$$|E(e^{itX})| = \left| \int_{-\infty}^{+\infty} e^{itx} p(x) dx \right| \leq \int_{-\infty}^{+\infty} |e^{itx}| p(x) dx = \int_{-\infty}^{+\infty} p(x) dx = 1$$

链 接
Lebesgue
积分性质

所以 $\varphi_X(t)$ 存在。

(2) 可以发现:

$$\varphi_X(0) = \int_{-\infty}^{+\infty} p(x) dx = 1$$

再由 (1) 的证明过程即可得出结论。

(3) 因为:

$$\varphi_X(t) = E(e^{itX}) = E[\cos(tX) + i \sin(tX)] = E[\cos(tX)] + i E[\sin(tX)]$$

所以:

$$\overline{\varphi_X(t)} = E[\cos(tX)] - i E[\sin(tX)] = E[\cos(-tX)] + i E[\sin(-tX)] = \varphi_X(-t)$$

(4) 因为 X_k 相互独立, 所以 $e^{it(\alpha_k X_k + \beta_k)}$ 之间也相互独立, $k = 1, 2, \dots, n$, 于是有:

$$\begin{aligned}\varphi_Y(t) &= E \left[\exp \left(it \sum_{k=1}^n (\alpha_k X_k + \beta_k) \right) \right] = E \left(\prod_{k=1}^n e^{it(\alpha_k X_k + \beta_k)} \right) \\ &= \prod_{k=1}^n E[e^{it(\alpha_k X_k + \beta_k)}] = \prod_{k=1}^n e^{it\beta_k} E(e^{it\alpha_k X_k}) = \prod_{k=1}^n e^{it\beta_k} \varphi_{X_k}(\alpha_k t)\end{aligned}$$

(5) **必要性:** 因为 X_k 相互独立, 所以 $e^{it_k X_k}$ 相互独立, $k = 1, 2, \dots, n$ 。由随机向量特征函数的定义可得:

$$\begin{aligned}\varphi_{X_1, \dots, X_n}(t_1, t_2, \dots, t_n) &= E \left[\exp \left(i \sum_{k=1}^n t_k X_k \right) \right] = E \left(\prod_{k=1}^n e^{it_k X_k} \right) \\ &= \prod_{k=1}^n E(e^{it_k X_k}) = \prod_{k=1}^n \varphi_{X_k}(t_k)\end{aligned}$$

充分性: 因为:

$$\begin{aligned}\varphi_{X_1, \dots, X_n}(t_1, t_2, \dots, t_n) &= E \left[\exp \left(i \sum_{k=1}^n t_k X_k \right) \right] \\ &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp \left(i \sum_{k=1}^n t_k x_k \right) p(x_1, \dots, x_n) dx_1 \cdots dx_n \\ \prod_{i=1}^n \varphi_{X_i}(t_i) &= \prod_{k=1}^n E(e^{it_k X_k}) \\ &= \prod_{k=1}^n \int_{-\infty}^{+\infty} e^{it_k x_k} p(x_k) dx_k \\ &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp \left(i \sum_{k=1}^n t_k x_k \right) p(x_1) p(x_2) \cdots p(x_n) dx_1 dx_2 \cdots dx_n\end{aligned}$$

若两式相等, 则有:

$$p(x_1, x_2, \dots, x_n) = p(x_1)p(x_2) \cdots p(x_n)$$

链接独立性
条件

由可得 X_k , $k = 1, 2, \dots, n$ 相互独立。

(6)

(7) 因为 $E(X^n)$ 存在, 所以:

$$\int_{-\infty}^{+\infty} |x|^n p(x) dx < +\infty$$

于是:

$$\left| \int_{-\infty}^{+\infty} i^n x^n e^{itx} p(x) dx \right| \leq \int_{-\infty}^{+\infty} |x|^n p(x) dx < +\infty$$

所以:

$$\varphi_X^{(n)}(t) = \int_{-\infty}^{+\infty} i^n x^n e^{itx} p(x) dx$$

存在。由定理 2.1 可知对 $1 \leq k \leq n$ 有 $E(X^k)$ 存在, 于是:

$$\varphi_X^{(k)}(0) = \int_{-\infty}^{+\infty} i^k x^k p(x) dx = i^k \int_{-\infty}^{+\infty} x^k p(x) dx = i^k E(X^k)$$

也存在。

(8) 注意到:

$$\varphi_X^{(n)}(t) = \int_{-\infty}^{+\infty} i^n x^n e^{itx} p(x) dx$$

因为 $\varphi_X(t)$ 在 $t=0$ 处最高具有 n 阶导数, 于是:

$$|\varphi_X^{(n)}(0)| = \left| \int_{-\infty}^{+\infty} i^n x^n p(x) dx \right| = \left| \int_{-\infty}^{+\infty} x^n p(x) dx \right| < +\infty$$

当 $n = 2k + 1, k \in \mathbb{N}$ 时, 有:

$$\int_{-\infty}^{+\infty} |x|^n p(x) dx > |\varphi_X^{(n)}(0)| = \left| \int_{-\infty}^{+\infty} x^n p(x) dx \right|$$

所以 $E(X^n)$ 不一定存在。当 $n = 2k, k \in \mathbb{N}^+$ 时, 有:

$$|\varphi_X^{(n)}(0)| = \left| \int_{-\infty}^{+\infty} x^n p(x) dx \right| = \int_{-\infty}^{+\infty} |x|^n p(x) dx < +\infty$$

需要证明对
小于的都存
在

存在, 于是 $E(X^n)$ 存在。由定理 2.1 可知, 此时 X 具有所有不超过 n 阶的原点矩。

(9) 对任意的 $t, h \in \mathbb{R}$ 和 $a > 0$, 有:

$$\begin{aligned} |\varphi(t+h) - \varphi(t)| &= \left| \int_{-\infty}^{+\infty} [e^{i(t+h)x} - e^{itx}] p(x) dx \right| \\ &= \left| \int_{-\infty}^{+\infty} (e^{ihx} - 1) e^{itx} p(x) dx \right| \\ &\leq \int_{-\infty}^{+\infty} |e^{ihx} - 1| e^{itx} |p(x)| dx \\ &= \int_{-\infty}^{+\infty} |e^{ihx} - 1| e^{itx} p(x) dx \\ &= \int_{-\infty}^{+\infty} |e^{ihx} - 1| p(x) dx \\ &= \int_{-a}^a |e^{ihx} - 1| p(x) dx + \int_{|x| \geq a} |e^{ihx} - 1| p(x) dx \\ &\leq \int_{-a}^a |e^{ihx} - 1| p(x) dx + \int_{|x| \geq a} (|e^{ihx}| + 1) p(x) dx \\ &= \int_{-a}^a |e^{ihx} - 1| p(x) dx + 2 \int_{|x| \geq a} p(x) dx \end{aligned}$$

对于任意的 $\varepsilon > 0$, 可以先选定一个充分大的 a , 使得:

$$2 \int_{|x| \geq a} p(x) dx < \frac{\varepsilon}{2}$$

对任意的 $x \in [-a, a]$, 只要取 $\delta = \frac{\varepsilon}{2a}$, 则当 $|h| < \delta$ 时, 就有:

$$\begin{aligned}
 |e^{ihx} - 1| &= \left| e^{ihx} - e^{i\frac{hx}{2}} e^{i\frac{-hx}{2}} \right| = \left| e^{i\frac{hx}{2}} (e^{i\frac{hx}{2}} - e^{i\frac{-hx}{2}}) \right| \\
 &= \left| e^{i\frac{hx}{2}} \right| \left| e^{i\frac{hx}{2}} - e^{i\frac{-hx}{2}} \right| \\
 &= \left| e^{i\frac{hx}{2}} - e^{i\frac{-hx}{2}} \right| \\
 &= \left| \cos \frac{hx}{2} + i \sin \frac{hx}{2} - \cos \frac{-hx}{2} - i \sin \frac{-hx}{2} \right| \\
 &= \left| 2i \sin \frac{hx}{2} \right| = 2 \left| \sin \frac{hx}{2} \right| \leq 2 \left| \frac{hx}{2} \right| \leq ha < \frac{\varepsilon}{2}
 \end{aligned}$$

于是对任意的 $t \in \mathbb{R}$, 有:

$$|\varphi(t+h) - \varphi(t)| < \int_{-a}^a \frac{\varepsilon}{2} p(x) dx + 2 \int_{|x| \geq a} p(x) dx < \frac{\varepsilon}{2} \int_{-\infty}^{+\infty} p(x) dx + \frac{\varepsilon}{2} = \varepsilon$$

即 $\varphi_X(t)$ 在 \mathbb{R} 上一致连续。

(10) 显然:

$$\begin{aligned}
 \sum_{i=1}^n \sum_{j=1}^n c_i \bar{c}_j \varphi_X(t_i - t_j) &= \sum_{k=1}^n \sum_{j=1}^n c_k \bar{c}_j \int_{-\infty}^{+\infty} e^{i(t_k - t_j)x} p(x) dx \\
 &= \int_{-\infty}^{+\infty} \sum_{k=1}^n \sum_{j=1}^n c_k \bar{c}_j e^{i(t_k - t_j)x} p(x) dx \\
 &= \int_{-\infty}^{+\infty} \left(\sum_{k=1}^n c_k e^{it_k x} \right) \left(\sum_{j=1}^n \bar{c}_j e^{-it_j x} \right) p(x) dx \\
 &= \int_{-\infty}^{+\infty} \left(\sum_{k=1}^n c_k e^{it_k x} \right) \left(\sum_{j=1}^n \overline{c_k e^{it_k x}} \right) p(x) dx \\
 &= \int_{-\infty}^{+\infty} \left| \sum_{k=1}^n c_k e^{it_k x} \right|^2 p(x) dx
 \end{aligned}$$

□

2.9 Fisher 信息量

Definition 2.17. 设 \mathbf{X} 是测度空间 (X, \mathcal{F}, μ) 上的一个随机向量, 其分布由 n 维参数 $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ 决定, \mathbf{X} 的概率函数为 $f(\mathbf{X}; \theta)$ 。若 $f(\mathbf{X}; \theta)$ 满足如下正则条件:

1. $f(\mathbf{X}; \theta)$ 关于 θ 的偏导数 a.e. 存在;
2. 对 $f(\mathbf{X}; \theta)$ 在 X 上的积分关于 θ 任一分量求导时都可以交换求导与积分的顺序;
3. $f(\mathbf{X}; \theta)$ 的定义域与 θ 无关。

则称:

$$[I(\theta)]_{ij} = E \left[\left(\frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta_i} \right) \left(\frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta_j} \right) \right]$$

为 \mathbf{X} 的 Fisher 信息矩阵 (Fisher information matrix, FIM)。

note 2.1. Fisher 信息量（即一维情况的信息矩阵）用来表明随机变量 \mathbf{X} 携带的关于参数 θ 的信息。如果它比较大，表示平均下来 θ 的微小变化会给 \mathbf{X} 的分布带来较大的变化，即 \mathbf{X} 的分布很依赖 θ 的具体取值，所以携带了较多关于 θ 的信息。

Property 2.9.1. 设 \mathbf{X} 是测度空间 (X, \mathcal{F}, μ) 上的一个随机向量，其 Fisher 信息矩阵具有如下性质：

1. Fisher 信息矩阵可以看作协方差矩阵：

$$I(\theta) = \text{Cov} \left[\frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta} \right]$$

2. 若 $\ln f(\mathbf{X}; \theta)$ 有关于 θ 的所有二阶导数，且对该二阶导数在 X 上的积分关于 θ 任一分量求导时都可以交换求导与积分的顺序，则：

$$[I(\theta)]_{ij} = -\mathbb{E} \left[\frac{\partial^2 \ln f(\mathbf{X}; \theta)}{\partial \theta_i \partial \theta_j} \right]$$

Proof. (1) 由正则条件可得：

$$\begin{aligned} \mathbb{E} \left[\frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta_i} \right] &= \int_X \frac{\frac{\partial f(\mathbf{X}; \theta)}{\partial \theta_i}}{f(\mathbf{X}; \theta)} f(\mathbf{X}; \theta) d\mu = \int_X \frac{\partial f(\mathbf{X}; \theta)}{\partial \theta_i} d\mu \\ &= \frac{\partial}{\partial \theta_i} \int_X f(\mathbf{X}; \theta) d\mu = \frac{\partial 1}{\partial \theta_i} = 0 \end{aligned}$$

(2) 由 (1) 和正则条件可得：

$$\begin{aligned} \frac{\partial}{\partial \theta_j} \mathbb{E} \left[\frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta_i} \right] &= 0 \\ \frac{\partial}{\partial \theta_j} \int_X \frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta_i} f(\mathbf{X}; \theta) d\mu &= 0 \\ \int_X \frac{\partial}{\partial \theta_j} \left[\frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta_i} f(\mathbf{X}; \theta) \right] d\mu &= 0 \\ \int_X \left[\frac{\partial^2 \ln f(\mathbf{X}; \theta)}{\partial \theta_i \partial \theta_j} f(\mathbf{X}; \theta) + \frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta_i} \frac{\partial f(\mathbf{X}; \theta)}{\partial \theta_j} \right] d\mu &= 0 \\ \int_X \left[\frac{\partial^2 \ln f(\mathbf{X}; \theta)}{\partial \theta_i \partial \theta_j} f(\mathbf{X}; \theta) + \frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta_i} \frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta_j} f(\mathbf{X}; \theta) \right] d\mu &= 0 \\ \mathbb{E} \left[\left(\frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta_i} \right) \left(\frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta_j} \right) \right] &= -\mathbb{E} \left[\frac{\partial^2 \ln f(\mathbf{X}; \theta)}{\partial \theta_i \partial \theta_j} \right] \end{aligned}$$

其中倒数第三行到倒数第二行是因为：

$$\frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta_i} f(\mathbf{X}; \theta) = \frac{\frac{\partial f(\mathbf{X}; \theta)}{\partial \theta_i}}{f(\mathbf{X}; \theta)} f(\mathbf{X}; \theta) = \frac{\partial f(\mathbf{X}; \theta)}{\partial \theta_i}$$

□

Chapter 3

正态分布与三大抽样分布

3.1 多元正态分布

3.1.1 多元正态分布的定义

Definition 3.1. 若一个随机向量 $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)^T \in \mathbb{R}^n$ 满足以下概率密度函数:

$$p(\mathbf{X}|\boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det \Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu})}$$

则称其为一个正态随机向量, 记作 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$ 。其中, $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T \in \mathbb{R}^n$, $\Sigma \in M_n(\mathbb{R})$ 且 $\Sigma > 0$ 。

Theorem 3.1. 对于正态随机向量的概率密度函数, $\boldsymbol{\mu}$ 和 Σ 分别是 \mathbf{X} 的均值向量和协方差矩阵。

Proof. 令:

$$\mathbf{Y} = \Sigma^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu})$$

则有 $\mathbf{X} = \Sigma^{\frac{1}{2}}\mathbf{Y} + \boldsymbol{\mu}$, 由求随机变量函数的分布中的变量变换法可知¹:

$$p(\mathbf{Y}) = p(\Sigma^{\frac{1}{2}}\mathbf{Y} + \boldsymbol{\mu})|\mathbf{J}|$$

其中 \mathbf{J} 为变换的 Jacobi 行列式:

$$\mathbf{J} = \begin{vmatrix} \frac{\partial \mathbf{X}_1}{\partial \mathbf{Y}_1} & \cdots & \frac{\partial \mathbf{X}_1}{\partial \mathbf{Y}_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{X}_n}{\partial \mathbf{Y}_1} & \cdots & \frac{\partial \mathbf{X}_n}{\partial \mathbf{Y}_n} \end{vmatrix} = \det \Sigma^{\frac{1}{2}} = (\det \Sigma)^{\frac{1}{2}}$$

那么 \mathbf{Y} 的概率密度函数为:

$$p(\mathbf{Y}) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\mathbf{Y}^T \mathbf{Y}} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{\mathbf{Y}_i^2}{2}}$$

¹这是由于 \mathbf{X} 到 \mathbf{Y} 的变换是一个线性变换, 线性变换让 \mathbf{X} 关于 \mathbf{Y} 有连续偏导数, 同时这个线性变换可逆, 满足变量变换法的两大条件。

写完变量变换法链接过来

对 \mathbf{Y}_i 求边缘分布可得 $\mathbf{Y}_i \sim N(0, 1)$, 并且可以发现 \mathbf{Y} 的 n 个分量的联合密度等于每个分量密度函数的乘积, 于是 \mathbf{Y} 的各个分量相互独立, 所以有:

$$E(\mathbf{Y}) = \mathbf{0}, \text{Cov}(\mathbf{Y}) = \mathbf{I}$$

结合 \mathbf{Y} 与 \mathbf{X} 的关系, 由性质 2.4.1 可得:

$$E(\mathbf{X}) = \boldsymbol{\mu}, \text{Cov}(\mathbf{X}) = \Sigma \quad \square$$

Definition 3.2. 正态随机向量 \mathbf{X} 的概率密度函数也可写作:

$$p(\mathbf{X}) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det \Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2} \text{tr}[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T \Sigma^{-1}]}$$

Proof. 只需注意到二次型的迹就是自身以及性质 4.7.2:

$$(\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}) = \text{tr}[(\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu})] = \text{tr}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1}] \quad \square$$

多元正态分布的等价定义

Definition 3.3. \mathbf{X} 为 n 维随机向量。若存在矩阵 $A \in M_{n \times r}(\mathbb{R})$ 使得 $\mathbf{X} = A\mathbf{U} + \boldsymbol{\mu}$, 其中 $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_r)^T$, $\mathbf{U}_i \sim N(0, 1)$ 且互相独立, $\boldsymbol{\mu}$ 为 n 维非随机实向量, 则称 \mathbf{X} 为服从均值为 $\boldsymbol{\mu}$ 、协方差矩阵为 $\Sigma = AA^T$ 的多元正态向量, 记为 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, 其中 $\Sigma \geq 0$ 。若 $|\Sigma| = 0$, 则称此时的分布为奇异正态分布。

Theorem 3.2. \mathbf{X} 是一个随机向量, 其协方差矩阵为正定矩阵, 则 \mathbf{X} 满足定义 3.1 的充分必要条件是满足定义 3.3, 即两种正态分布的定义在随机向量的协方差矩阵是正定矩阵的情形下是等价的。

Proof. (1) 充分性: 设 \mathbf{X} 满足定义 3.3, 因为 \mathbf{U} 中的元素服从标准正态分布且彼此独立, 所以有:

$$E(\mathbf{U}) = \mathbf{0}, \text{Cov}(\mathbf{U}) = I$$

同时:

$$p(\mathbf{U}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{\mathbf{U}_i^2}{2}} = \frac{1}{(2\pi)^{\frac{n}{2}} [\det \text{Cov}(\mathbf{U})]^{\frac{1}{2}}} e^{-\frac{1}{2} \mathbf{U}^T [\text{Cov}(\mathbf{U})]^{-1} \mathbf{U}}$$

因为 $\mathbf{X} = A\mathbf{U} + \boldsymbol{\mu}$, 由和性质 2.4.1(3)(4)(5) 可得:

$$E(\mathbf{X}) = \boldsymbol{\mu}, \text{Cov}(\mathbf{X}) = AA^T$$

因为 $\text{Cov}(\mathbf{X}) > 0$, 由定理 1.44 可得 $\text{rank}[\text{Cov}(\mathbf{X})] = n$ 。因为 $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$, 所以 $\text{rank}(A) = n$, 即 $r = n$, A 是一个 n 阶可逆矩阵, 于是 $\mathbf{U} = A^{-1}(\mathbf{X} - \boldsymbol{\mu})$ 。由求随机变量函数的分布中的变量变换法可知:

期望的性质

写完矩阵的秩做链接

写完变量变换法链接过来

$$\begin{aligned}
P(\mathbf{X}) &= P[A^{-1}(\mathbf{X} - \boldsymbol{\mu})] |\mathbf{J}| \\
&= \frac{1}{(2\pi)^{\frac{n}{2}} \{\det \text{Cov}[A^{-1}(\mathbf{X} - \boldsymbol{\mu})]\}^{\frac{1}{2}}} \\
&\quad \cdot e^{-\frac{1}{2}[A^{-1}(\mathbf{X} - \boldsymbol{\mu})]^T \{\text{Cov}[A^{-1}(\mathbf{X} - \boldsymbol{\mu})]\}^{-1} [A^{-1}(\mathbf{X} - \boldsymbol{\mu})]} |\det A^{-1}| \\
&= \frac{1}{(2\pi)^{\frac{n}{2}} [\det \text{Cov}(\mathbf{X})]^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^T [\text{Cov}(\mathbf{X})]^{-1} (\mathbf{X} - \boldsymbol{\mu})}
\end{aligned}$$

即 \mathbf{X} 满足定义 3.4。

(2) 必要性：设 \mathbf{X} 满足定义 3.1，此时只要选择 $A = \Sigma^{\frac{1}{2}}$ 即可得到 \mathbf{X} 满足定义 3.3。□

3.1.2 多元正态分布的性质

Theorem 3.3. 设 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, $\Sigma \geq 0$, $B \in M_{m \times n}(\mathbb{R})$, $c \in \mathbb{R}^m$, 则 $\mathbf{Y} = B\mathbf{X} + c \sim N(B\boldsymbol{\mu} + c, B\Sigma B^T)$ 。

Proof. 因为 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, 由定义 3.3 可得, 存在 $A \in M_{n \times r}(\mathbb{R})$, $\boldsymbol{\mu} \in \mathbb{R}^n$ 使得:

$$\mathbf{X} = A\mathbf{U} + \boldsymbol{\mu}, \quad AA^T = \Sigma, \quad \mathbf{U} \sim N(\mathbf{0}, I)$$

于是:

$$\mathbf{Y} = B(A\mathbf{U} + \boldsymbol{\mu}) + c = BA\mathbf{U} + B\boldsymbol{\mu} + c$$

注意到 $BA(BA)^T = BAA^TB^T = B\Sigma B^T$, 由定义 3.3 可得 $\mathbf{Y} \sim N(B\boldsymbol{\mu} + c, B\Sigma B^T)$ 。□

Corollary 3.1. 由上述定理可以得到如下推论:

1. 设 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, $\Sigma > 0$, 则 $\mathbf{Y} = \Sigma^{-\frac{1}{2}}\mathbf{X} \sim N_n(\Sigma^{-\frac{1}{2}}\boldsymbol{\mu}, I_n)$;
2. 设 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \sigma^2 I_n)$, Q 为正交矩阵, 则 $Q\mathbf{X} \sim N_n(Q\boldsymbol{\mu}, \sigma^2 I_n)$;
3. 设 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, $c \in \mathbb{R}^n$, 则 $c^T\mathbf{X} \sim N(c^T\boldsymbol{\mu}, c^T\Sigma c)$;
4. 设 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$, $\Sigma = (\sigma_{ij})$, 则 $\mathbf{X}_i \sim N(\mu_i, \sigma_{ii})$, $i = 1, 2, \dots, n$ 。
5. 设 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$, $\Sigma = (\sigma_{ij})$, $i_1 < i_2 < \dots < i_k$, 则有 $(\mathbf{X}_{i_1}, \mathbf{X}_{i_2}, \dots, \mathbf{X}_{i_k})^T \sim N(\boldsymbol{\mu}_0, \Sigma_0)$, 其中:

$$\boldsymbol{\mu}_0 = \begin{pmatrix} \mu_{i_1} \\ \mu_{i_2} \\ \vdots \\ \mu_{i_k} \end{pmatrix}, \quad \Sigma_0 = \begin{pmatrix} \sigma_{i_1 i_1} & \sigma_{i_1 i_2} & \cdots & \sigma_{i_1 i_k} \\ \sigma_{i_2 i_1} & \sigma_{i_2 i_2} & \cdots & \sigma_{i_2 i_k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{i_k i_1} & \sigma_{i_k i_2} & \cdots & \sigma_{i_k i_k} \end{pmatrix}$$

Proof. (1) 由性质 4.7.5(3) 可知 $\Sigma^{-\frac{1}{2}}$ 是对称阵, 所以:

$$\Sigma^{-\frac{1}{2}}\Sigma(\Sigma^{-\frac{1}{2}})^T = \Sigma^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}} = I_n$$

(2) 显然:

$$Q\sigma^2 I Q^T = \sigma^2 Q Q^T = \sigma^2 I_n$$

(3) 可直接得到;

(4) 对 \mathbf{X}_i , 取 $c = (0, \dots, 0, 1, 0, \dots, 0)^T$, 其中 c 的第 i 位为 1 其余全是 0, 于是:

$$c^T \mathbf{X} = \mathbf{X}_i, \quad c^T \boldsymbol{\mu} = \mu_i, \quad c^T \Sigma c = \sigma_{ii}$$

所以 $\mathbf{X}_i \sim N(\mu_i, \sigma_{ii})$, $i = 1, 2, \dots, n$ 。

(5) 取:

$$A = \begin{pmatrix} e_{i_1}^T \\ e_{i_2}^T \\ \vdots \\ e_{i_k}^T \end{pmatrix}$$

其中 e_{i_j} 为单位列向量, 只在第 i_j 位取 1, 其余位置上元素为 0, $j = 1, 2, \dots, k$ 。于是有:

$$A\boldsymbol{\mu} = (\mu_{i_1}, \mu_{i_2}, \dots, \mu_{i_k})^T = \boldsymbol{\mu}_0$$

$$A\Sigma A^T = \begin{pmatrix} e_{i_1}^T \Sigma e_{i_1} & e_{i_1}^T \Sigma e_{i_2} & \cdots & e_{i_1}^T \Sigma e_{i_k} \\ e_{i_2}^T \Sigma e_{i_1} & e_{i_2}^T \Sigma e_{i_2} & \cdots & e_{i_2}^T \Sigma e_{i_k} \\ \vdots & \vdots & \ddots & \vdots \\ e_{i_k}^T \Sigma e_{i_1} & e_{i_k}^T \Sigma e_{i_2} & \cdots & e_{i_k}^T \Sigma e_{i_k} \end{pmatrix} = \begin{pmatrix} \sigma_{i_1 i_1} & \sigma_{i_1 i_2} & \cdots & \sigma_{i_1 i_k} \\ \sigma_{i_2 i_1} & \sigma_{i_2 i_2} & \cdots & \sigma_{i_2 i_k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{i_k i_1} & \sigma_{i_k i_2} & \cdots & \sigma_{i_k i_k} \end{pmatrix} = \Sigma_0$$

由上一定理可得 $(\mathbf{X}_{i_1}, \mathbf{X}_{i_2}, \dots, \mathbf{X}_{i_k})^T \sim N(\boldsymbol{\mu}_0, \Sigma_0)$ 。 □

Theorem 3.4. 设 \mathbf{X} 是一个随机向量, 则 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$ 当且仅当它的特征函数为:

$$\varphi_{\mathbf{X}}(t) = \exp \left(it^T \boldsymbol{\mu} - \frac{t^T \Sigma t}{2} \right), \quad t \in \mathbb{R}^n$$

Proof. 若 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, 则由定义 3.3 可知, 存在矩阵 $A \in M_{n \times r}(\mathbb{R})$ 使得 $\mathbf{X} = A\mathbf{U} + \boldsymbol{\mu}$, 其中 $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_r)^T$, $\mathbf{U}_i \sim N(0, 1)$ 且互相独立, $\boldsymbol{\mu}$ 为 n 维非随机实向量, $\Sigma = AA^T$ 。由性质 2.8.1(5) 可得:

$$\varphi_{\mathbf{U}}(t) = \prod_{i=1}^n \varphi_{\mathbf{U}_i}(t_i) = \prod_{i=1}^n e^{-\frac{t_i^2}{2}} = e^{-\frac{t^T t}{2}}, \quad t \in \mathbb{R}^n$$

于是:

$$\begin{aligned} \varphi_{\mathbf{X}}(t) &= E(e^{it^T \mathbf{X}}) = E[e^{it^T (A\mathbf{U} + \boldsymbol{\mu})}] = e^{it^T \boldsymbol{\mu}} E(e^{it^T A\mathbf{U}}) \\ &= e^{it^T \boldsymbol{\mu}} \varphi_{\mathbf{U}}(A^T t) = e^{it^T \boldsymbol{\mu}} e^{-\frac{t^T A A^T t}{2}} = e^{it^T \boldsymbol{\mu}} e^{-\frac{t^T \Sigma t}{2}} = \exp \left(it^T \boldsymbol{\mu} - \frac{t^T \Sigma t}{2} \right) \end{aligned}$$

由性质 2.8.1(6), 概率分布与特征函数之间是一一对应的关系, 于是结论成立。 □

Theorem 3.5. 设 \mathbf{X} 是一个 n 维随机向量, 则 \mathbf{X} 服从 n 维多元正态分布的充分必要条件为对于任意的 $\alpha \in \mathbb{R}^n$, $\alpha^T \mathbf{X}$ 服从正态分布。

Proof. (1) 必要性: 由推论 3.1(3) 直接得到。

(2) 充分性: 由定理 3.4 可知此时 $\alpha^T \mathbf{X}$ 的特征函数为:

$$\varphi_{\alpha^T \mathbf{X}}(t) = \exp\left(it\mu - \frac{1}{2}t^2\sigma^2\right)$$

期望的性质 其中 μ 和 σ^2 分别为 $\alpha^T \mathbf{X}$ 的均值与方差。由和性质 2.4.1(3) 可得:

$$\mu = E(\alpha^T \mathbf{X}) = \alpha^T E(\mathbf{X}), \quad \sigma^2 = \text{Cov}(\alpha^T \mathbf{X}) = \alpha^T \text{Cov}(\mathbf{X}) \alpha$$

于是有:

$$\varphi_{\alpha^T \mathbf{X}}(t) = \exp\left(it\alpha^T E(\mathbf{X}) - \frac{t\alpha^T \text{Cov}(\mathbf{X})\alpha t}{2}\right)$$

由 α 的任意性, 上式可写作:

$$\varphi_{\mathbf{X}}(\beta) = \exp\left(i\beta^T E(\mathbf{X}) - \frac{\beta^T \text{Cov}(\mathbf{X})\beta}{2}\right)$$

由性质 2.8.1(6) 和定理 3.4 可知 \mathbf{X} 服从多元正态分布。 \square

Lemma 3.1. 设 \mathbf{X} 和 \mathbf{Y} 分别为 m 维正态随机向量和 n 维正态随机向量且相互独立, 则:

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N_{m+n} \left(\begin{pmatrix} \mu_{\mathbf{X}} \\ \mu_{\mathbf{Y}} \end{pmatrix}, \begin{pmatrix} \text{Cov}(\mathbf{X}) & \mathbf{0} \\ \mathbf{0} & \text{Cov}(\mathbf{Y}) \end{pmatrix} \right)$$

Proof. 由定义 3.1 即可得到。 \square

Theorem 3.6. 设 $\mathbf{X} \sim N_n(\mu, \Sigma_{\mathbf{X}})$ 和 $\mathbf{Y} \sim N_n(\nu, \Sigma_{\mathbf{Y}})$ 且相互独立, 则 $\mathbf{X} + \mathbf{Y} \sim N_n(\mu + \nu, \Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Y}})$ 。

Proof. 因为:

$$\mathbf{X} + \mathbf{Y} = \begin{pmatrix} I_n & I_n \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$$

由引理 3.1 和定理 3.3 即可得到结论。 \square

Theorem 3.7. 设 $\mathbf{X} = (X_1, X_2, \dots, X_n)^T \sim N_n(\mu, \Sigma)$, 其中:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}$$

则对任意的 $m \leq n$, $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m)^T$, \mathbf{X}_j 是 r_j 维随机向量, $\sum_{j=1}^m r_j = n$, 若 \mathbf{X}_j 不相关, 则 \mathbf{X}_j 相互独立, $j = 1, 2, \dots, m$ 。即对于服从正态分布的随机变量或随机向量而言, 不相关和独立等价。

Proof. 由推论 3.1(5) 可知 $\mathbf{X}_j = (X_{j1}, X_{j2}, \dots, X_{jr_j})^T \sim N_{r_j}(\mu_j, \Sigma_j)$ 。由定理 3.4 可知:

$$\varphi_{\mathbf{X}_j}(t_j) = \exp\left(it_j^T \mu_j - \frac{t_j^T \Sigma_j t_j}{2}\right), \quad t_j \in \mathbb{R}^{r_j}$$

将 μ 和 Σ 写成分块的形式:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1m} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{m1} & \Sigma_{m2} & \cdots & \Sigma_{mm} \end{pmatrix}$$

因为 \mathbf{X}_j 不相关, 所以 $\text{Cov}(\mathbf{X}_j, \mathbf{X}_k) = \Sigma_{jk} = \mathbf{0}$, $j \neq k$, $j, k = 1, 2, \dots, m$ 。于是:

$$\Sigma = \begin{pmatrix} \Sigma_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Sigma_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Sigma_m \end{pmatrix}$$

由定理 3.4 可得:

$$\begin{aligned} \varphi_{\mathbf{X}}(t) &= \exp\left(it^T \mu - \frac{t^T \Sigma t}{2}\right) = \exp\left[it \sum_{j=1}^m t_j^T \mu_j - \frac{\sum_{j=1}^m t_j^T \Sigma_j t_j}{2}\right] \\ &= \prod_{j=1}^m \exp\left(it_j^T \mu_j - \frac{t_j^T \Sigma_j t_j}{2}\right) = \prod_{j=1}^m \varphi_{\mathbf{X}_j}(t_j) \end{aligned}$$

由性质 2.8.1(5) 可知 \mathbf{X}_j 相互独立, $j = 1, 2, \dots, m$ 。 □

正态随机向量的二次型

Theorem 3.8. 设 $\mathbf{X} \sim N_n(\mu, \Sigma)$, $\Sigma > 0$, A 为 n 阶非随机实对称阵, 则:

$$E(\mathbf{X}^T A \mathbf{X}) = \mu^T A \mu + \text{tr}(A \Sigma), \quad \text{Var}(\mathbf{X}^T A \mathbf{X}) = 2 \text{tr}[(A \Sigma)^2] + 4 \mu^T A \Sigma A \mu$$

Proof. 期望可直接由定理 2.3 得到。记 $\mathbf{Y} = \Sigma^{-\frac{1}{2}} \mathbf{X}$, 由推论 3.1(1) 可知 $\mathbf{Y} \sim N_n(\Sigma^{-\frac{1}{2}} \mu, I_n)$, 根据定理 3.7, \mathbf{Y} 的各分量相互独立。注意到 \mathbf{Y} 的各分量的三阶中心矩和四阶中心矩分别为 0 和 3, 由定理 2.4、性质 4.7.5(3) 和性质 4.7.2(3) 可得:

$$\begin{aligned} \text{Var}(\mathbf{X}^T A \mathbf{X}) &= \text{Var}(\mathbf{Y}^T \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \mathbf{Y}) \\ &= 3 \sum_{i=1}^n (\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}})_{ii}^2 + 2 \text{tr}(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}) - 3 \sum_{i=1}^n (\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}})_{ii}^2 \\ &\quad + 4 \mu^T \Sigma^{-\frac{1}{2}} \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \Sigma^{-\frac{1}{2}} \mu \\ &= 2 \text{tr}(\Sigma^{\frac{1}{2}} A \Sigma A \Sigma^{\frac{1}{2}}) + 4 \mu^T A \Sigma A \mu \\ &= 2 \text{tr}(A \Sigma A \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}}) + 4 \mu^T A \Sigma A \mu \\ &= 2 \text{tr}[(A \Sigma)^2] + 4 \mu^T A \Sigma A \mu \end{aligned}$$
□

Theorem 3.9. 设 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, $\Sigma > 0$, 则 $(\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_n^2$ 。

Proof. 因为 $\Sigma > 0$, 所以存在 $\Sigma^{-\frac{1}{2}}$ 。由推论 3.1(1) 可得:

$$\Sigma^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu}) \sim N_n(\mathbf{0}, I_n)$$

于是根据性质 4.7.5(1) 和 (3) 可得:

$$\begin{aligned} (\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}) &= (\mathbf{X} - \boldsymbol{\mu})^T (\Sigma^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}}) (\mathbf{X} - \boldsymbol{\mu}) \\ &= (\mathbf{X} - \boldsymbol{\mu})^T (\Sigma^{-\frac{1}{2}})^T \Sigma^{-\frac{1}{2}} (\mathbf{X} - \boldsymbol{\mu}) \\ &= [\Sigma^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu})]^T \Sigma^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi_n^2 \quad \square \end{aligned}$$

Theorem 3.10. 设 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, I_n)$, $A \in M_n(K)$ 是一个非随机实对称矩阵, 则 $\mathbf{X}^T A \mathbf{X} \sim \chi_{r, \boldsymbol{\mu}^T A \boldsymbol{\mu}}^2$ 的充分必要条件为 A 是一个幂等阵且 $\text{rank}(A) = r$ 。

Proof. (1) 充分性: 因为 A 是一个幂等阵, 由性质 1.8.1(1) 可知 A 的特征值只能为 0 或 1。根据性质 1.6.2(3) 可知存在正交矩阵 Q 使得:

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

令 $\mathbf{Y} = Q\mathbf{X}$, 由推论 3.1(2) 可知 $\mathbf{Y} \sim N_n(Q\boldsymbol{\mu}, I_n)$ 。对 \mathbf{Y} 和 Q 进行分块:

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \quad Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}$$

其中 \mathbf{Y}_1 为 r 维随机向量, Q_1 为 $r \times n$ 矩阵, 所以:

$$\begin{aligned} \mathbf{X}^T A \mathbf{X} &= \mathbf{X}^T Q^{-1} \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q \mathbf{X} = \mathbf{X}^T Q^T \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q \mathbf{X} \\ &= \mathbf{Y}^T \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1^T & \mathbf{Y}_2^T \end{pmatrix} \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \mathbf{Y}_1^T \mathbf{Y}_1 \sim \chi_{r, \lambda}^2 \end{aligned}$$

其中:

$$\lambda = (Q_1 \boldsymbol{\mu})^T Q_1 \boldsymbol{\mu} = \boldsymbol{\mu}^T Q_1^T Q_1 \boldsymbol{\mu} = \boldsymbol{\mu}^T A \boldsymbol{\mu}$$

这是因为推论 3.1(5) 和:

$$A = Q^{-1} \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q = Q^T \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q = \begin{pmatrix} Q_1^T & Q_2^T \end{pmatrix} \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = Q_1^T Q_1$$

(2) 必要性: 设 $\text{rank}(A) = t$ 。因为 A 是 Hermitian 矩阵, 由性质 1.6.2(3) 可知存在正交阵 Q 使得:

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

其中 $A = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_t\}$, $\lambda_1, \lambda_2, \dots, \lambda_t$ 是 A 的非零特征值。若能证得 $\lambda_i = 1$, $i = 1, 2, \dots, t$ 且 $t = r$, 则 A 是一个幂等阵且 $\text{rank}(A) = r$ 。注意到:

$$\mathbf{X}^T A \mathbf{X} = \mathbf{X}^T Q^T \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q \mathbf{X}$$

令 $\mathbf{Y} = Q \mathbf{X}$, 由推论 3.1(2) 可知 $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n) \sim N_n(Q\boldsymbol{\mu}, I_n)$, 根据定理 3.7 可得 \mathbf{Y}_j 之间彼此独立。令 $\mathbf{c} = Q\boldsymbol{\mu} = (c_1, c_2, \dots, c_n)^T$, 由推论 3.1(5) 可知 $\mathbf{Y}_j \sim N(c_j, 1)$ 。而:

$$\mathbf{X}^T A \mathbf{X} = \mathbf{Y}^T \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Y} = \sum_{j=1}^t \lambda_j \mathbf{Y}_j^2$$

由性质 2.8.1(4) 和性质 3.2.1(3) 可知:

$$\begin{aligned} \varphi_{\mathbf{X}^T A \mathbf{X}}(t) &= \prod_{j=1}^t \varphi_{\lambda_j \mathbf{Y}_j^2}(t) = \prod_{j=1}^t (1 - 2it)^{-\frac{1}{2}} \exp \left\{ \frac{i\lambda_j t c_j^2}{1 - 2i\lambda_j t} \right\} \\ &= (1 - 2it)^{-\frac{t}{2}} \prod_{j=1}^t \exp \left\{ \frac{i\lambda_j t c_j^2}{1 - 2i\lambda_j t} \right\} \end{aligned}$$

因为 $\mathbf{X}^T A \mathbf{X} \sim \chi_{r, \boldsymbol{\mu}^T A \boldsymbol{\mu}}^2$, 所以:

$$\varphi_{\mathbf{X}^T A \mathbf{X}}(t) = (1 - 2it)^{-\frac{r}{2}} \exp \left\{ \frac{it\boldsymbol{\mu}^T A \boldsymbol{\mu}}{1 - 2it} \right\}$$

由性质 2.8.1(6) 可知:

$$(1 - 2it)^{-\frac{t}{2}} \prod_{j=1}^t \exp \left\{ \frac{i\lambda_j t c_j^2}{1 - 2i\lambda_j t} \right\} = (1 - 2it)^{-\frac{r}{2}} \exp \left\{ \frac{it\boldsymbol{\mu}^T A \boldsymbol{\mu}}{1 - 2it} \right\}$$

所以 $t = r$, 同时有:

$$\sum_{j=1}^t \frac{i\lambda_j t c_j^2}{1 - 2i\lambda_j t} = \frac{it\boldsymbol{\mu}^T A \boldsymbol{\mu}}{1 - 2it}$$

即:

$$\begin{aligned} \sum_{j=1}^t \frac{i\lambda_j t c_j^2}{1 - 2i\lambda_j t} &= \frac{1}{1 - 2it} it\boldsymbol{\mu}^T Q^T \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q \boldsymbol{\mu} \\ \sum_{j=1}^t \frac{\lambda_j c_j^2}{1 - 2i\lambda_j t} &= \frac{1}{1 - 2it} \mathbf{c}^T \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{c} \\ \sum_{j=1}^t \frac{\lambda_j c_j^2}{1 - 2i\lambda_j t} &= \frac{1}{1 - 2it} \sum_{j=1}^t \lambda_j c_j^2 \\ \frac{\lambda_j c_j^2}{1 - 2i\lambda_j t} &= \frac{\lambda_j c_j^2}{1 - 2it}, \quad j = 1, 2, \dots, t \end{aligned}$$

所以 $\lambda_j = 1$ 。 □

Corollary 3.2. 设 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, $\Sigma > 0$, $A \in M_n(K)$ 是一个非随机实对称矩阵, 则 $\mathbf{X}^T A \mathbf{X} \sim \chi_{r, \boldsymbol{\mu}^T A \boldsymbol{\mu}}^2$ 的充分必要条件为 $A \Sigma A = A$, $\text{rank}(A) = r$ 。

Proof. 因为 $\Sigma > 0$, 所以存在 $\Sigma^{-\frac{1}{2}}$ 。考虑随机向量 $\mathbf{Y} = \Sigma^{-\frac{1}{2}} \mathbf{X}$, 由定理 3.3 可知 $\mathbf{Y} \sim N_n(\Sigma^{-\frac{1}{2}} \boldsymbol{\mu}, I_n)$ 。注意到:

$$\mathbf{X}^T A \mathbf{X} = \mathbf{X}^T \Sigma^{-\frac{1}{2}} \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \Sigma^{-\frac{1}{2}} \mathbf{X}$$

由性质 4.7.5(3) 可得:

$$\mathbf{X}^T A \mathbf{X} = \mathbf{X}^T (\Sigma^{-\frac{1}{2}})^T \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \Sigma^{-\frac{1}{2}} \mathbf{X} = \mathbf{Y} \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \mathbf{Y}$$

由定理 3.10 可得 $\mathbf{Y} \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \mathbf{Y} \sim \chi_{r, \boldsymbol{\mu}^T A \boldsymbol{\mu}}^2$ 的充分必要条件为 $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$ 是一个对称阵且:

$$(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}})^2 = \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}, \text{rank}(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}) = r, (\Sigma^{-\frac{1}{2}} \boldsymbol{\mu})^T \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \Sigma^{-\frac{1}{2}} \boldsymbol{\mu} = \boldsymbol{\mu}^T A \boldsymbol{\mu}$$

第三式显然成立。因为 $\Sigma > 0$, 所以 $\text{rank}(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}) = \text{rank}(A)$ 。注意到:

$$\begin{aligned} (\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}})^2 &= \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \Leftrightarrow \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} = \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \\ &\Leftrightarrow \Sigma^{\frac{1}{2}} A \Sigma A \Sigma^{\frac{1}{2}} = \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \Leftrightarrow A \Sigma A = A \end{aligned} \quad \square$$

Theorem 3.11. 设 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, I_n)$, $A \in M_n(K)$ 是一个实对称矩阵, $B \in M_{m \times n}(K)$ 。若 $BA = \mathbf{0}$, 则 $B\mathbf{X}$ 与 $\mathbf{X}^T A \mathbf{X}$ 相互独立。

Proof. 因为 A 是一个实对称矩阵, 由性质 1.6.2(3) 可知存在正交矩阵 Q 使得:

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

其中 $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r)$, $\lambda_i \neq 0$, $i = 1, 2, \dots, r$, $\text{rank}(A) = r$ 。因为 $BA = \mathbf{0}$, 所以有 $BQ Q^T A Q = BAQ = \mathbf{0}$, 于是:

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

设:

$$C = BQ = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

则:

$$BQ \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} C_{11} \Lambda & \mathbf{0} \\ C_{21} \Lambda & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

于是有 $C_{11} = \mathbf{0}$, $C_{21} = \mathbf{0}$ 。对 C 和 Q 做对应分块:

$$C = BQ = \begin{pmatrix} \mathbf{0} & C_1 \end{pmatrix}, Q = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix}$$

于是:

$$B = CQ^T = \begin{pmatrix} \mathbf{0} & C_1 \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix} = C_1 Q_2^T$$

而：

$$A = Q \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^T = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix} = Q_1 \Lambda Q_1^T$$

记 $\mathbf{Y} = Q^T \mathbf{X}$ ，由推论 3.1(2) 可得：

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} Q_1^T \mathbf{X} \\ Q_2^T \mathbf{X} \end{pmatrix} \sim N_n(Q^T \boldsymbol{\mu}, \sigma^2 I_n)$$

由定理 3.7 可知 \mathbf{Y}_1 与 \mathbf{Y}_2 独立。因为：

$$\begin{aligned} B\mathbf{X} &= C_1 Q_2^T \mathbf{X} = C_1 \mathbf{Y}_2 \\ \mathbf{X}^T A \mathbf{X} &= \mathbf{X}^T Q_1 \Lambda Q_1^T \mathbf{X} = \mathbf{Y}_1^T \Lambda \mathbf{Y}_1 \end{aligned}$$

所以 $B\mathbf{X}$ 与 $\mathbf{X}^T A \mathbf{X}$ 独立。 \square

Corollary 3.3. 设 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, $\Sigma > 0$, A 为 n 阶对称阵。若 $C\Sigma A = \mathbf{0}$, 则 $C\mathbf{X}$ 与 $\mathbf{X}^T A \mathbf{X}$ 独立。

Proof. \square

Theorem 3.12. 设 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, I_n)$, A, B 为 n 阶对称阵。若 $AB = \mathbf{0}$, 则 $\mathbf{X}^T A \mathbf{X}$ 与 $\mathbf{X}^T B \mathbf{X}$ 独立。

Proof. 因为 $AB = \mathbf{0}$ 且 A, B 都是对称阵, 所以 $BA = B^T A^T$ \square

3.1.3 矩阵正态分布的定义

密度函数定义

Definition 3.4. 若 $m \times n$ 随机矩阵 \mathbf{X} 满足以下概率密度函数：

$$p(\mathbf{X}) = \frac{1}{(2\pi)^{\frac{mn}{2}} (\det U)^{\frac{n}{2}} (\det V)^{\frac{m}{2}}} e^{-\frac{1}{2} \text{tr}[V^{-1}(X-M)^T U^{-1}(X-M)]}$$

其中, $M \in M_{m \times n}(\mathbb{R})$, $U \in M_m(\mathbb{R})$, $V \in M_n(\mathbb{R})$, $U, V > 0$ 。此时称 \mathbf{X} 服从矩阵正态分布, 记作 $\mathbf{X} \sim MN(M, U, V)$ 。

向量化定义

Definition 3.5. 若随机矩阵 \mathbf{X} 满足 $\text{vec}(\mathbf{X}) \sim N(\text{vec}(M), V \otimes U)$, 其中, $M \in M_{m \times n}(\mathbb{R})$, $U \in M_m(\mathbb{R})$, $V \in M_n(\mathbb{R})$, $U, V \geq 0$ 。此时称 \mathbf{X} 服从矩阵正态分布, 记作 $\mathbf{X} \sim MN(M, U, V)$ 。

Theorem 3.13. 设 \mathbf{X} 是一个 $m \times n$ 随机矩阵, 其行协方差矩阵 U 和列协方差矩阵 V 都是正定矩阵, 则 \mathbf{X} 满足定义 3.4 的充分必要条件为满足定义 3.5。

Proof. 由性质 4.7.2(3)、性质 4.7.3(1)(2)、性质 4.7.1(2)(3) 可得:

$$\begin{aligned}
 \text{tr}[V^{-1}(\mathbf{X} - M)^T U^{-1}(\mathbf{X} - M)] &= \text{tr}[(\mathbf{X} - M)^T U^{-1}(\mathbf{X} - M)V^{-1}] \\
 &= \text{vec}(\mathbf{X} - M)^T \text{vec}[U^{-1}(\mathbf{X} - M)V^{-1}] \\
 &= \text{vec}(\mathbf{X} - M)^T [(V^{-1})^T \otimes U^{-1}] \text{vec}(\mathbf{X} - M) \\
 &= \text{vec}(\mathbf{X} - M)^T [(V^T)^{-1} \otimes U^{-1}] \text{vec}(\mathbf{X} - M) \\
 &= \text{vec}(\mathbf{X} - M)^T (V^{-1} \otimes U^{-1}) \text{vec}(\mathbf{X} - M) \\
 &= [\text{vec}(\mathbf{X}) - \text{vec}(M)]^T (V \otimes U)^{-1} [\text{vec}(\mathbf{X}) - \text{vec}(M)]
 \end{aligned}$$

需要补充证明, 但这里涉及到了 Jordan 标准形, 学完再来补。

因为 $\det(V \otimes U) = (\det V)^m (\det U)^n$, 所以 $(\det U)^{\frac{n}{2}} (\det V)^{\frac{m}{2}}$ 可化作 $[\det(V \otimes U)]^{\frac{1}{2}}$ 。□

Corollary 3.4. 如果正态随机矩阵 $\mathbf{X} \sim MN(M, U, V)$ 中的每个元素都服从标准正态分布, 则 $M = \mathbf{0}$, $V \otimes U = I_{mn}$ 。

由此我们看到, M 就是正态随机矩阵 X 的均值矩阵, 仍然不明确的是 U, V 到底是什么, 只能说 $V \otimes U$ 对应着 X 被向量化后的协方差矩阵, 那就先来研究一下 $\text{Cov}(\mathbf{X}_{ij}, \mathbf{X}_{kl})$ 到底对应着 $V \otimes U$ 中的哪个元素。联想正态随机向量中两个元素的协方差在协方差矩阵中的位置, 我们需要找到 \mathbf{X}_{ij} 和 \mathbf{X}_{kl} 在 $\text{vec}(\mathbf{X})$ 中的索引, 注意到向量化算子 vec 是按列拉直, 那么 \mathbf{X}_{ij} 和 \mathbf{X}_{kl} 分别在 $\text{vec}(\mathbf{X})$ 的第 $(j-1)m + i$ 位和第 $(l-1)m + k$ 位, 于是有:

$$\text{Cov}(\mathbf{X}_{ij}, \mathbf{X}_{kl}) = (V \otimes U)_{(j-1)s+i, (l-1)s+k} = V_{jl} U_{ik}$$

如果 U 是一个对角阵, 那么 $i \neq k$ 时有 $U_{ik} = 0$, 就会导致:

$$\text{Cov}(\mathbf{X}_{ij}, \mathbf{X}_{kl}) = V_{jl} U_{ik} = 0, i \neq k$$

这表明此时只要 X 中的元素处于不同行, 它们就不相关。

如果 V 是一个对角阵, 那么 $j \neq l$ 时有 $V_{jl} = 0$, 就会导致:

$$\text{Cov}(\mathbf{X}_{ij}, \mathbf{X}_{kl}) = V_{jl} U_{ik} = 0, j \neq l$$

这表明此时只要 X 中的元素处于不同列, 它们就不相关。

对于元素 \mathbf{X}_{ij} , 有:

$$\text{Var}(\mathbf{X}_{ij}) = \text{Cov}(\mathbf{X}_{ij}, \mathbf{X}_{ij}) = V_{jj} U_{ii}$$

这表明此时协方差由 $V_{jj} U_{ii}$ 控制。

对于同一行的元素, 有:

$$\text{Cov}(\mathbf{X}_{ij}, \mathbf{X}_{il}) = V_{jl} U_{ii}$$

这表明此时协方差由 V_{jl} 控制。

对于同一列的元素, 有:

$$\text{Cov}(\mathbf{X}_{ij}, \mathbf{X}_{kj}) = V_{jj} U_{ik}$$

这表明此时协方差由 U_{ik} 控制。

线性变换定义

Definition 3.6. \mathbf{X} 为 $m \times n$ 随机矩阵。若存在矩阵 $A \in M_{q \times n}(\mathbb{R})$, $B \in M_{m \times p}(K)$ 使得 $\mathbf{X} = B\mathbf{Y}A^T + M$, 其中 \mathbf{Y} 是一个 $p \times q$ 随机矩阵, $\mathbf{Y}_{ij} \sim N(0, 1)$ 且互相独立, $i = 1, 2, \dots, p$, $j = 1, 2, \dots, q$, $M \in M_{p \times q}(\mathbb{R})$, 则称 \mathbf{X} 服从矩阵正态分布, 记作 $\mathbf{X} \sim MN(M, U, V)$ 。其中, $U = BB^T$, $V = AA^T$ 。

Theorem 3.14. \mathbf{X} 是一个 $m \times n$ 随机矩阵, 则 \mathbf{X} 满足定义 3.5 的充分必要条件是满足定义 3.6。

Proof. (1) 必要性: 设 \mathbf{X} 满足定义 3.6, 由 \mathbf{Y} 的定义可知:

$$\text{vec}(\mathbf{Y}) \sim N_{pq}(\mathbf{0}, I_{pq})$$

由性质 4.7.3(1)(3) 可得:

$$\text{vec}(\mathbf{X}) = \text{vec}(B\mathbf{Y}A^T + M) = \text{vec}(B\mathbf{Y}A^T) + \text{vec}(M) = (A \otimes B) \text{vec}(\mathbf{Y}) + M$$

由性质 2.4.1(3) 和性质 4.7.1(4)(2) 可得:

$$E[\text{vec}(\mathbf{X})] = E[(A \otimes B) \text{vec}(\mathbf{Y}) + M] = (A \otimes B) E[\text{vec}(\mathbf{Y})] + M = M,$$

$$\begin{aligned} \text{Cov}[\text{vec}(\mathbf{X})] &= \text{Cov}[(A \otimes B) \text{vec}(\mathbf{Y})] \\ &= (A \otimes B) \text{Cov}[\text{vec}(\mathbf{Y})] (A \otimes B)^T \\ &= (A \otimes B) I_{pq} (A^T \otimes B^T) \\ &= AA^T \otimes BB^T. \end{aligned}$$

因为 $\text{vec}(\mathbf{X}) = (A \otimes B) \text{vec}(\mathbf{Y}) + M$, 而 $\text{vec}(\mathbf{Y}) \sim N_{pq}(\mathbf{0}, I_{pq})$, 由定理 3.3 可知:

$$\text{vec}(\mathbf{X}) \sim N(\text{vec}(M), AA^T \otimes BB^T)$$

令 $V = AA^T$, $U = BB^T$, 则有 $\text{vec}(\mathbf{X}) \sim N(\text{vec}(M), V \otimes U)$, 即 \mathbf{X} 满足定义 3.5。

(2) 充分性: 设 \mathbf{X} 满足定义 3.5, 因为 $U, V \geq 0$, 所以存在 $U^{\frac{1}{2}}$ 和 $V^{\frac{1}{2}}$, 令 $B = U^{\frac{1}{2}}$, $A = V^{\frac{1}{2}}$, 于是 $\text{vec}(\mathbf{X}) \sim N(\text{vec}(M), V \otimes U)$ 可写作 $\text{vec}(\mathbf{X}) \sim N(\text{vec}(M), AA^T \otimes BB^T)$ 。设 \mathbf{Y} 是一个随机矩阵, 其中的每一个元素都服从标准正态分布且互相独立, 则 $\text{vec}(\mathbf{X}) = (A \otimes B) \text{vec}(\mathbf{Y}) + M$ 。由性质 4.7.3(1)(3) 可知:

$$\text{vec}(\mathbf{X}) = \text{vec}(B\mathbf{Y}A^T + M)$$

于是 $\mathbf{X} = B\mathbf{Y}A^T + M$, 即 \mathbf{X} 满足定义 3.6。 □

3.1.4 矩阵正态分布的性质

Theorem 3.15. 设 \mathbf{X} 为 $m \times n$ 随机矩阵且服从矩阵正态分布 $MN(M, U, V)$, $P \in M_{s \times m}(R)$, $Q \in M_{n \times t}(R)$, 则 $P\mathbf{X}Q^T \sim$

Proof. 由定义 3.6 可知 $\mathbf{X} = \mathbf{B}\mathbf{Y}\mathbf{A}^T + \mathbf{M}$, 于是:

$$\mathbf{P}\mathbf{X}\mathbf{Q}^T = \mathbf{P}\mathbf{B}\mathbf{Y}\mathbf{A}^T\mathbf{Q}^T + \mathbf{P}\mathbf{M}\mathbf{Q}^T$$

此时 $\mathbf{P}\mathbf{B}\mathbf{B}^T\mathbf{P}^T = \mathbf{P}\mathbf{U}\mathbf{P}^T$, $\mathbf{Q}\mathbf{A}\mathbf{A}^T\mathbf{Q}^T = \mathbf{Q}\mathbf{V}\mathbf{Q}^T$, 由定义 3.6 即可得到结论. \square

3.2 χ^2 分布, t 分布和 F 分布

3.2.1 χ^2 分布

Definition 3.7. 设 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, I_n)$, 则随机变量 $\mathbf{Y} = \mathbf{X}^T\mathbf{X}$ 的分布称为自由度为 n 、非中心参数为 $\lambda = \boldsymbol{\mu}^T\boldsymbol{\mu}$ 的 χ^2 分布, 记为 $\mathbf{Y} \sim \chi_{n,\lambda}^2$ 。当 $\lambda = 0$ 时, 称 \mathbf{Y} 的分布为中心 χ^2 分布, 记为 $Y \sim \chi_n^2$ 。

Property 3.2.1. χ^2 分布具有如下性质:

1. 设 $Y_i \sim \chi_{n_i,\lambda_i}^2$, $i = 1, 2, \dots, k$ 相互独立, 则:

$$\sum_{i=1}^k Y_i \sim \chi_{n,\lambda}^2, \quad \text{其中 } n = \sum_{i=1}^k n_i, \lambda = \sum_{i=1}^k \lambda_i$$

2. 设 $Y \sim \chi_{n,\lambda}^2$, 则 $E(Y) = n + \lambda$, $\text{Var}(Y) = 2n + 4\lambda$;

3. 设 $Y \sim \chi_{n,\lambda}^2$, $\mathbf{X} \sim N_n(\boldsymbol{\mu}, I_n)$, $\mathbf{Y} = \mathbf{X}^T\mathbf{X}$, 则:

$$\varphi_{\mathbf{Y}}(t) = (1 - 2it)^{-\frac{n}{2}} \exp\left\{\frac{it\lambda}{1 - 2it}\right\}$$

Proof. (1) 设 $Y_i = \mathbf{X}_i^T\mathbf{X}_i$, 其中 $\mathbf{X}_i \sim N_{n_i}(\boldsymbol{\mu}_i, I_{n_i})$ 。令 $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k)^T$, 则有

$$\sum_{i=1}^k Y_i = \sum_{i=1}^k \mathbf{X}_i^T\mathbf{X}_i = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k)(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k)^T = \mathbf{X}^T\mathbf{X}$$

因为 Y_i 相互独立, 所以 \mathbf{X}_i 也相互独立, 于是 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, I_n)$, 其中:

$$n = \sum_{i=1}^k n_i, \quad \boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_n)^T$$

因此有:

$$\sum_{i=1}^k Y_i \sim \chi_{n,\lambda}^2, \quad \lambda = \boldsymbol{\mu}^T\boldsymbol{\mu} = \sum_{i=1}^k \boldsymbol{\mu}_i^T\boldsymbol{\mu}_i = \sum_{i=1}^k \lambda_i$$

(2) 因为 $Y \sim \chi_{n,\lambda}^2$, 由定义可知 Y 可以表示为:

$$Y = \sum_{i=1}^n X_i^2, \quad X_i \sim N(\mu_i, 1), \quad \sum_{i=1}^n \mu_i^2 = \lambda$$

其中 X_i 相互独立。由性质 2.2.1(1) 可知:

$$E(Y) = E\left(\sum_{i=1}^n X_i^2\right) = \sum_{i=1}^n E(X_i^2) = \sum_{i=1}^n \{\text{Var}(X_i) + [E(X_i)]^2\} = \sum_{i=1}^n (1 + \mu_i^2) = n + \lambda$$

链接独立方差等于和

因为 X_i 相互独立, 由可知:

$$\begin{aligned}\text{Var}(Y) &= \text{Var}\left(\sum_{i=1}^n X_i^2\right) = \sum_{i=1}^n \text{Var}(X_i^2) = \sum_{i=1}^n \{E(X_i^4) - [E(X_i^2)]^2\} \\ &= \sum_{i=1}^n E(X_i^4) - \sum_{i=1}^n [E(X_i^2)]^2\end{aligned}$$

由性质 2.2.1(1) 可知:

$$E(X_i^2) = \text{Var}(X_i) + [E(X_i)]^2 = 1 + \mu_i^2$$

所以:

$$\sum_{i=1}^n [E(X_i^2)]^2 = \sum_{i=1}^n (\mu_i^4 + 2\mu_i^2 + 1) = \sum_{i=1}^n \mu_i^4 + 2 \sum_{i=1}^n \mu_i^2 + n = \sum_{i=1}^n \mu_i^4 + 2\lambda + n$$

而:

$$E(X_i^4) = \mu_i^4 + 6\mu_i^2 + 3$$

于是:

$$\begin{aligned}\text{Var}(Y) &= \sum_{i=1}^n E(X_i^4) - \sum_{i=1}^n [E(X_i^2)]^2 \\ &= \sum_{i=1}^n \mu_i^4 + 6 \sum_{i=1}^n \mu_i^2 + 3n - \sum_{i=1}^n \mu_i^4 - 2\lambda - n \\ &= 6\lambda + 3n - 2\lambda - n = 2n + 4\lambda\end{aligned}$$

(3) 因为 $\mathbf{X} \sim N_n(\boldsymbol{\mu}, I_n)$, 由定理 3.7 可知 \mathbf{X}_i 相互独立, 所以 \mathbf{X}_i^2 相互独立。因为 $\mathbf{Y} = \mathbf{X}^T \mathbf{X} = \sum_{i=1}^n \mathbf{X}_i^2$, 由性质 2.8.1(4) 可知:

$$\varphi_{\mathbf{Y}}(t) = \prod_{i=1}^n \varphi_{\mathbf{X}_i^2}(t)$$

下面来求 $\varphi_{\mathbf{X}_i^2}$ 。

由推论 3.1(5) 可知 $\mathbf{X}_i \sim N(\mu_i, 1)$, 于是:

$$\begin{aligned}\varphi_{\mathbf{X}_i^2}(t) &= E(e^{it\mathbf{X}_i^2}) \\ &= \int_{-\infty}^{+\infty} e^{itx^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left\{-\frac{x^2 - 2\mu_i x + \mu_i^2}{2} + itx^2\right\} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left\{-\frac{x^2}{2}(1 - 2it) + \mu_i x - \frac{\mu_i^2}{2}\right\} dx \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{\mu_i^2}{2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{x^2}{2}(1 - 2it) + \mu_i x\right\} dx\end{aligned}$$

这是一个 Gaussian 积分, 由 Gaussian 积分公式可得:

$$\begin{aligned}\varphi_{\mathbf{X}_i^2}(t) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{\mu_i^2}{2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{x^2}{2}(1-2it) + \mu_i x\right\} dx \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{\mu_i^2}{2}} \sqrt{\frac{2\pi}{1-2it}} e^{\frac{\mu_i^2}{2-4it}} = (1-2it)^{-\frac{1}{2}} \exp\left\{\frac{\mu_i^2}{2-4it} - \frac{\mu_i^2}{2}\right\} \\ &= (1-2it)^{-\frac{1}{2}} \exp\left\{\frac{it\mu_i^2}{1-2it}\right\}\end{aligned}$$

于是:

$$\varphi_{\mathbf{Y}} = \prod_{i=1}^n (1-2it)^{-\frac{1}{2}} \exp\left\{\frac{it\mu_i^2}{1-2it}\right\} = (1-2it)^{-\frac{n}{2}} \exp\left\{\frac{it\lambda}{1-2it}\right\}$$

□

3.2.2 t 分布

Definition 3.8. 设随机变量 $X \sim N(0, 1)$, $Y \sim \chi_n^2$ 且 X 与 Y 独立, 则称:

$$T = \frac{X}{\sqrt{Y/n}}$$

为自由度是 n 的 t 变量, 其分布称为自由度为 n 的 t 分布, 记为 $T \sim t_n$ 。

3.2.3 F 分布

Definition 3.9. 设随机变量 $X \sim \chi_m^2$, $Y \sim \chi_n^2$ 且 X 与 Y 独立, 则称:

$$F = \frac{X/m}{Y/n}$$

为自由度是 m 和 n 的 F 变量, 其分布称为自由度为 m 和 n 的 F 分布, 记为 $F \sim F_{m,n}$ 。

Property 3.2.2. F 分布具有如下性质:

1. 若 $F \sim F_{m,n}$, 则有 $\frac{1}{F} \sim F_{n,m}$;
2. 若 $T \sim t_n$, 则有 $T^2 \sim F_{1,n}$;
3. $F_{m,n}(1-\alpha) = \frac{1}{F_{n,m}(\alpha)}$;

Proof. (1) 由 F 分布的定义直接可得。

(2) 设:

$$T = \frac{X}{\sqrt{Y/n}}$$

其中 $X \sim N(0, 1)$, $Y \sim \chi_n^2$ 且 X 与 Y 独立, 于是:

$$T^2 = \frac{X^2}{Y/n} = \frac{X^2/1}{Y/n}$$

注意到 $X^2 \sim \chi_1^2$ 且有 X^2 与 Y 独立, 由 F 分布的定义即可得到 $T^2 \sim F_{1,n}$ 。

(3) 由分位数的定义:

$$\begin{aligned}
 P[F > F_{m,n}(1-\alpha)] &= 1-\alpha \\
 P\left[\frac{X/m}{Y/n} > F_{m,n}(1-\alpha)\right] &= 1-\alpha \\
 P\left[\frac{Y/n}{X/m} < \frac{1}{F_{m,n}(1-\alpha)}\right] &= 1-\alpha \\
 P\left[\frac{Y/n}{X/m} \geq \frac{1}{F_{m,n}(1-\alpha)}\right] &= \alpha \\
 P\left[\frac{Y/n}{X/m} > \frac{1}{F_{m,n}(1-\alpha)}\right] &= \alpha
 \end{aligned}$$

即:

$$F_{m,n}(1-\alpha) = \frac{1}{F_{n,m}(\alpha)}$$

□

Chapter 4

线性模型

4.1 一般线性模型

Definition 4.1. 称以下模型为线性模型 (*linear model*):

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

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

Definition 4.2. 称方程 $X^T X \beta = X^T y$ 为正则方程 (*normal equation*)。

Theorem 4.1. 对于定义 4.1, $\hat{\beta} = (X^T X)^- X^T y$ 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 1.2 可知方程 $X^T X\beta = X^T y$ 是相容的, 根据定理 1.22 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 4.1. 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 1.44(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

4.1.1 参数估计

回归系数

Definition 4.3. 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 4.1.1. 对于定义 4.1, $c^T \beta$ 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 BLUE。

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) 直接可得。

(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(7) 可得:

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

(6) 因为 $c^T \beta, d^T \beta$ 是可估函数, 由 (1) 可知存在 α, γ 使得 $c = X^T \alpha, d = X^T \gamma$ 。由性质 2.4.1(3) 和性质 1.5.1(6)(7) 可知:

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

在上第三行中把 c^T, d 展开为 $\alpha^T X, X^T \gamma$, 由性质 1.5.1(4) 即可知 $\text{Cov}(c^T \hat{\beta}, d^T \hat{\beta})$ 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 1.5.1(6) 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 1.5.1(7):

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

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 4.4. 对于定义 4.1, 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。

残差

Definition 4.5. 称 $\hat{e} = y - X\hat{\beta}$ 为残差向量 (*residual vector*), 称 $\hat{e}^T \hat{e}$ 为残差平方和 (*sum of squared residuals, SSE*), 记为 SSE。

Property 4.1.2. 对于定义 4.1, $\hat{\beta}$ 为正则方程的解, 则残差向量 \hat{e} 满足:

$$1. E(\hat{e}) = 0, \text{Cov}(\hat{e}) = \sigma^2(I - P_X);$$

$$2. SSE = y^T(I_n - P_X)y.$$

Proof. (1) 由性质 1.8.2(2) 可知向 $\mathcal{M}(X)$ 的正交投影阵 $P_X = X(X^T X)^{-1}X^T$, 根据性质 1.8.2(4) 可知 $I_n - P_X$ 是对称幂等阵, 所以由性质 2.4.1(3) 可得:

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

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

(2) 由 (1) 的证明过程可知:

$$\hat{e} = (I_n - P_X)y$$

且 $I_n - P_X$ 是一个对称幂等阵, 于是:

$$\hat{e}^T \hat{e} = y^T(I_n - P_X)^T(I_n - P_X)y = y^T(I_n - P_X)(I_n - P_X)y = y^T(I_n - P_X)y \quad \square$$

误差方差

Theorem 4.2. 对于定义 4.1, $\hat{\beta}$ 为正则方程的解, $\text{rank}(X) = r$, 则:

$$\hat{\sigma}^2 = \frac{SSE}{n - r}$$

是 σ^2 的无偏估计。

Proof. 注意到 $(I_n - P_X)X = X - X = \mathbf{0}$, 由性质 4.1.2(2) 和定理 2.3 可得:

$$\begin{aligned} E(SSE) &= E[y^T(I_n - P_X)y] = \beta^T X^T(I_n - P_X)X\beta + \text{tr}[(I_n - P_X)\sigma^2 I_n] \\ &= \sigma^2 \text{tr}(I_n - P_X) \end{aligned}$$

由性质 1.8.2(4)、性质 1.8.1(2)(3) 和性质 1.8.2(1) 可得:

$$\text{tr}(I_n - P_X) = \text{rank}(I_n - P_X) = n - \text{rank}(P_X) = n - \text{rank}(X) = n - r$$

即:

$$E\left(\frac{SSE}{n - r}\right) = \sigma^2 \quad \square$$

Definition 4.6. 称 $\hat{\sigma}^2$ 为 σ^2 的 LS 估计。

4.1.2 约束最小二乘估计

Theorem 4.3. 对于定义 4.1, 假设:

$$A\beta = b, \quad A \in M_{k \times p}(K), \quad \text{rank}(A) = k, \quad \mathcal{M}(A^T) \subseteq \mathcal{M}(X^T)$$

且 $A\beta = b$ 相容, 则:

$$\hat{\beta}_A = \hat{\beta} - (X^T X)^- A^T [A(X^T X)^- A^T]^{-1} (A\hat{\beta} - b)$$

为 β 在约束 $A\beta = b$ 下的约束 LS 解, $A\hat{\beta}_A$ 为 $A\beta$ 的约束 LSE。

Proof. 使用 Lagrange 乘子法构造辅助函数 (λ 为 Lagrange 乘子, 乘子前加上系数 2 是为了下面不出现分数, 对结果没有影响):

$$\begin{aligned} F(\beta, \lambda) &= \|y - X\beta\|^2 + 2\lambda^T (A\beta - b) \\ &= y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta + 2\lambda^T A\beta - 2\lambda^T b \end{aligned}$$

于是:

$$\frac{\partial F(\beta, \lambda)}{\partial \beta} = -2X^T y + 2X^T X\beta + 2A^T \lambda$$

令上式为 0, 得到:

$$X^T X\beta = X^T y - A^T \lambda$$

于是约束下的解即为方程组：

$$\begin{cases} X^T X \beta = X^T y - A^T \lambda \\ A \beta = b \end{cases}$$

的解，将其记为 $\hat{\beta}_A, \hat{\lambda}$ 。因为 $\mathcal{M}(A^T) \subseteq \mathcal{M}(X^T)$ ，所以方程组是相容的。由定理 1.22 可知：

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

代入方程组的第二个方程可得：

$$A \hat{\beta} - A (X^T X)^- A^T \hat{\lambda} = b$$

由可知：

$$\hat{\lambda} = [A (X^T X)^- A^T]^{-1} (A \hat{\beta} - b)$$

$A (X^T X)^- A^T$
的可逆性

于是：

$$\hat{\beta}_A = \hat{\beta} - (X^T X)^- A^T [A (X^T X)^- A^T]^{-1} (A \hat{\beta} - b)$$

下证明这个解确实是最小二乘解。

做分解：

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

由性质 1.5.1(5)(6) 可得：

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

于是：

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

由性质 1.5.1(6)(7) 以及 $\mathcal{M}(A^T) \subseteq \mathcal{M}(X^T)$ 可得：

$$\begin{aligned} (\hat{\beta} - \hat{\beta}_A)^T X^T X(\hat{\beta}_A - \beta) &= [(X^T X)^- A^T \hat{\lambda}]^T X^T X(\hat{\beta}_A - \beta) = \hat{\lambda}^T A (X^T X)^- X^T X(\hat{\beta}_A - \beta) \\ &= \hat{\lambda}^T A(\hat{\beta}_A - \beta) = \hat{\lambda}^T (A \hat{\beta}_A - A \beta) = 0 \end{aligned}$$

所以：

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

即对任意满足 $A \beta = b$ 的 β 都有：

$$\|y - X \beta\|^2 \geq \|y - X \hat{\beta}\|^2 + \|X(\hat{\beta} - \hat{\beta}_A)\|^2$$

等号成立当且仅当 $\beta = \hat{\beta}_A$ ，于是 $\hat{\beta}_A$ 是 LSE。

□

误差方差

Theorem 4.4. 在定理 4.3 的假设下, 在参数区域 $A\beta = b$ 上,

$$\hat{\sigma}_A^2 = \frac{\|y - X\hat{\beta}_A\|^2}{n - r + k} = \frac{SSE_A}{n - r + k}$$

是 σ^2 的无偏估计。

Proof. 由定理 4.3 可知:

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

根据定理 4.2 可知:

$$E(\|y - X\hat{\beta}\|^2) = (n - r)\sigma^2$$

由性质 1.5.1(6)(7)、 $\mathcal{M}(A^T) \subseteq \mathcal{M}(X^T)$

$$\begin{aligned} \|X(\hat{\beta} - \hat{\beta}_A)\| &= (\hat{\beta} - \hat{\beta}_A)^T X^T X (\hat{\beta} - \hat{\beta}_A) \\ &= \{(X^T X)^- A^T [A(X^T X)^- A^T]^{-1} (A\hat{\beta} - b)\}^T \\ &\quad \cdot X^T X (X^T X)^- A^T [A(X^T X)^- A^T]^{-1} (A\hat{\beta} - b) \\ &= (A\hat{\beta} - b)^T [A(X^T X)^- A^T]^{-1} A(X^T X)^- \\ &\quad \cdot X^T X (X^T X)^- A^T [A(X^T X)^- A^T]^{-1} (A\hat{\beta} - b) \\ &= (A\hat{\beta} - b)^T [A(X^T X)^- A^T]^{-1} A(X^T X)^- A^T [A(X^T X)^- A^T]^{-1} (A\hat{\beta} - b) \\ &= (A\hat{\beta} - b)^T [A(X^T X)^- A^T]^{-1} (A\hat{\beta} - b) \end{aligned}$$

因为 $\mathcal{M}(A^T) \subseteq \mathcal{M}(X^T)$, 所以由性质 4.1.1(1) 可知 $A\beta$ 的每一个元素都是可估函数, 于是由定理 2.3 和性质 4.1.1(5)(6) 可知:

$$\begin{aligned} E(\|X(\hat{\beta} - \hat{\beta}_A)\|) &= E\{(A\hat{\beta} - b)^T [A(X^T X)^- A^T]^{-1} (A\hat{\beta} - b)\} \\ &= (A\beta - b)^T [A(X^T X)^- A^T]^{-1} (A\beta - b) \\ &\quad + \text{tr}\{[A(X^T X)^- A^T]^{-1} \text{Cov}(A\hat{\beta} - b)\} \\ &= \text{tr}\{[A(X^T X)^- A^T]^{-1} \sigma^2 A(X^T X)^- A^T\} = \sigma^2 \text{tr}(I_k) = k\sigma^2 \end{aligned}$$

所以:

$$E(\|y - X\hat{\beta}_A\|^2) = (n - r + k)\sigma^2$$

即在参数区域 $A\beta = b$ 上,

$$\hat{\sigma}_A^2 = \frac{\|y - X\hat{\beta}_A\|^2}{n - r + k}$$

是 σ^2 的无偏估计。 □

4.1.3 实际计算

Theorem 4.5. 对于无约束条件以及约束 $A\beta = \mathbf{0}$, 有:

$$\begin{aligned} SSE &= \|y - X\hat{\beta}\|^2 = y^T y - \hat{\beta}^T X^T y \\ SSE_A &= \|y - X\hat{\beta}_A\|^2 = y^T y - \hat{\beta}_A^T X^T y \end{aligned}$$

Proof. 由性质 1.5.1(5) 可知:

$$\begin{aligned} SSE &= (y - X\hat{\beta})^T (y - X\hat{\beta}) = y^T y - y^T X\hat{\beta} - \hat{\beta}^T X^T y + \hat{\beta}^T X^T X\hat{\beta} \\ &= y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X(X^T X)^{-1} X^T y = y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T y \\ &= y^T y - \hat{\beta}^T X^T y \end{aligned}$$

由定理 4.3 可知:

$$\begin{cases} X^T X\hat{\beta}_A = X^T y - A^T \hat{\lambda} \\ A\hat{\beta}_A = \mathbf{0} \end{cases}$$

其中 λ 为 Lagrange 乘子, 于是有:

$$\begin{aligned} SSE_A &= (y - X\hat{\beta}_A)^T (y - X\hat{\beta}_A) = y^T y - y^T X\hat{\beta}_A - \hat{\beta}_A^T X^T y + \hat{\beta}_A^T X^T X\hat{\beta}_A \\ &= y^T y - \hat{\beta}_A^T X^T y + \hat{\beta}_A^T X^T X\hat{\beta}_A - \hat{\beta}_A^T X^T y = y^T y - \hat{\beta}_A^T X^T y + \hat{\beta}_A^T (X^T X\hat{\beta}_A - X^T y) \\ &= y^T y - \hat{\beta}_A^T X^T y - \hat{\beta}_A^T A^T \hat{\lambda} = y^T y - \hat{\beta}_A^T X^T y \quad \square \end{aligned}$$

Definition 4.7. 称 $\hat{\beta}^T X^T y$ 为 **回归平方和** (*regression sum of squares, RSS*), 记为 $RSS(\beta)$ 。称 $\hat{\beta}_A^T X^T y$ 为约束条件 $A\beta = \mathbf{0}$ 下的回归平方和, 记为 $RSS_A(\beta)$ 。

note 4.1. 回归平方和表示了数据平方和 $y^T y$ 中能够由因变量 y 与自变量 X_1, X_2, \dots, X_p 的线性关系解释的部分。

4.2 正态线性模型

4.2.1 参数估计

Definition 4.8. 称以下模型为 **正态线性模型** (*normal linear model*):

$$\begin{cases} y = X\beta + \varepsilon \\ \varepsilon \sim N_n(\mathbf{0}, \sigma^2 I_n) \end{cases}$$

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

Property 4.2.1. 对于定义 4.8, 设 $c^T \beta$ 为可估函数, 则:

1. LS 估计 $c^T \hat{\beta}$ 是 $c^T \beta$ 的 MLE , $\hat{\sigma}^2 = \frac{n-r}{n} \hat{\sigma}^2$ 是 σ^2 的 MLE 。若模型在定理 4.3 的约束下, 则 LS 估计 $c^T \hat{\beta}_A$ 是 $c^T \beta$ 的 MLE , $\hat{\sigma}_A^2 = \frac{n-r+k}{n} \hat{\sigma}_A^2$ 是 σ^2 的 MLE ;
2. $c^T \hat{\beta} \sim N[c^T \beta, \sigma^2 c^T (X^T X)^{-1} c]$, $\frac{(n-r)\hat{\sigma}^2}{\sigma^2} = \frac{SSE}{\sigma^2} \sim \chi_{n-r}^2$;
3. $c^T \hat{\beta}$ 与 $\hat{\sigma}^2$ 相互独立;
4. $T_1 = y^T y$, $T_2 = X^T y$ 为完全充分统计量;
5. $c^T \hat{\beta}$ 是 $c^T \beta$ 唯一的 $MVUE$, $\hat{\sigma}^2$ 为 σ^2 唯一的 $MVUE$ 。

Proof. (1) 对于定义 4.8, 其似然函数为:

$$L(\beta, \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}}} \exp \left(-\frac{1}{2\sigma^2} \|y - X\beta\|^2 \right)$$

于是对数似然函数为:

$$\ln L(\beta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \|y - X\beta\|^2$$

固定 σ^2 时, 由最小二乘法原理可知:

$$\|y - X\hat{\beta}\|^2 = \min \|y - X\beta\|^2$$

当 $\beta = \hat{\beta}$ 时有:

$$\ln L(\hat{\beta}, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \|y - X\hat{\beta}\|^2$$

由极值点的必要条件可知:

$$\begin{aligned} \frac{d \ln L(\hat{\beta}, \sigma^2)}{d\sigma^2} &= -\frac{n}{2\sigma^2} + \frac{\|y - X\hat{\beta}\|^2}{2\sigma^4} \\ \hat{\sigma}^2 &= \frac{1}{n} \|y - X\hat{\beta}\|^2 \end{aligned}$$

时对数似然函数取极值, 注意到:

$$\begin{aligned} \frac{d^2 \ln L(\hat{\beta}, \sigma^2)}{d(\sigma^2)^2} &= \frac{n}{2\sigma^4} - \frac{\|y - X\hat{\beta}\|^2}{\sigma^6} \\ \frac{d^2 \ln L(\hat{\beta}, \sigma^2)}{d(\sigma^2)^2} \Big|_{\sigma^2 = \hat{\sigma}^2} &= \frac{n^3}{2\|y - X\hat{\beta}\|^4} - \frac{\|y - X\hat{\beta}\|^2 n^3}{\|y - X\hat{\beta}\|^6} \\ &= \frac{n^3}{2\|y - X\hat{\beta}\|^4} - \frac{n^3}{\|y - X\hat{\beta}\|^4} = -\frac{n^3}{2\|y - X\hat{\beta}\|^4} < 0 \end{aligned}$$

于是此处取极大值。因为:

$$\hat{\sigma}^2 = \frac{1}{n} \|y - X\hat{\beta}\|^2 = \frac{n-r}{n} \hat{\sigma}^2$$

所以 $\hat{\sigma}^2$ 是 σ^2 的 MLE 。由上可知 $\hat{\beta}$ 是 β 的 MLE , 根据??(1) 可得 $c^T \hat{\beta}$ 是 $c^T \beta$ 的 MLE 。

约束条件下的情况与上述证明过程类似。

(2) 因为 $\varepsilon \sim N_n(\mathbf{0}, \sigma^2 I_n)$, 而 $c^T \hat{\beta} = c^T (X^T X)^{-1} X^T y = c^T (X^T X)^{-1} X^T (X\beta + \varepsilon)$ 。因为 $c^T \beta$ 是可估函数, 所以由性质 4.1.1(1) 可知存在 α 使得 $c = X^T \alpha$, 根据性质 1.5.1(7) 可知:

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

由性质 1.5.1(6)(7) 可知:

$$c^T (X^T X)^{-1} X^T [c^T (X^T X)^{-1} X^T]^T = c^T (X^T X)^{-1} X^T X (X^T X)^{-1} c = c^T (X^T X)^{-1} c$$

于是由定理 3.3 可得:

$$c^T \hat{\beta} \sim N[c^T \beta, c^T (X^T X)^{-1} c]$$

因为 $(I_n - P_X)X = \mathbf{0}$, 根据性质 1.8.2(4) 可得 $I_n - P_X$ 是对称阵, 所以由性质 4.1.2(2) 可知:

$$\begin{aligned} \frac{n-r}{\sigma^2} \hat{\sigma}^2 &= \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{\sigma^2} = \frac{y^T (I_n - P_X) y}{\sigma^2} \\ &= \frac{(X\beta + \varepsilon)^T (I_n - P_X) (X\beta + \varepsilon)}{\sigma^2} \\ &= \frac{(X\beta + \varepsilon)^T (I_n - P_X) X\beta + (X\beta + \varepsilon)^T (I_n - P_X) \varepsilon}{\sigma^2} \\ &= \frac{(X\beta + \varepsilon)^T (I_n - P_X) \varepsilon}{\sigma^2} = \frac{\beta^T X^T (I_n - P_X) \varepsilon + \varepsilon^T (I_n - P_X) \varepsilon}{\sigma^2} \\ &= \frac{\beta^T [(I_n - P_X)^T X]^T \varepsilon + \varepsilon^T (I_n - P_X) \varepsilon}{\sigma^2} = \frac{\varepsilon^T (I_n - P_X) \varepsilon}{\sigma^2} \end{aligned}$$

因为 $\varepsilon \sim N_n(\mathbf{0}, \sigma^2 I_n)$, 由定理 3.15 可知 $\frac{\varepsilon}{\sigma} \sim N_n(\mathbf{0}, I_n)$ 。根据性质 1.8.2(4) 可知 $I_n - P_X$ 是对称幂等阵, 由性质 1.8.1(3) 和性质 1.8.2(1) 可得 $\text{rank}(I_n - P_X) = n - r$, 于是根据定理 3.10 可得:

$$\frac{n-r}{\sigma^2} \hat{\sigma}^2 = \frac{\varepsilon^T (I_n - P_X) \varepsilon}{\sigma^2} \sim \chi_{n-r}^2$$

(3) 由性质 4.1.2(2) 可知:

$$c^T \hat{\beta} = c^T (X^T X)^{-1} X^T y, \quad \hat{\sigma}^2 = \frac{y^T (I_n - P_X) y}{n-r}$$

由性质 1.8.2(4) 可知 $I_n - P_X$ 为对称阵, 所以 $\frac{I_n - P_X}{n-r}$ 也是对称阵。因为:

$$\begin{aligned} c^T (X^T X)^{-1} X^T \frac{I_n - P_X}{n-r} &= \frac{1}{n-r} c^T (X^T X)^{-1} X^T (I_n - P_X) \\ &= \frac{1}{n-r} c^T (X^T X)^{-1} [(I_n - P_X) X]^T = \mathbf{0} \end{aligned}$$

由定理 3.11 可知 $c^T \hat{\beta}$ 与 $\hat{\sigma}^2$ 独立。

(4)

□

4.2.2 假设检验

Theorem 4.6. 对于定义 4.8, 假设:

$$A\beta = b, \quad A \in M_{k \times p}(K), \quad \text{rank}(A) = k, \quad \mathcal{M}(A^T) \subseteq \mathcal{M}(X^T)$$

且 $A\beta = b$ 相容, 则:

1. 似然比检验 $H_0: A\beta = b, H_1: A\beta \neq b$ 的似然比为:

$$\lambda(y) = \left(\frac{SSE_A}{SSE} \right)^{\frac{n}{2}}$$

2. $\frac{SSE_A - SSE}{\sigma^2} \sim \chi_{k, \alpha}^2$, 其中:

$$\alpha = \frac{(A\beta - b)^T [A(X^T X)^{-1} A^T]^{-1} (A\beta - b)}{\sigma^2}$$

3. $SSE_A - SSE$ 与 SSE 相互独立;

4. 当 $A\beta = b$ 为真时,

$$F = \frac{(SSE_A - SSE)/k}{SSE/(n-r)} \sim F_{k, n-r}$$

且上式左侧为 $\lambda(y)$ 的单调增函数;

5. 似然比检验 $H_0: A\beta = b, H_1: A\beta \neq b$ 的拒绝域为 $\{F: F > F_{k, n-r}(\alpha)\}$ 。

Proof. (1) 由性质 4.2.1(1) 可知:

$$\begin{aligned} \sup_{\beta, \sigma} L(\beta, \sigma^2; y) &= L(\hat{\beta}, \tilde{\sigma}^2; y) = (2\pi)^{-\frac{n}{2}} (\tilde{\sigma}^2)^{-\frac{n}{2}} \exp \left(-\frac{\|y - X\hat{\beta}\|^2}{2\tilde{\sigma}^2} \right) \\ &= (2\pi)^{-\frac{n}{2}} \left(\frac{\|y - X\hat{\beta}\|^2}{n} \right)^{-\frac{n}{2}} \exp \left(-\frac{n\|y - X\hat{\beta}\|^2}{2\|y - X\hat{\beta}\|^2} \right) \\ &= (2\pi)^{-\frac{n}{2}} \left(\frac{\|y - X\hat{\beta}\|^2}{n} \right)^{-\frac{n}{2}} \exp \left(-\frac{n}{2} \right) = \left(\frac{2\pi e}{n} \right)^{-\frac{n}{2}} \|y - X\hat{\beta}\|^{-n} \\ \sup_{A\beta=b, \sigma^2} L(\beta, \sigma^2; y) &= L(\hat{\beta}_A, \tilde{\sigma}_A^2; y) = \left(\frac{2\pi e}{n} \right)^{-\frac{n}{2}} \|y - X\hat{\beta}_A\|^{-n} \end{aligned}$$

于是:

$$\lambda(y) = \frac{L(\hat{\beta}, \tilde{\sigma}^2; y)}{L(\hat{\beta}_A, \tilde{\sigma}_A^2; y)} = \left(\frac{SSE_A}{SSE} \right)^{\frac{n}{2}}$$

(2) 根据性质 1.5.1(6)(7) 和性质 1.8.2(4) 以及 $(I_n - P_X)X = \mathbf{0}$ 对 SSE_A 作分解:

$$\begin{aligned}
 SSE_A &= \|y - X\hat{\beta}_A\|^2 = \left\| y - X\{\hat{\beta} - (X^T X)^{-1} A^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - b)\} \right\|^2 \\
 &= \left\| y - X\hat{\beta} + X(X^T X)^{-1} A^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - b) \right\|^2 \\
 &= \|y - X\hat{\beta}\|^2 + 2(y - X\hat{\beta})^T X(X^T X)^{-1} A^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - b) \\
 &\quad + \{X(X^T X)^{-1} A^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - b)\}^T \\
 &\quad \cdot X(X^T X)^{-1} A^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - b) \\
 &= \|y - X\hat{\beta}\|^2 + 2[y - X(X^T X)^{-1} X^T y]^T X(X^T X)^{-1} A^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - b) \\
 &\quad + (A\hat{\beta} - b)^T [A(X^T X)^{-1} A^T]^{-1} A(X^T X)^{-1} X^T \\
 &\quad \cdot X(X^T X)^{-1} A^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - b) \\
 &= \|y - X\hat{\beta}\|^2 + 2[(I_n - P_X)y]^T X(X^T X)^{-1} A^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - b) \\
 &\quad + (A\hat{\beta} - b)^T [A(X^T X)^{-1} A^T]^{-1} A(X^T X)^{-1} A^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - b) \\
 &= \|y - X\hat{\beta}\|^2 + 2y^T (I_n - P_X) X(X^T X)^{-1} A^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - b) \\
 &\quad + (A\hat{\beta} - b)^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - b) \\
 &= \|y - X\hat{\beta}\|^2 + (A\hat{\beta} - b)^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - b)
 \end{aligned}$$

所以有:

$$SSE_A - SSE = (A\hat{\beta} - b)^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - b)$$

由性质 4.2.1(2) 可知:

$$A\hat{\beta} - b \sim N_k[A\beta - b, \sigma^2 A(X^T X)^{-1} A^T]$$

根据可知 $A(X^T X)^{-1} A^T$ 存在平方根阵及平方根阵的逆矩阵, 由推论 3.1(1) 可知:

$$\frac{[A(X^T X)^{-1} A^T]^{-\frac{1}{2}}}{\sigma} (A\hat{\beta} - b) \sim N_k\{[A(X^T X)^{-1} A^T]^{-\frac{1}{2}} (A\beta - b), I_n\}$$

$A(X^T X)^{-1} A^T$
的正定性

于是由 χ^2 分布的定义可得:

$$\frac{SSE_A - SSE}{\sigma^2} = \frac{\{[A(X^T X)^{-1} A^T]^{-\frac{1}{2}} (A\hat{\beta} - b)\}^T [A(X^T X)^{-1} A^T]^{-\frac{1}{2}} (A\hat{\beta} - b)}{\sigma^2} \sim \chi_{k, \alpha}^2$$

其中:

$$\begin{aligned}
 \alpha &= \{[A(X^T X)^{-1} A^T]^{-\frac{1}{2}} (A\beta - b)\}^T [A(X^T X)^{-1} A^T]^{-\frac{1}{2}} (A\beta - b) \\
 &= (A\beta - b)^T [A(X^T X)^{-1} A^T]^{-1} (A\beta - b)
 \end{aligned}$$

(3) 由性质 4.1.2(2) 可知 $SSE = y^T (I_n - P_X) y$, 由 (2) 的过程可得:

$$\begin{aligned}
 SSE_A - SSE &= (A\hat{\beta} - b)^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - b) \\
 &= [A(X^T X)^{-1} X^T y - b]^T [A(X^T X)^{-1} A^T]^{-1} [A(X^T X)^{-1} X^T y - b] \\
 &= y^T X(X^T X)^{-1} A^T [A(X^T X)^{-1} A^T]^{-1} A(X^T X)^{-1} X^T y \\
 &\quad - 2b^T [A(X^T X)^{-1} A^T]^{-1} A(X^T X)^{-1} X^T y + b^T [A(X^T X)^{-1} A^T]^{-1} b
 \end{aligned}$$

因为 $(I_n - P_X)X = \mathbf{0}$, 由性质 1.8.2(4) 可得:

$$\begin{aligned} (I_n - P_X)X(X^T X)^{-1}A^T[A(X^T X)^{-1}A^T]^{-1}A(X^T X)^{-1}X^T &= \mathbf{0} \\ 2b^T[A(X^T X)^{-1}A^T]^{-1}A(X^T X)^{-1}X^T(I_n - P_X) & \\ = 2b^T[A(X^T X)^{-1}A^T]^{-1}A(X^T X)^{-1}[(I_n - P_X)X]^T &= \mathbf{0} \end{aligned}$$

于是根据定理 3.11 和定理 3.12 可知 $SSE_A - SSE$ 与 SSE 独立。

(4) 当 $A\beta = b$ 为真时, 由 (2) 可知 $\frac{SSE_A - SSE}{\sigma^2} \sim \chi_k^2$ 。根据 (3) 和性质 4.2.1(2) 可得:

$$\frac{(SSE_A - SSE)/(k\sigma^2)}{SSE/[(n-r)\sigma^2]} = \frac{(SSE_A - SSE)/k}{SSE/(n-r)} \sim F_{k, n-r} \quad \square$$

由 (1) 可得:

$$\frac{(SSE_A - SSE)/k}{SSE/(n-r)} = \frac{n-r}{k} [\lambda^{\frac{2}{n}}(y) - 1]$$

所以它是 $\lambda(y)$ 的单调增函数。

(5) 由 (1)(4) 可立即得出。

4.2.3 置信域

置信椭圆

Theorem 4.7. 对于定义 4.8, 若不能接受假设 $A\beta = \mathbf{0}$, 则 $A\beta$ 置信度为 $1 - \alpha$ 的置信椭圆为:

$$\{A\beta : (A\beta - A\hat{\beta})^T[A(X^T X)^{-1}A^T]^{-1}(A\beta - A\hat{\beta}) \leq k\hat{\sigma}^2 F_{k, n-r}(\alpha)\}$$

Proof. 由性质 4.2.1(2) 可知:

$$A\hat{\beta} \sim N[A\beta, \sigma^2 A(X^T X)^{-1}A^T]$$

所以:

$$\frac{A\hat{\beta} - A\beta}{\sigma} \sim N[\mathbf{0}, A(X^T X)^{-1}A^T]$$

因为:

$$[A(X^T X)^{-1}A^T]^{-1}A(X^T X)^{-1}A^T[A(X^T X)^{-1}A^T]^{-1} = [A(X^T X)^{-1}A^T]^{-1}$$

所以由推论 3.2 可知:

$$\frac{(A\hat{\beta} - A\beta)^T[A(X^T X)^{-1}A^T]^{-1}(A\hat{\beta} - A\beta)}{\sigma^2} \sim \chi_k^2$$

根据性质 4.2.1(2)(3) 可得:

$$\begin{aligned} & \frac{(A\hat{\beta} - A\beta)^T[A(X^T X)^{-1}A^T]^{-1}(A\hat{\beta} - A\beta)}{k\sigma^2} \bigg/ \frac{(n-r)\hat{\sigma}^2}{(n-r)\sigma^2} \\ &= \frac{(A\hat{\beta} - A\beta)^T[A(X^T X)^{-1}A^T]^{-1}(A\hat{\beta} - A\beta)}{k\hat{\sigma}^2} \sim F_{k, n-r} \end{aligned}$$

所以对任意的 $0 < \alpha < 1$, 有:

$$P \left[\frac{(A\hat{\beta} - A\beta)^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - A\beta)}{k\hat{\sigma}^2} \leq F_{k, n-r}(\alpha) \right] = 1 - \alpha$$

即 $A\beta$ 置信度为 $1 - \alpha$ 的置信椭球为:

$$\{A\beta : (A\beta - A\hat{\beta})^T [A(X^T X)^{-1} A^T]^{-1} (A\beta - A\hat{\beta}) \leq k\hat{\sigma}^2 F_{k, n-r}(\alpha)\} \quad \square$$

note 4.2. 这里其实是一个未知方差构造 F 分布的思想。

4.3 误差协方差推广

在很多情况下线性模型误差的协方差矩阵都不是 $\sigma^2 I_n$ 的形式。

4.3.1 广义最小二乘估计

Definition 4.9. 称以下模型为广义线性模型 (*generalized linear model*):

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

其中 y 为 $n \times 1$ 观测向量, X 为 $n \times p$ 设计矩阵, β 为 $p \times 1$ 未知参数向量, ε 为随机误差, $\sigma^2 \Sigma$ 为误差协方差矩阵且 $\Sigma > 0$ 。

推导 4.2. 因为 $\Sigma > 0$, 所以存在 $\Sigma^{-\frac{1}{2}}$ 。令:

$$y^* = \Sigma^{-\frac{1}{2}} y, \quad X^* = \Sigma^{-\frac{1}{2}} X, \quad \varepsilon^* = \Sigma^{-\frac{1}{2}} \varepsilon$$

由性质 2.4.1(3) 可知定义 4.9 可化作:

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

于是我们可以将广义线性模型化作线性模型来处理, 对于线性模型与正态线性模型的那些结论, 广义线性模型也可类似得到。

4.3.2 最小二乘统一理论

4.4 似然比检验

假设

$$H_0 : \theta \in \Theta_1 \quad H_1 : \theta \notin \Theta_1$$

原理

Definition 4.10. 称：

$$\lambda(y) = \frac{\sup_{\theta \in \Theta} L(\theta; y)}{\sup_{\theta \in \Theta_1} L(\theta; y)}$$

为，其中 $L(\theta; y)$ 为似然函数。

推导 4.3. 若 $\lambda(y)$ 较大，则说明 $\theta \in \Theta_1$ 时出现数据 y 的可能性较小，于是拒绝域应形如 $\{y : \lambda(y) \geq c\}$ 。可以寻找统计量 $T(y)$ ，它是 $\lambda(y)$ 的单调增函数，于是检验的拒绝域可取为 $\{y : T(y) \geq c\}$ 。

附录

4.5 数域

Definition 4.11. 复数集的一个子集 K 如果满足:

1. $1 \in K$;

2. $a, b \in K \Rightarrow a \pm b, ab \in K$, $a, b \in K$ 且 $b \neq 0 \Rightarrow \frac{a}{b} \in K$ 。

则称 K 是一个数域 (*number field*)。

Theorem 4.8. 任意数域都包含有理数域。

Proof. 任取数域 K , 由数域的定义可得 $1 \in K$, 那么根据数域对其内数加法与减法的封闭性, $\mathbb{Z} \subseteq K$ 。于是任意分数 $\frac{a}{b} \in K$, $b \neq 0$, 即 $\mathbb{Q} \subseteq K$ 。 \square

4.6 等价关系

Definition 4.12. 对于任意两个非空集合 S, M , 称:

$$\{(a, b) : a \in S, b \in M\}$$

为集合 S 与集合 M 的笛卡尔积 (*Cartesian product*), 记为 $S \times M$ 。其中两个元素 (a_1, b_1) 与 (a_2, b_2) 如果满足 $a_1 = a_2, b_1 = b_2$, 则称二者相等, 记作 $(a_1, b_1) = (a_2, b_2)$ 。

Definition 4.13. 设 S 是一个非空集合, 把 $S \times S$ 的一个子集 W 叫作 S 上的一个二元关系 (*binary relation*)。如果 $(a, b) \in W$, 则称 a 与 b 有 W 关系; 如果 $(a, b) \notin W$, 则称 a 与 b 没有 W 关系。当 a 与 b 有 W 关系时, 记作 aWb , 或 $a \sim b$ 。

Definition 4.14. 集合 S 上的一个二元关系 \sim 如果具有如下性质: 对 $\forall a, b, c \in S$, 有:

1. $a \sim a$ (反身性);

2. $a \sim b \Rightarrow b \sim a$ (对称性);

3. $a \sim b, b \sim c \Rightarrow a \sim c$ (传递性)。

那么称 \sim 是集合 S 上的一个等价关系 (*equivalence relationship*)。

Definition 4.15. 设 \sim 是集合 S 上的一个等价关系, $a \in S$, 令:

$$\bar{a} \stackrel{\text{def}}{=} \{x \in S : x \sim a\}$$

称 \bar{a} 是由 a 确定的等价类 (equivalence class), 称 a 是等价类 \bar{a} 的一个代表。

Property 4.6.1. 等价类具有如下基本性质:

1. $a \in \bar{a}$;
2. $x \in \bar{a} \Leftrightarrow x \sim a$;
3. $x \sim y \Leftrightarrow \bar{x} = \bar{y}$ 。

Proof. (1) 等价关系具有反身性。(2) 由等价类的定义可直接得出。

(3) 充分性: 因为 $x \in \bar{x}$ 且 $\bar{x} = \bar{y}$, 所以 $x \in \bar{y}$, 由 \bar{y} 的定义, $x \sim y$ 。

必要性: 任取 $a \in \bar{x}$, 则 $a \sim x$ 。因为 $x \sim y$, 由等价关系的传递性, $a \sim y$, 即 $a \in \bar{y}$ 。由 a 的任意性, $\bar{x} \subseteq \bar{y}$ 。同理可证得 $\bar{y} \subseteq \bar{x}$, 所以 $\bar{x} = \bar{y}$ 。□

Corollary 4.1. 用不同代表表示的等价类是一样的, 即代表的选择与等价类本身无关。

Proof. 由等价类基本性质 (3) 可直接得到。□

Theorem 4.9. 设 \sim 是集合 S 上的一个等价关系。对 $\forall a, b \in S$, 有 $\bar{a} = \bar{b}$ 或 $\bar{a} \cap \bar{b} = \emptyset$ 。

Proof. 如果 $\bar{a} \neq \bar{b}$, 假设此时 $\bar{a} \cap \bar{b} \neq \emptyset$, 取 $c \in \bar{a} \cap \bar{b}$, 则有 $c \sim a$ 且 $c \sim b$ 。由等价关系的对称性与传递性可得 $a \sim b$, 根据等价类的基本性质 (3), 此时应有 $\bar{a} = \bar{b}$, 矛盾, 所以 $\bar{a} \cap \bar{b} = \emptyset$ 。□

Definition 4.16. 如果集合 S 可以表示为一些非空子集的并集, 且这些子集不相交, 即:

$$\exists S_i \subseteq S, \bigcup_{i \in I} S_i = S, S_i \cap S_j = \emptyset, i \neq j, i, j \in I$$

其中 I 是指标集。称集合 $\{S_i : i \in I\}$ 是 S 的一个划分 (partition), 记作 $\pi(S)$ 。

Theorem 4.10. 设 \sim 是集合 S 上的一个等价关系, 则所有等价类组成的集合是 S 的一个划分, 记作 $\pi_{\sim}(S)$ 。

Proof. 对 $\forall a_i \in S$, 其中 i 是指标集, 有 $a_i \in \bar{a}_i$, 于是 $S = \bigcup_{i \in I} \bar{a}_i$ 。由定理 4.9 可得, 若 $i \neq j$, $\bar{a}_i \cap \bar{a}_j = \emptyset$, 从而所有等价类组成的集合是 S 的一个划分。□

Definition 4.17. 设 \sim 是集合 S 上的一个等价关系, 所有等价类组成的集合称为 S 对于关系 \sim 的商集 (quotient set)。

Definition 4.18. 设 \sim 是集合 S 上的一个等价关系, 一种量或者一种表达式如果对于同一个等价类里的元素是相等的, 那么称这种量或表达式是一个不变量; 恰好能完全绝对等价类的一组不变量称为完全不变量。

4.7 矩阵

4.7.1 Kronecker 乘积

Definition 4.19. 给定两个矩阵 $A = (a_{ij}) \in M_{m \times n}(K)$ 和 $B \in M_{p \times q}(K)$, 它们的 *Kronecker* 乘积 $A \otimes B$ 是一个大小为 $mp \times nq$ 的矩阵, 定义为:

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}$$

$a_{ij}B$ 表示矩阵 B 乘以标量 a_{ij} , $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$ 。

Property 4.7.1. *Kronecker* 乘积具有如下性质:

1. $I_m \otimes I_n = I_{mn}$;
2. 设 $A \in M_{m \times n}(K)$, $B \in M_{p \times q}(K)$, $C \in M_{n \times k}(K)$, $D \in M_{q \times r}(K)$, 则 $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$;
3. 设 $A \in M_m(K)$, $B \in M_n(K)$, 则 $A \otimes B$ 可逆的充分必要条件为 A, B 都可逆, 此时有:

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

4. 设 $A = (a_{ij}) \in M_{m \times n}(K)$, $B = (b_{kl}) \in M_{p \times q}(K)$, 则 $(A \otimes B)^T = A^T \otimes B^T$;

Proof. (1) 由 *Kronecker* 乘积的定义:

$$I_m \otimes I_n = \begin{pmatrix} I_n & 0 & \cdots & 0 \\ 0 & I_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_n \end{pmatrix} = I_{mn}$$

(2) 由 Kronecker 乘积的定义:

$$\begin{aligned}
 (A \otimes B)(C \otimes D) &= \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix} \begin{pmatrix} c_{11}D & c_{12}D & \cdots & c_{1k}D \\ c_{21}D & c_{22}D & \cdots & c_{2k}D \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1}D & c_{n2}D & \cdots & c_{nk}D \end{pmatrix} \\
 &= \begin{pmatrix} \sum_{i=1}^n a_{1i}c_{i1}BD & \sum_{i=1}^n a_{1i}c_{i2}BD & \cdots & \sum_{i=1}^n a_{1i}c_{ik}BD \\ \sum_{i=1}^n a_{2i}c_{i1}BD & \sum_{i=1}^n a_{2i}c_{i2}BD & \cdots & \sum_{i=1}^n a_{2i}c_{ik}BD \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n a_{mi}c_{i1}BD & \sum_{i=1}^n a_{mi}c_{i2}BD & \cdots & \sum_{i=1}^n a_{mi}c_{ik}BD \end{pmatrix} \\
 &= (AC) \otimes (BD)
 \end{aligned}$$

(3) 必要性: 假设此时 A 不可逆, 则存在非零向量 x 使得 $Ax = \mathbf{0}$ 。取非零向量 y , 则 $(x \otimes y)$ 不是一个零向量。由 (2) 可得:

$$(A \otimes B)(x \otimes y) = (Ax) \otimes (By) = \mathbf{0} \otimes (By) = \mathbf{0}$$

因为 $A \otimes B$ 可逆, 所以不存在非零向量 z 使得 $(A \otimes B)z = \mathbf{0}$, 但此时有 $(A \otimes B)(x \otimes y) = \mathbf{0}$, 矛盾, 所以 A 可逆。同理可得 B 可逆。

充分性: 由 (1)(2) 可得:

$$\begin{aligned}
 (A \otimes B)(A^{-1} \otimes B^{-1}) &= (AA^{-1}) \otimes (BB^{-1}) = I_m \otimes I_n = I_{mn} \\
 (A^{-1} \otimes B^{-1})(A \otimes B) &= (A^{-1}A) \otimes (B^{-1}B) = I_m \otimes I_n = I_{mn}
 \end{aligned}$$

所以 $A \otimes B$ 可逆, 逆矩阵就是 $(A^{-1} \otimes B^{-1})$ 。

(4) 对任意的 $1 \leq i \leq m, 1 \leq j \leq n, 1 \leq k \leq p, 1 \leq l \leq q$, 元素 $a_{ij}b_{kl}$ 在 $A \otimes B$ 中的行标为 $(i-1)p+k$, 列标为 $(j-1)q+l$ 。于是 $a_{ij}b_{kl}$ 在 $(A \otimes B)^T$ 中的列标为 $(i-1)p+k$, 行标为 $(j-1)q+l$ 。而 $A^T \otimes B^T$ 列标为 $(i-1)p+k$ 、行标为 $(j-1)q+l$ 的元素为 $a_{ij}b_{kl}$, 于是 $(A \otimes B)^T = A^T \otimes B^T$ 。□

4.7.2 迹

Definition 4.20. $A = (a_{ij}) \in M_n(K)$ 的主对角线上的元素之和称为 A 的迹 (trace), 记作 $\text{tr}(A)$, 即:

$$\text{tr}(A) = \sum_{i=1}^n a_{ii}$$

Property 4.7.2. 设 $A, B \in M_n(K), k \in K$, 矩阵的迹具有如下性质:

$$I. \text{tr}(A + B) = \text{tr}(A) + \text{tr}(B);$$

$$2. \operatorname{tr}(kA) = k \operatorname{tr}(A);$$

$$3. \operatorname{tr}(AB) = \operatorname{tr}(BA).$$

Proof. (1)(2) 是显然的;

(3) 显然:

$$\operatorname{tr}(AB) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} b_{ji} = \sum_{j=1}^n \sum_{i=1}^n b_{ji} a_{ij} = \operatorname{tr}(BA)$$

□

4.7.3 向量化算子

Definition 4.21. 设 $A = (a_{ij}) \in M_{s \times m}(K)$, 则:

$$\operatorname{vec}(A) = (a_{11}, a_{21}, \dots, a_{s1}, a_{12}, a_{22}, \dots, a_{s2}, \dots, a_{1m}, a_{2m}, \dots, a_{sm})^T$$

$$\operatorname{rvec}(A) = (a_{11}, a_{12}, \dots, a_{1m}, a_{21}, a_{22}, \dots, a_{2m}, \dots, a_{s1}, a_{s2}, \dots, a_{sm})$$

称 vec 与 rvec 为向量化算子。

交换矩阵的定义及其性质

Definition 4.22. 对 $A \in M_{s \times m}(K)$, 存在唯一的 $K_{sm} \in M_{sm \times sm}(K)$, 使得:

$$K_{sm} \operatorname{vec}(A) = \operatorname{vec}(A^T)$$

称 K_{sm} 为交换矩阵 (*commutation matrix*)。

向量化算子的性质

Property 4.7.3. 向量化算子具有如下性质:

1. 设 $A, B \in M_{s \times m}(K)$, 则 $\forall k_1, k_2 \in K, \operatorname{vec}(k_1 A + k_2 B) = k_1 \operatorname{vec}(A) + k_2 \operatorname{vec}(B)$, 即向量化算子是线性算子;

2. 设 $A, B \in M_{s \times m}(K)$, 则 $\operatorname{tr}(A^T B) = \operatorname{vec}(A)^T \operatorname{vec}(B)$;

3. 设 $A \in M_{s \times m}(K)$, $B \in M_{m \times n}(K)$, $C \in M_{n \times p}(K)$, $\operatorname{vec}(ABC) = (C^T \otimes A) \operatorname{vec}(B)$;

Proof. (1) 是显然的;

(2) 因为:

$$\operatorname{tr}(A^T B) = \sum_{i=1}^m \sum_{j=1}^s a_{ji} b_{ji}$$

所以:

$$\begin{aligned}
 \text{vec}(A)^T \text{vec}(B) &= (a_{11}, a_{21}, \dots, a_{s1}, a_{12}, a_{22}, \dots, a_{s2}, \dots, a_{1m}, a_{2m}, \dots, a_{sm}) \\
 &\quad \cdot (b_{11}, b_{21}, \dots, b_{s1}, b_{12}, b_{22}, \dots, b_{s2}, \dots, b_{1m}, b_{2m}, \dots, b_{sm})^T \\
 &= \sum_{i=1}^m \sum_{j=1}^s a_{ji} b_{ji} \\
 &= \text{tr}(A^T B)
 \end{aligned}$$

(3) 设 $B = (B_1, B_2, \dots, B_n)$, $C = (C_1, C_2, \dots, C_p)$, 则 ABC 的第 k 列:

$$\begin{aligned}
 ABC[:, k] &= A(B_1, B_2, \dots, B_n)C_k = A \sum_{i=1}^n B_i c_{ik} \\
 &= (c_{1k}A, c_{2k}A, \dots, c_{nk}A) \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{pmatrix} = (C_k^T \otimes A) \text{vec}(B)
 \end{aligned}$$

于是:

$$\begin{aligned}
 \text{vec}(ABC) &= \begin{pmatrix} ABC[:, 1] \\ ABC[:, 2] \\ \vdots \\ ABC[:, p] \end{pmatrix} = \begin{pmatrix} (C_1^T \otimes A) \text{vec}(B) \\ (C_2^T \otimes A) \text{vec}(B) \\ \vdots \\ (C_p^T \otimes A) \text{vec}(B) \end{pmatrix} \\
 &= \begin{pmatrix} C_1^T \otimes A \\ C_2^T \otimes A \\ \vdots \\ C_p^T \otimes A \end{pmatrix} \text{vec}(B) = (C^T \otimes A) \text{vec}(B)
 \end{aligned}$$

□

4.7.4 平方根阵

Definition 4.23. 设对称阵 $A \in M_n(\mathbb{R})$, 其特征值记为 λ_i , $i = 1, 2, \dots, n$ 。因为 A 是一个实对称阵, 由??(3) 可知存在正交矩阵 Q 使得 $A = Q^T \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}Q$ 。若 $A \geq 0$, 由定理 1.45(3) 的第五条可知 $\lambda_i \geq 0$, $i = 1, 2, \dots, n$, 记:

$$A^{\frac{1}{2}} = \text{diag}\{\lambda_1^{\frac{1}{2}}, \lambda_2^{\frac{1}{2}}, \dots, \lambda_n^{\frac{1}{2}}\}$$

称:

$$A^{\frac{1}{2}} = Q^T A^{\frac{1}{2}} Q$$

为 A 的平方根阵。

Property 4.7.4. 设对称阵 $A \in M_n(\mathbb{R})$ 且 $A \geq 0$, $A^{\frac{1}{2}}$ 具有如下性质:

1. $(A^{\frac{1}{2}})^2 = A$;

2. $A^{\frac{1}{2}} \geq 0$;

3. $A^{\frac{1}{2}}$ 是对称阵;

Proof. 由 $A^{\frac{1}{2}}$ 的定义, 有 $A^{\frac{1}{2}} = Q^T \Lambda^{\frac{1}{2}} Q$ 和 $A = Q^T \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} Q$, 其中 Q 是一个正交矩阵, $\Lambda^{\frac{1}{2}} = \text{diag}\{\lambda_1^{\frac{1}{2}}, \lambda_2^{\frac{1}{2}}, \dots, \lambda_n^{\frac{1}{2}}\}$, $\lambda_i, i = 1, 2, \dots, n$ 为 A 的特征值.

(1) 显然:

$$(A^{\frac{1}{2}})^2 = Q^T \Lambda^{\frac{1}{2}} Q Q^T \Lambda^{\frac{1}{2}} Q = Q^T (\Lambda^{\frac{1}{2}})^2 Q = Q^T \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} Q = A$$

(2) 因为 Q 是正交矩阵, 所以 $Q^T = Q^{-1}$. 于是:

$$Q A^{\frac{1}{2}} Q^{-1} = \Lambda^{\frac{1}{2}}$$

由定理 1.27 的必要性可知, $\lambda_i^{\frac{1}{2}}$ 为 $A^{\frac{1}{2}}$ 的特征值, 而 $\lambda_i \geq 0$, 由定理 1.45(3) 的第五条可知 $A^{\frac{1}{2}} \geq 0$.

(3) 显然:

$$(A^{\frac{1}{2}})^T = (Q^T \Lambda^{\frac{1}{2}} Q)^T = Q^T (\Lambda^{\frac{1}{2}})^T Q = Q^T \Lambda^{\frac{1}{2}} Q = A^{\frac{1}{2}}$$

□

平方根阵的逆矩阵

Theorem 4.11. 设对称阵 $A \in M_n(\mathbb{R})$, 其特征值记为 $\lambda_i, i = 1, 2, \dots, n$. 因为 A 是一个实对称阵, 由 ??(3) 可知存在正交矩阵 Q 使得 $A = Q^T \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} Q$. 若 $A > 0$, 由定理 1.44(3) 的第五条可知 $\lambda_i > 0, i = 1, 2, \dots, n$, 记:

$$\Lambda^{-\frac{1}{2}} = \text{diag}\{\lambda_1^{-\frac{1}{2}}, \lambda_2^{-\frac{1}{2}}, \dots, \lambda_n^{-\frac{1}{2}}\}$$

则:

$$A^{-\frac{1}{2}} = Q^T \Lambda^{-\frac{1}{2}} Q$$

是 $A^{\frac{1}{2}}$ 的逆矩阵。

Proof. 显然:

$$A^{\frac{1}{2}} A^{-\frac{1}{2}} = Q^T \text{diag}\{\lambda_1^{\frac{1}{2}}, \lambda_2^{\frac{1}{2}}, \dots, \lambda_n^{\frac{1}{2}}\} Q Q^T \text{diag}\{\lambda_1^{-\frac{1}{2}}, \lambda_2^{-\frac{1}{2}}, \dots, \lambda_n^{-\frac{1}{2}}\} Q = I \quad \square$$

Property 4.7.5. 设对称阵 $A \in M_n(\mathbb{R})$ 且 $A > 0$, $A^{\frac{1}{2}}$ 具有如下性质:

1. $(A^{-\frac{1}{2}})^2 = A^{-1}$;

2. $A^{-\frac{1}{2}} > 0$;

3. $A^{-\frac{1}{2}}$ 是对称阵;

Proof. 由 $A^{-\frac{1}{2}}$ 的定义, 有 $A^{-\frac{1}{2}} = Q^T \Lambda^{-\frac{1}{2}} Q$ 和 $A = Q^T \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} Q$, 其中 Q 是一个正交矩阵, $\Lambda^{-\frac{1}{2}} = \text{diag}\{\lambda_1^{-\frac{1}{2}}, \lambda_2^{-\frac{1}{2}}, \dots, \lambda_n^{-\frac{1}{2}}\}$, $\lambda_i, i = 1, 2, \dots, n$ 为 A 的特征值。

(1) 显然:

$$(A^{-\frac{1}{2}})^2 = Q^T \Lambda^{-\frac{1}{2}} Q Q^T \Lambda^{-\frac{1}{2}} Q = Q^T \text{diag}\{\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_n^{-1}\} Q$$

而:

$$\begin{aligned} A^{-1} &= (Q^T \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} Q)^{-1} = Q^{-1} \text{diag}\{\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_n^{-1}\} (Q^T)^{-1} \\ &= Q^T \text{diag}\{\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_n^{-1}\} Q \end{aligned}$$

所以 $(A^{-\frac{1}{2}})^2 = A^{-1}$ 。

(2) 因为 Q 是正交矩阵, 所以 $Q^T = Q^{-1}$ 。于是:

$$Q A^{-\frac{1}{2}} Q^{-1} = \Lambda^{-\frac{1}{2}}$$

由定理 1.27 的必要性可知, $\lambda_i^{-\frac{1}{2}}$ 为 $A^{-\frac{1}{2}}$ 的特征值, 而 $\lambda_i > 0$, 由定理 1.44(3) 的第五条可知 $A^{\frac{1}{2}} > 0$ 。

(3) 显然:

$$(A^{-\frac{1}{2}})^T = (Q^T \Lambda^{-\frac{1}{2}} Q)^T = Q^T (\Lambda^{-\frac{1}{2}})^T Q = Q^T \Lambda^{-\frac{1}{2}} Q = A^{-\frac{1}{2}}$$

□