

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

域的概念	1
齐次方程组的解与逆矩阵	11
回头改证明，同时注意数域问题	39
Hermitian 转置不改变秩	39
证明有问题	49
Schmidt 正交化链接	64
可逆矩阵行列式链接	70
可逆矩阵行列式链接	71
有空证明	73
存在性证明	77
K^n 子空间为闭集	77
期望的线性性	83
Jensen 不等式链接	87
期望的线性性质，Lebesgue 积分	88
链接 Lebesgue 积分性质	90
链接独立性条件	91
需要证明对小于的都存在	92
写完变量变换法链接过来	95
期望的性质	96
写完矩阵的秩做链接	96
写完变量变换法链接过来	96
期望的性质	99
需要补充证明，但这里涉及到了 Jordan 标准形，学完再来补。	105
链接独立方差等于和	108
考虑链接什么过来	112
链接随机变量函数的分布中的增补变量法	116
复相关系数性质	120
链接复相关系数的性质	120
写线性模型的时候再把多重共线性推导链接过来	121

期望的性质	124
证明	127
似然比检验	127
自由度的计算	127
行列式等于特征值的积，行列式大于 0 矩阵可逆	130
$A(X^T X)^{-1} A^T$ 的可逆性	134
$A(X^T X)^{-1} A^T$ 的正定性	142
$A(X^T X)^{-1} A^T$ 的正定性	144
T 的可逆性证明	146

目录

第一章 线性空间	1
1.1 线性空间	1
1.1.1 线性空间的概念与基本性质	1
1.1.2 线性相关与线性无关	3
1.1.3 子空间	11
1.1.4 线性空间的同构	18
1.1.5 商空间	20
1.2 线性映射	23
1.2.1 线性映射的定义与基本性质	23
1.2.2 线性映射的矩阵表示	26
1.3 线性变换	28
1.3.1 线性变换的性质	29
1.3.2 线性变换的特征值与特征向量	30
1.3.3 不变子空间	33
1.3.4 投影变换	33
1.4 内积空间	35
第二章 矩阵	36
2.1 矩阵空间	36
2.1.1 矩阵的运算	36
2.1.2 矩阵的行列式	38
2.1.3 矩阵的秩	38
2.2 矩阵的向量空间	39
2.3 线性方程组	39
2.3.1 初等方法	41
2.3.2 秩与子空间	41
2.3.3 解的结构	42
2.4 矩阵的等价关系	43
2.4.1 相抵	43

2.4.2	相似	44
2.4.3	合同	44
2.5	相抵的应用	47
2.5.1	广义逆	47
2.5.2	Moore-Penrose 广义逆	50
2.5.3	线性方程组的解	54
2.6	相似的应用	56
2.6.1	特征值与特征向量	56
2.6.2	矩阵的对角化	60
2.6.3	Hermitian 矩阵的对角化	62
2.7	合同的应用——二次型	65
2.7.1	二次型的规范形	66
2.7.2	正定二次型与正定矩阵	68
2.8	特殊矩阵	74
2.8.1	幂等阵	74
2.8.2	正交投影阵	76
2.9	矩阵的分解	77
2.9.1	SVD 分解	77
第三章	随机变量的数字特征	79
3.1	期望	79
3.2	方差	79
3.3	矩	80
3.3.1	原点矩	80
3.3.2	中心矩	81
3.4	协方差	81
3.5	二次型	83
3.6	矩母函数	87
3.7	累积量生成函数	88
3.8	特征函数	89
3.9	Fisher 信息量	93
第四章	正态分布与三大抽样分布	95
4.1	多元正态分布	95
4.1.1	多元正态分布的定义	95
4.1.2	多元正态分布的性质	97
4.1.3	矩阵正态分布的定义	104
4.1.4	矩阵正态分布的性质	106
4.2	χ^2 分布, t 分布和 F 分布	107

4.2.1	χ^2 分布	107
4.2.2	t 分布	109
4.2.3	F 分布	109
第五章	something	111
5.1	抽样分布	111
5.2	次序统计量	114
5.3	充分统计量	117
5.4	Delta method	117
5.5	主成分分析	117
5.5.1	总体主成分分析	117
5.5.2	样本主成分分析	121
5.5.3	注意事项	121
5.6	因子分析	122
5.6.1	参数估计方法	124
5.6.2	因子旋转	126
5.6.3	模型检验	127
5.6.4	因子得分	128
第六章	线性模型	129
6.1	一般线性模型	129
6.1.1	参数估计	130
6.1.2	约束最小二乘估计	133
6.1.3	实际计算	136
6.1.4	预测	136
6.2	正态线性模型	138
6.2.1	参数估计	138
6.2.2	假设检验	141
6.2.3	置信域	143
6.2.4	区间预测	145
6.3	误差协方差推广	146
6.3.1	广义最小二乘估计	146
6.3.2	最小二乘统一理论	147

Chapter 1

线性空间

1.1 线性空间

1.1.1 线性空间的概念与基本性质

Definition 1.1. 设 S 是一个非空集合, $S \times S$ 是 S 与自身的一个 Cartesian product (定义可见??), 则 $f: S \times S \rightarrow S$ 称为 S 上一个二元代数运算 (binary algebraic operation), 简称为 S 上的一个运算。

线性空间的定义

Definition 1.2. 设 X 是一个非空集合, F 是一个域。如果 X 上有一个运算, 即 $f: (\alpha, \beta) \rightarrow \gamma(\alpha, \beta, \gamma \in X)$, 将该运算称为加法 (addition), 把 γ 称为 α 与 β 的和 (sum), 记作 $\alpha + \beta = \gamma$; 同时 F 与 X 有一个运算, 即 $g: (k, \alpha) \rightarrow \delta(k \in F, \alpha, \delta \in X)$, 将该运算称为纯量乘法 (scalar multiplication), 把 δ 称为 k 与 α 的纯量乘积 (scalar multiple), 记作 $k\alpha = \delta$ 。若上述两个运算还满足以下 8 条运算法则:

域的概念

1. $\forall \alpha, \beta \in X, \alpha + \beta = \beta + \alpha$;
2. $\forall \alpha, \beta, \gamma \in X, (\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$;
3. X 中有一个元素, 记作 $\mathbf{0}$, 称为 X 的零元 (zero vector), 它使得:

$$\forall \alpha \in X, \alpha + \mathbf{0} = \alpha$$

4. 对于任意的 $\alpha \in X$, 存在与之对应的 $\beta \in X$, 称为 α 的负元 (additive inverse), 记作 $-\alpha$, 它使得:

$$\alpha + \beta = \mathbf{0}$$

5. F 中的单位元 (multiplicative identity) 1 满足 $\forall \alpha \in X, 1\alpha = \alpha$;
6. $\forall \alpha \in X, k, l \in F, (kl)\alpha = k(l\alpha)$;

$$7. \forall \alpha \in X, k, l \in F, (k+l)\alpha = k\alpha + l\alpha;$$

$$8. \forall \alpha, \beta \in X, k \in F, k(\alpha + \beta) = k\alpha + k\beta.$$

那么称 X 是域 F 上的一个线性空间 (*linear space*), 称 X 中的元素为向量 (*vector*)。

线性空间的基本性质

Property 1.1.1. 域 F 上的线性空间 X 具有如下性质:

1. X 中的零元是唯一的;
2. X 中每个元素的负元是唯一的;
3. $\forall \alpha \in X, 0\alpha = \mathbf{0}$;
4. $\forall k \in F, k\mathbf{0} = \mathbf{0}$;
5. 设 $k \in F, \alpha \in X$ 。如果 $k\alpha = \mathbf{0}$, 那么 $k = 0$ 或 $\alpha = \mathbf{0}$ 。
6. $\forall \alpha \in X, (-1)\alpha = -\alpha$;

Proof. (1) 假设 X 中有两个零元 $\mathbf{0}_1, \mathbf{0}_2$ 且 $\mathbf{0}_1 \neq \mathbf{0}_2$, 由线性空间运算法则 (3) 可得:

$$\mathbf{0}_1 + \mathbf{0}_2 = \mathbf{0}_1, \mathbf{0}_2 + \mathbf{0}_1 = \mathbf{0}_2$$

而由线性空间运算法则 (1) 可得:

$$\mathbf{0}_1 + \mathbf{0}_2 = \mathbf{0}_2 + \mathbf{0}_1$$

于是 $\mathbf{0}_1 = \mathbf{0}_2$, 产生矛盾, 所以 X 中的零元是唯一的。

(2) 任取 X 中的一个元素 α , 假设它有两个负元 β_1, β_2 。由线性空间运算法则 (4)(3)(2) 可得:

$$\begin{aligned} (\beta_1 + \alpha) + \beta_2 &= \mathbf{0} + \beta_2 = \beta_2 \\ (\beta_1 + \alpha) + \beta_2 &= \beta_1 + (\alpha + \beta_2) = \beta_1 + \mathbf{0} = \beta_1 \end{aligned}$$

所以 $\beta_1 = \beta_2$, 产生矛盾。由 α 的任意性, X 中每个元素的负元都是唯一的。

(3) 由线性空间运算法则 (7) 可得:

$$0\alpha + 0\alpha = (0 + 0)\alpha = 0\alpha$$

两边同时加上 -0α 可得:

$$0\alpha + 0\alpha + (-0\alpha) = 0\alpha + (-0\alpha)$$

由线性空间运算法则 (2)(4) 和 (3) 可得:

$$0\alpha = \mathbf{0}$$

(4) 由线性空间运算法则 (8) 和 (3) 可得:

$$k\mathbf{0} + k\mathbf{0} = k(\mathbf{0} + \mathbf{0}) = k\mathbf{0}$$

两边加上 $-k\mathbf{0}$ 再由线性空间运算法则 (2)(4) 和 (3) 可得:

$$k\mathbf{0} = \mathbf{0}$$

(5) 如果 $k \neq 0$, 依次由线性空间运算法则 (5)、(6) 和线性空间基本性质 (4) 可得:

$$\alpha = 1\alpha = (k^{-1}k)\alpha = k^{-1}(k\alpha) = k^{-1}\mathbf{0} = \mathbf{0}$$

(6) 由线性空间运算法则 (5) 与 (7) 以及线性空间基本性质 (3) 可得:

$$\alpha + (-1)\alpha = 1\alpha + (-1)\alpha = (1 - 1)\alpha = 0\alpha = \mathbf{0}$$

再由负元的定义, $(-1)\alpha = -\alpha$ 。 □

Definition 1.3. 设 X 是域 F 上的线性空间。由性质 1.1.1(2), 定义 $f: (\alpha, \beta) \rightarrow \alpha + (-\beta) \in X (\alpha, \beta \in X)$, 将该运算称为**减法 (subtraction)**, 把 $\alpha + (-\beta)$ 称为 α 与 β 的**差 (difference)**, 记作 $\alpha - \beta = \alpha + (-\beta)$ 。

1.1.2 线性相关与线性无关

Definition 1.4. X 是域 F 上的线性空间。按照一定顺序写出的有限多个向量 (允许有相同的向量) 称为 X 的一个**向量组 (set of vectors)**, 如 $\alpha_1, \alpha_2, \dots, \alpha_n$ 。

Definition 1.5. X 是域 F 上的线性空间。对于 X 中的一组向量 $\alpha_1, \alpha_2, \dots, \alpha_n$ 和 F 中的一组元素 k_1, k_2, \dots, k_n , 作纯量乘法和加法得到:

$$\sum_{i=1}^n k_i \alpha_i \in X$$

称该向量为 $\alpha_1, \alpha_2, \dots, \alpha_n$ 的一个**线性组合 (linear combination)**。

Definition 1.6. X 是域 F 上的线性空间。若 $\beta \in X$ 可以表示成向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 的一个线性组合, 则称 β 可以由 $\alpha_1, \alpha_2, \dots, \alpha_n$ **线性表出**。若 $\beta \in X$ 可以由向量集 $W \subseteq X$ 中有限多个向量构成的向量组线性表出, 则称 β 可以由向量集 W **线性表出**。

Definition 1.7. X 是域 F 上的线性空间。按照如下方式定义 X 中对象的**线性相关 (linearly dependent)**与**线性无关 (linearly independent)**:

研究对象	线性相关	线性无关
X 中的向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$	F 中有不全为 0 的元素 k_1, k_2, \dots, k_n 使得 $\sum_{i=1}^n k_i \alpha_i = \mathbf{0}$	从 $\sum_{i=1}^n k_i \alpha_i = \mathbf{0}$ 可以推出 $k_1 = k_2 = \dots = k_n = 0$
空集		定义空集是线性无关的
X 的非空有限子集	给这个子集的元素一种编号 所得的向量组线性相关	给这个子集的元素一种编号所得的 向量组线性无关
X 的无限子集 W	W 有一个有限子集线性相关	W 的任一有限子集都线性无关

Property 1.1.2. 设 X 是域 F 上的线性空间，则：

1. 如果 X 中向量组的一个部分组线性相关，那么这个向量组线性相关；
2. 包含 $\mathbf{0}$ 的向量组是线性相关的；
3. 元素个数大于 1 的向量集 W 线性相关当且仅当 W 中至少有一个向量可以由其余向量中的有限多个线性表出，从而 W 线性无关当且仅当 W 中的每一个向量都不能由其余向量中的有限多个线性表出。

Proof. (1) 取 X 中的向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ ，其部分组 $\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_m}$ ， $i_m, m \leq n$ 线性相关，即 F 中存在不全为 0 的一组元素 $k_{i_1}, k_{i_2}, \dots, k_{i_m}$ 使得：

$$k_{i_1} \alpha_{i_1} + k_{i_2} \alpha_{i_2} + \dots + k_{i_m} \alpha_{i_m} = \mathbf{0}$$

在：

$$l_1 \alpha_1 + l_2 \alpha_2 + \dots + l_n \alpha_n$$

中取 $l_j = k_{i_j}$ ， $j = 1, 2, \dots, m$ ，其余系数为 0，则 l_1, l_2, \dots, l_n 不全为 0 同时上式值为 $\mathbf{0}$ ，即向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性相关。

(2) 由性质 1.1.1(4) 可得 $\mathbf{0}$ 作为向量组的部分组是线性相关的，由 (1) 即可得出结果。

(3) 由定义直接得到。 □

Property 1.1.3. 设 X 是域 F 上的线性空间，则：

1. 向量 $\beta \in X$ 可以由向量集 $W \subseteq X$ 线性表出，则表示方法唯一的充分必要条件是 W 线性无关；
2. 若向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关，则向量 β 可以由向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性表出的充分必要条件为 $\alpha_1, \alpha_2, \dots, \alpha_n, \beta$ 线性相关。

Proof. (1) 充分性：因为 β 可以由向量集 W 线性表出，假设此时 β 有以下两种表出方式：

$$\beta = k_1 \alpha_1 + \dots + k_r \alpha_r + k_{r+1} u_1 + \dots + k_{r+s} u_s$$

$$\beta = l_1 \alpha_1 + \dots + l_r \alpha_r + l_{r+1} v_1 + \dots + l_{r+t} v_t$$

其中 $\alpha_1, \alpha_2, \dots, \alpha_r, u_1, u_2, \dots, u_s, v_1, v_2, \dots, v_t \in W$, $k_1, \dots, k_{r+s}, l_1, \dots, l_{r+t} \in F$, $r, s, t \geq 0$ 。二式作差可得:

$$\mathbf{0} = (k_1 - l_1)\alpha_1 + \dots + (k_r - l_r)\alpha_r + k_{r+1}u_1 + \dots + k_{r+s}u_s - l_{r+1}v_1 - \dots - l_{r+t}v_t$$

因为 W 线性无关, 所以向量组 $\alpha_1, \alpha_2, \dots, \alpha_r, u_1, u_2, \dots, u_s, v_1, v_2, \dots, v_t$ 线性无关, 于是:

$$k_1 - l_1 = 0, \dots, k_r - l_r = 0, k_{r+1} = 0, \dots, k_{r+s} = 0, l_{r+1} = 0, \dots, l_{r+t} = 0$$

所以两个表出方式完全相同, β 由 W 线性表出的表示方法唯一。

必要性: 如果 W 线性相关, 则 W 有一个有限子集 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 线性相关, 于是 F 中有不全为 0 的元素 k_1, k_2, \dots, k_n 使得:

$$k_1\alpha_1 + k_2\alpha_2 + \dots + k_n\alpha_n = \mathbf{0}$$

由于 β 可以由 W 线性表出, 所以:

$$\beta = l_1\alpha_1 + \dots + l_n\alpha_n + l_{n+1}v_1 + \dots + l_{n+s}v_s$$

其中 $l_i \in F$, $i = 1, 2, \dots, n$; $v_j \in X$, $j = 1, 2, \dots, s$ 。将上两式相加可得:

$$\beta = (l_1 + k_1)\alpha_1 + \dots + (l_n + k_n)\alpha_n + l_{n+1}v_1 + \dots + l_{n+s}v_s$$

因为 k_1, k_2, \dots, k_n 不全为 0, 所以有序元素组:

$$(l_1, \dots, l_n, l_{n+1}, \dots, l_{n+s}) \neq (l_1 + k_1, \dots, l_n + k_n, l_{n+1}, \dots, l_{n+s})$$

于是 β 由 W 线性表出的方式不唯一, 矛盾, 所以 W 线性无关。

(2) 充分性: 因为 $\alpha_1, \alpha_2, \dots, \alpha_n, \beta$ 线性相关, 所以域 F 中存在不全为 0 的元素 k_1, k_2, \dots, k_n, l 使得:

$$k_1\alpha_1 + k_2\alpha_2 + \dots + k_n\alpha_n + l\beta = \mathbf{0}$$

若 $l = 0$, 则域 F 中存在不全为 0 的元素 k_1, k_2, \dots, k_n 使得:

$$k_1\alpha_1 + k_2\alpha_2 + \dots + k_n\alpha_n = \mathbf{0}$$

这与向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关矛盾, 所以 $l \neq 0$ 。于是:

$$\beta = -\frac{k_1}{l}\alpha_1 - \frac{k_2}{l}\alpha_2 - \dots - \frac{k_n}{l}\alpha_n$$

即向量 β 可以由向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性表出。

必要性: 因为 β 可以由向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性表出, 所以域 F 中存在不全为 0 的元素 k_1, k_2, \dots, k_n, l 使得:

$$\beta = k_1\alpha_1 + k_2\alpha_2 + \dots + k_n\alpha_n$$

移项即可得到:

$$k_1\alpha_1 + k_2\alpha_2 + \dots + k_n\alpha_n - \beta = \mathbf{0}$$

因为 $-1 \neq 0$, 所以 $\alpha_1, \alpha_2, \dots, \alpha_n, \beta$ 线性相关。 □

极大线性无关组与秩

Definition 1.8. 设 X 是域 F 上的一个线性空间。向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 中的一个部分组若满足下述条件:

1. 本身线性无关;
2. 若该部分组不等于向量组, 则从向量组的其余向量中任取一个向量添加进该部分组都将使部分组线性相关。若该部分组等于向量组, 则跳过此条件。

则称该部分组是向量组的一个极大线性无关组 (*maximal linearly independent system*)。

Definition 1.9. 如果向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 的每一个向量都可以由向量组 $\beta_1, \beta_2, \dots, \beta_r$ 线性表出, 那么称向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 可以由向量组 $\beta_1, \beta_2, \dots, \beta_r$ 线性表出。如果向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 和向量组 $\beta_1, \beta_2, \dots, \beta_r$ 可以互相线性表出, 则称两个向量组等价 (*equivalent*), 记作:

$$\{\alpha_1, \alpha_2, \dots, \alpha_s\} \cong \{\beta_1, \beta_2, \dots, \beta_r\}$$

Theorem 1.1. X 是域 K 上的一个线性空间。 X 中向量组的等价是 X 中向量组的一个等价关系。

Proof. (1) 反身性与 (2) 对称性显然成立。

(3) 传递性: 若向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 可以由向量组 $\beta_1, \beta_2, \dots, \beta_r$ 线性表出, 且向量组 $\beta_1, \beta_2, \dots, \beta_r$ 可以由向量组 $\gamma_1, \gamma_2, \dots, \gamma_t$ 线性表出, 于是有:

$$\alpha_i = \sum_{j=1}^r k_{ji} \beta_j, \quad i = 1, 2, \dots, s \quad \beta_j = \sum_{l=1}^t q_{lj} \gamma_l, \quad j = 1, 2, \dots, r$$

于是:

$$\alpha_i = \sum_{j=1}^r k_{ji} \left(\sum_{l=1}^t q_{lj} \gamma_l \right) = \sum_{j=1}^r \sum_{l=1}^t k_{ji} q_{lj} \gamma_l = \sum_{l=1}^t \left(\sum_{j=1}^r k_{ji} q_{lj} \right) \gamma_l, \quad i = 1, 2, \dots, s$$

即向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 可以由向量组 $\gamma_1, \gamma_2, \dots, \gamma_t$ 线性表出。传递性得证。 \square

Theorem 1.2. 设 X 是域 F 上的一个线性空间, 则:

1. 若向量组 $\beta_1, \beta_2, \dots, \beta_r$ 可以由向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 线性表出, 同时 $r > s$, 那么向量组 $\beta_1, \beta_2, \dots, \beta_r$ 线性相关;
2. 若向量组 $\beta_1, \beta_2, \dots, \beta_r$ 可以由向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 线性表出, 同时向量组 $\beta_1, \beta_2, \dots, \beta_r$ 线性无关, 就有 $r \leq s$;
3. 等价的线性无关的向量组所含向量的个数相同。

Proof. (1) 考虑方程:

$$k_1\beta_1 + k_2\beta_2 + \cdots + k_r\beta_r = \mathbf{0}$$

因为向量组 $\beta_1, \beta_2, \dots, \beta_r$ 可以由向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 线性表出, 所以:

$$\begin{cases} \beta_1 = a_{11}\alpha_1 + a_{12}\alpha_2 + \cdots + a_{1s}\alpha_s \\ \beta_2 = a_{21}\alpha_1 + a_{22}\alpha_2 + \cdots + a_{2s}\alpha_s \\ \vdots \\ \beta_r = a_{r1}\alpha_1 + a_{r2}\alpha_2 + \cdots + a_{rs}\alpha_s \end{cases}$$

则有:

$$\begin{aligned} k_1\beta_1 + k_2\beta_2 + \cdots + k_r\beta_r &= k_1(a_{11}\alpha_1 + a_{12}\alpha_2 + \cdots + a_{1s}\alpha_s) \\ &\quad + k_2(a_{21}\alpha_1 + a_{22}\alpha_2 + \cdots + a_{2s}\alpha_s) + \cdots \\ &\quad + k_r(a_{r1}\alpha_1 + a_{r2}\alpha_2 + \cdots + a_{rs}\alpha_s) \\ &= (k_1a_{11} + k_2a_{21} + \cdots + k_ra_{r1})\alpha_1 \\ &\quad + (k_1a_{12} + k_2a_{22} + \cdots + k_ra_{r2})\alpha_2 + \cdots \\ &\quad + (k_1a_{1s} + k_2a_{2s} + \cdots + k_ra_{rs})\alpha_s \\ &= \mathbf{0} \end{aligned}$$

将上式看作 k_1, k_2, \dots, k_r 的线性方程, 考虑如下齐次线性方程组:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1s} \\ a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rs} \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_r \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

因为 $s < r$, 由定理 2.5 可知上述齐次线性方程组必有非零解。取它的一个非零解 (k_1, k_2, \dots, k_r) , 由性质 1.1.1(3) 和线性空间运算法则 (3) 即可得:

$$k_1\beta_1 + k_2\beta_2 + \cdots + k_r\beta_r = 0\alpha_1 + 0\alpha_2 + \cdots + 0\alpha_s = \mathbf{0}$$

于是向量组 $\beta_1, \beta_2, \dots, \beta_r$ 线性相关。

(2) 是 (1) 的逆否命题。

(3) 可由 (2) 直接得到。

□

Property 1.1.4. 设 X 是域 F 上的一个线性空间, 则:

1. 一个向量组与它的任意一个极大线性无关组等价;
2. 一个向量组的任意两个极大线性无关组等价;
3. 一个向量组的任意两个极大线性无关组所含向量的个数相同。

Proof. (1) 设 $\alpha_1, \alpha_2, \dots, \alpha_s$ 是一个向量组, 任取它的一个极大线性无关组 $\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_n}$ 。显然 $\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_n}$ 可以由 $\alpha_1, \alpha_2, \dots, \alpha_s$ 线性表出。由极大线性无关组的定义与性质 1.1.3(2) 的充分性可直接得到 $\alpha_1, \alpha_2, \dots, \alpha_s$ 可以由 $\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_n}$ 线性表出。

(2) 由 (1) 以及向量组等价的对称性与传递性可直接推出。

(3) 由定理 1.2(3) 直接得到。 \square

Definition 1.10. 向量组的极大线性无关组所含向量的个数称为这个向量组的秩 (rank)。把向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 的秩记作 $\text{rank}\{\alpha_1, \alpha_2, \dots, \alpha_s\}$ 。全由零向量组成的向量组的秩规定为 0。

Property 1.1.5. 设 X 是域 F 上的一个线性空间, 则:

1. 向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 线性无关的充分必要条件是它的秩等于它所含向量的个数。
2. 如果向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 可以由向量组 $\beta_1, \beta_2, \dots, \beta_r$ 线性表出, 那么:

$$\text{rank}\{\alpha_1, \alpha_2, \dots, \alpha_s\} \leq \text{rank}\{\beta_1, \beta_2, \dots, \beta_r\}$$

3. 等价的向量组具有相等的秩。

Proof. (1) $\alpha_1, \alpha_2, \dots, \alpha_s$ 线性无关 \iff 极大线性无关组就是自身 \iff 秩等于所含向量的个数。

(2) 任取向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 的一个极大线性无关组 $\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_n}$, 再取向量组 $\beta_1, \beta_2, \dots, \beta_r$ 的一个极大线性无关组 $\beta_{j_1}, \beta_{j_2}, \dots, \beta_{j_m}$ 。因为向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 可以由向量组 $\beta_1, \beta_2, \dots, \beta_r$ 线性表出, 所以 $\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_n}$ 可以由 $\beta_{j_1}, \beta_{j_2}, \dots, \beta_{j_m}$ 线性表出, 因为 $\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_n}$ 线性无关, 由定理 1.2(2) 可得, $n \leq m$, 即 $\text{rank}\{\alpha_1, \alpha_2, \dots, \alpha_s\} \leq \text{rank}\{\beta_1, \beta_2, \dots, \beta_r\}$ 。

(3) 由 (2) 可直接推得。 \square

基与维数

Definition 1.11. 设 X 是域 F 上的一个线性空间。 X 中的向量集 S 若满足下述条件:

1. 本身线性无关;
2. X 中的每一个向量都可以由 S 中有限多个向量线性表出。

则称 S 是向量组的一个基 (basis)。

Theorem 1.3. 任一域 F 上的任一线性空间 X 都有一组基。

该定理的证明不提供, 涉及 Zorn 引理。

Definition 1.12. X 是域 F 上的线性空间。如果 X 的一组基是由有限多个向量组成的, 那么称 X 是有限维的 (finite-dimensional); 如果 X 有一组基含有无穷多个向量, 则称 X 是无限维的 (infinite-dimensional)。

Theorem 1.4. 如果域 F 上的线性空间 X 是有限维的, 那么 X 的任意两个基所含向量的个数相同。

Proof. 由有限维线性空间的定义, X 存在一组基只有有限多个向量, 记为 $\alpha_1, \alpha_2, \dots, \alpha_n$ 。再取 X 的一组基 S , 从中取出 $n+1$ 个向量 $\beta_1, \beta_2, \dots, \beta_{n+1}$ (考虑 S 可能含有无数个向量的情况)。因为 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 X 的基, 所以 $\beta_1, \beta_2, \dots, \beta_{n+1}$ 可以由 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性表出。因为 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关, 所以 $\beta_1, \beta_2, \dots, \beta_{n+1}$ 线性相关, 而此时 S 也应线性相关, 矛盾, 所以 S 中的元素小于 $n+1$ 个。设 $S = \{\beta_1, \beta_2, \dots, \beta_m\}$, 则 S 与 $\alpha_1, \alpha_2, \dots, \alpha_n$ 都线性无关且二者等价, 由定理 1.2(3), 它们具有相同的秩, 即所含向量个数相同。由 S 的任意性与向量组等价的传递性、对称性, X 的任意两个基所含向量的个数相同。 \square

Corollary 1.1. 如果域 F 上的线性空间 X 是无限维的, 那么 X 的任意一组基都含有无穷多个向量。

Proof. 如果 X 有一组基由有限多个向量组成, 那么 X 也是一个有限维线性空间, 而有限维线性空间所有基所含向量的个数都相同, 那么 X 就不可能有一组基含有无穷多个向量, 与无穷维线性空间的定义矛盾。 \square

Definition 1.13. X 是域 F 上的线性空间。如果 X 是有限维的, 那么把 X 的基所含向量的个数称为 X 的维数 (dimension), 记作 $\dim V$; 如果 X 是无限维的, 那么记 $\dim V = +\infty$ 。

Property 1.1.6. 设 X 是域 F 上的 n 维线性空间, 则:

1. X 中任意 $n+1$ 个向量都线性相关。
2. 任意 n 个线性无关的向量都是 X 的一组基;
3. 如果 X 中任一向量都可以由 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性表出, 则 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 X 的一组基;
4. X 中任意一个线性无关的向量组都可以扩充成 X 的一组基。

Proof. (1) 任意 $n+1$ 个向量都可以由一组基线性表出, 而每一组基所含向量个数都是 n , 由定理 1.2(1), 这 $n+1$ 个向量线性相关。

(2) 任取 X 中 n 个线性无关的向量 $\alpha_1, \alpha_2, \dots, \alpha_n$, 对任意的 $\beta \in X$, 由上一个定理, 向量组 $\alpha_1, \alpha_2, \dots, \alpha_n, \beta$ 线性相关。由性质 1.1.3(2), β 可以由 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性表出。由基的定义, $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 X 的一组基。由 $\alpha_1, \alpha_2, \dots, \alpha_n$ 的任意性, 命题成立。

(3) 取 X 的一组基 $\beta_1, \beta_2, \dots, \beta_n$, 则 $\beta_1, \beta_2, \dots, \beta_n$ 可以由 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性表出。由性质 1.1.5(2), $n = \text{rank}\{\beta_1, \beta_2, \dots, \beta_n\} \leq \text{rank}\{\alpha_1, \alpha_2, \dots, \alpha_n\} \leq n$, 所以 $\text{rank}\{\alpha_1, \alpha_2, \dots, \alpha_n\} = n$, $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关。由 (2), $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 X 的一组基。

(4) 任取 X 中一个线性无关的向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$, 若 $r = n$, 由 (2) 可知, $\alpha_1, \alpha_2, \dots, \alpha_r$ 是 X 的一组基; 若 $r < n$, 则 X 中必定存在一个元素不能由 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性表出, 将其记为 α_{r+1} , 否则的话 $\alpha_1, \alpha_2, \dots, \alpha_r$ 就是 X 的一组基, 进而 X 的维数应是 r , 矛盾。不断重复上述过程即可得到一个向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$, 这就是 X 的一组基。 \square

坐标与坐标变换

Definition 1.14. X 是域 F 上的 n 维线性空间, $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 X 的一组基, 由性质 1.1.3(I) 可知 X 中任一向量 α 由 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性表出的方式唯一:

$$\alpha = a_1\alpha_1 + a_2\alpha_2 + \cdots + a_n\alpha_n$$

把系数构成的 n 元有序数组写成列向量的形式, 得到 $(a_1, a_2, \dots, a_n)^T$, 该列向量被称为 α 在基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的坐标 (coordinate)。

接下来我们要讨论的是 X 中某一向量在不同基下的坐标之间有什么关系, 首先我们需要定义什么叫不同的基。

Definition 1.15. X 是域 F 上的 n 维线性空间。 X 中的两个向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 与 $\beta_1, \beta_2, \dots, \beta_n$ 如果满足 $\alpha_i = \beta_i, i = 1, 2, \dots, n$, 那么称这两个向量组相等。

Definition 1.16. X 是域 F 上的 n 维线性空间。给定 V 的两个基:

$$\alpha_1, \alpha_2, \dots, \alpha_n \quad \beta_1, \beta_2, \dots, \beta_n$$

因为 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 V 的一组基, 所以有:

$$\begin{cases} \beta_1 = a_{11}\alpha_1 + a_{12}\alpha_2 + \cdots + a_{1n}\alpha_n \\ \beta_2 = a_{21}\alpha_1 + a_{22}\alpha_2 + \cdots + a_{2n}\alpha_n \\ \dots\dots\dots \\ \beta_n = a_{n1}\alpha_1 + a_{n2}\alpha_2 + \cdots + a_{nn}\alpha_n \end{cases}$$

模仿矩阵乘法的定义将上式写作:

$$(\beta_1, \beta_2, \dots, \beta_n) = (\alpha_1, \alpha_2, \dots, \alpha_n) \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix}$$

将上式右端的矩阵记作 A , 称它是基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 到基 $\beta_1, \beta_2, \dots, \beta_n$ 的过渡矩阵 (transition matrix)。于是上式可以写作:

$$(\beta_1, \beta_2, \dots, \beta_n) = (\alpha_1, \alpha_2, \dots, \alpha_n)A$$

由于这种写法是模仿矩阵乘法的定义, 所以矩阵乘法所满足的运算法则对于这种写法也成立。

Theorem 1.5. X 是域 F 上的 n 维线性空间, $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 X 的一组基, 且向量组 $\beta_1, \beta_2, \dots, \beta_n$ 满足:

$$(\beta_1, \beta_2, \dots, \beta_n) = (\alpha_1, \alpha_2, \dots, \alpha_n)A$$

则 $\beta_1, \beta_2, \dots, \beta_n$ 是 X 的一组基当且仅当 A 是可逆矩阵。

齐次方程组
的解与逆矩
阵

Proof. 由可得:

$\beta_1, \beta_2, \dots, \beta_n$ 是 X 的一组基 \iff 由 $\sum_{i=1}^n k_i \beta_i = \mathbf{0}$ 可推出 $k_1 = k_2 = \dots = k_n = 0$

$$\iff \text{由 } (\alpha_1, \alpha_2, \dots, \alpha_n) A \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{pmatrix} = \mathbf{0} \text{ 可推出 } k_1 = k_2 = \dots = k_n = 0$$

$$\iff \text{由 } A \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{pmatrix} = \mathbf{0} \text{ 可推出 } k_1 = k_2 = \dots = k_n = 0$$

\iff 齐次线性方程组 $Ax = \mathbf{0}$ 只有零解

$\iff A$ 可逆 □

Theorem 1.6. X 是域 F 上的 n 维线性空间, $\alpha_1, \alpha_2, \dots, \alpha_n$ 和 $\beta_1, \beta_2, \dots, \beta_n$ 是 X 上的两个基, A 是由基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 到基 $\beta_1, \beta_2, \dots, \beta_n$ 的过渡矩阵。若 X 中向量 α 在这两个基下的坐标分别为:

$$x = (x_1, x_2, \dots, x_n)^T, \quad y = (y_1, y_2, \dots, y_n)^T$$

则有:

$$y = A^{-1}x$$

Proof. 因为:

$$\begin{aligned} \alpha &= (\alpha_1, \alpha_2, \dots, \alpha_n)(x_1, x_2, \dots, x_n)^T \\ &= (\beta_1, \beta_2, \dots, \beta_n)(y_1, y_2, \dots, y_n)^T \\ &= (\alpha_1, \alpha_2, \dots, \alpha_n)A(y_1, y_2, \dots, y_n)^T \end{aligned}$$

所以:

$$(x_1, x_2, \dots, x_n)^T = A(y_1, y_2, \dots, y_n)^T$$

即 $x = Ay$ 。因为 $\alpha_1, \alpha_2, \dots, \alpha_n$ 和 $\beta_1, \beta_2, \dots, \beta_n$ 是 X 上的两个基, 所以 A 可逆, 于是 $y = A^{-1}x$ 。 □

1.1.3 子空间

子空间的定义及其判别

Definition 1.17. X 是域 F 上的一个线性空间, E 是 X 的一个非空子集。若 E 对于 X 上的加法和纯量乘法也构成域 F 上的一个线性空间, 则称 E 是 X 的一个线性子空间 (linear subspace), 简称为子空间 (subspace)。

Theorem 1.7. X 是域 F 上的一个线性空间, E 是 X 的一个非空子集。 E 是 X 的一个子空间的充分必要条件是 E 对 X 中的加法和纯量乘法封闭, 即:

$$\alpha, \beta \in E \Rightarrow \alpha + \beta \in E$$

$$k \in F, \alpha \in E \Rightarrow k\alpha \in E$$

Proof. (1) 必要性: 因为 E 是 X 的子空间, 由子空间的定义, E 中的加法和纯量乘法就是 X 中的加法和纯量乘法, 由加法和纯量乘法的定义, E 对 X 中的加法和纯量乘法封闭。

(2) 充分性: 由条件可知此时已经对 E 定义了加法和纯量乘法, 且就是 X 中的加法和纯量乘法, 还需要证明的是 E 对于它自身的加法和纯量乘法满足线性空间的八条运算法则。对 $\forall \alpha, \beta, \gamma \in E, \forall k, l \in F$:

1. 因为 $\alpha, \beta, \gamma \in E$, 所以 $\alpha, \beta, \gamma \in X$, 对于 X 上的加法, 有 $\alpha + \beta = \beta + \alpha, (\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$, 而 X 上的加法就是 E 上的加法, 所以 E 上的加法满足线性空间运算法则 (1)(2); 同理, E 上的纯量乘法满足线性空间运算法则 (5)(6)(7)(8);
2. 因为 E 不是空集, 所以存在 $\delta \in E$ 。因为 $\delta \in X$, 所以由 X 中的纯量乘法可得 $0\delta = \mathbf{0}_X \in E$ 。对任意的 $\alpha \in E$, 有 $\alpha \in X$, 根据 X 中的加法有 $\alpha + \mathbf{0}_X = \alpha$, 于是 $\mathbf{0}_X$ 是 E 中的零元, 即 E 满足线性空间运算法则 (3);
3. 因为 E 对纯量乘法封闭, 所以 $(-1)\alpha \in E$ 。因为 $\alpha + (-1)\alpha = [1 + (-1)]\alpha = 0\alpha = \mathbf{0}$ (第一步到第二步由 1, 第三步到第四步因为 $\alpha \in X$), 所以 $(-1)\alpha$ 是 α 的负元, 即 E 满足线性空间运算法则 (4)。

□

子空间的性质

Theorem 1.8. 设 X 是域 F 上的一个线性空间, E 是 X 的任意一个子空间, 则有 $\dim(E) \leq \dim(X)$ 。 X 是有限维时等号成立当且仅当 $E = X$ 。

Proof. 由性质 1.1.6(4) 可得, E 的一组基可以扩充成 X 的一组基, 所以 $\dim(E) \leq \dim(X)$ 。当 $E = X$ 时, 显然 $\dim(E) = \dim(X)$ 。当 $\dim(E) = \dim(X)$ 时, 由性质 1.1.6(2) 可知 E 的一组基就是 X 的一组基, 所以 X 中的任一向量可以由 E 的基线性表出, 于是 $X \subseteq E$, 从而 $E = X$ 。 □

张成的子空间

Definition 1.18. 设 X 是域 F 上的一个线性空间, $\alpha_1, \alpha_2, \dots, \alpha_n \in X$, 称:

$$W = \left\{ \sum_{i=1}^n k_i \alpha_i : k_i \in F \right\}$$

为向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 张成的线性子空间, 记作 $\langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$ 。

Property 1.1.7. 设 X 是域 F 上的一个线性空间, $\alpha_1, \alpha_2, \dots, \alpha_n \in X$. $\langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$ 具有如下性质:

1. $\langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$ 是 X 的一个子空间;
2. $\langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$ 是 X 中包含 $\alpha_1, \alpha_2, \dots, \alpha_n$ 的最小的子空间;
3. $\alpha_1, \alpha_2, \dots, \alpha_n$ 的极大线性无关组是 $\langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$ 的基;
4. $\dim \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle = \text{rank}\{\alpha_1, \alpha_2, \dots, \alpha_n\}$;
5. 若 $\beta_1, \beta_2, \dots, \beta_m \in X$, 则有:

$$\langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle = \langle \beta_1, \beta_2, \dots, \beta_m \rangle \iff \{\alpha_1, \alpha_2, \dots, \alpha_n\} \cong \{\beta_1, \beta_2, \dots, \beta_m\}$$

6. 若 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 X 的一组基, 则 $X = \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$.

Proof. (1) 显然 $\langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$ 中的元素对 X 中的加法与纯量乘法封闭。

(2) 由于子空间对加法与纯量乘法的封闭性, 包含 $\alpha_1, \alpha_2, \dots, \alpha_n$ 的子空间必然包含 $\langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$ 。

(3) 显然。

(4) 由 (3) 直接得到。

(5) 充分性显然, 必要性由反证法可得。

(6) 显然。 □

子空间的交与和

Theorem 1.9. 设 X 是域 F 上的一个线性空间, I 是一个指标集, 对任意的 $i \in I$ 有 X_i 是 X 的子空间, 则

$$\bigcap_{i \in I} X_i = \{\alpha : \alpha \in X_i, \forall i \in I\}$$

也是 X 的子空间。

Proof. 任取 $\alpha, \beta \in \bigcap_{i \in I} X_i$ 和 $k_1, k_2 \in F$ 。对任意的 $i \in I$, 因为 X_i 是 X 的子空间, $\alpha, \beta \in X_i$, 所以 $k_1\alpha + k_2\beta \in X_i$, 于是 $k_1\alpha + k_2\beta \in \bigcap_{i \in I} X_i$ 。由定理 1.7 可知 $\bigcap_{i \in I} X_i$ 是 X 的子空间。□

Definition 1.19. 设 X 是域 F 上的一个线性空间, X_1, X_2, \dots, X_n 是 X 的子空间。定义:

$$\sum_{i=1}^n X_i = \left\{ \sum_{i=1}^n \alpha_i : \alpha_i \in X_i \right\}$$

Theorem 1.10. 设 X 是域 F 上的一个线性空间, X_1, X_2, \dots, X_n 是 X 的子空间, $n \in \mathbb{N}^+$, 则 $X_1 + X_2 + \dots + X_n$ 也是 X 的子空间。

Proof. 任取 $\alpha, \beta \in X_1 + X_2 + \cdots + X_n$ 和 $k_1, k_2 \in F$, 则:

$$\alpha = \gamma_1 + \gamma_2 + \cdots + \gamma_n, \quad \beta = \delta_1 + \delta_2 + \cdots + \delta_n$$

其中 $\gamma_i, \delta_i \in X_i, i = 1, 2, \dots, n$, 于是:

$$k_1\alpha + k_2\beta = \sum_{i=1}^n (k_i\gamma_i + k_2\delta_i)$$

因为 X_i 是 X 的子空间, 所以 $k_i\gamma_i + k_2\delta_i \in X_i, i = 1, 2, \dots, n$, 于是 $k_1\alpha + k_2\beta \in X_1 + X_2 + \cdots + X_n$ 。由定理 1.7 可得 $X_1 + X_2 + \cdots + X_n$ 是 X 的子空间。□

Lemma 1.1. 设 X 是域 F 上的一个线性空间, $\alpha_1, \alpha_2, \dots, \alpha_m, \beta_1, \beta_2, \dots, \beta_n \in X$, 则:

$$\langle \alpha_1, \alpha_2, \dots, \alpha_m \rangle + \langle \beta_1, \beta_2, \dots, \beta_n \rangle = \langle \alpha_1, \alpha_2, \dots, \alpha_m, \beta_1, \beta_2, \dots, \beta_n \rangle$$

Proof. 由子空间与子空间和的定义可直接得到。□

Theorem 1.11. 设 X 是域 F 上的一个线性空间, X_1, X_2 为 X 的有限维子空间, 则 $X_1 \cap X_2, X_1 + X_2$ 也是有限维的, 并且有:

$$\dim(X_1) + \dim(X_2) = \dim(X_1 + X_2) + \dim(X_1 \cap X_2)$$

Proof. 显然 $X_1 \cap X_2$ 是有限维的, 设 $X_1, X_2, X_1 \cap X_2$ 的维数分别为 n_1, n_2, m 。取 $X_1 \cap X_2$ 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_m$, 把它分别扩充为 X_1 和 X_2 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_m, \beta_1, \beta_2, \dots, \beta_{n_1-m}$ 与 $\alpha_1, \alpha_2, \dots, \alpha_m, \gamma_1, \gamma_2, \dots, \gamma_{n_2-m}$ 。由性质 1.1.7(6) 和引理 1.1 可得:

$$\begin{aligned} X_1 + X_2 &= \langle \alpha_1, \alpha_2, \dots, \alpha_m, \beta_1, \beta_2, \dots, \beta_{n_1-m} \rangle + \langle \alpha_1, \alpha_2, \dots, \alpha_m, \gamma_1, \gamma_2, \dots, \gamma_{n_2-m} \rangle \\ &= \langle \alpha_1, \alpha_2, \dots, \alpha_m, \beta_1, \beta_2, \dots, \beta_{n_1-m}, \gamma_1, \gamma_2, \dots, \gamma_{n_2-m} \rangle \end{aligned}$$

根据性质 1.1.7(4) 可得 $\dim(X_1 + X_2) \leq n_1 + n_2 - m$, 即 $X_1 + X_2$ 是有限维的。

设:

$$k_1\alpha_1 + \cdots + k_m\alpha_m + l_1\beta_1 + \cdots + l_{n_1-m}\beta_{n_1-m} + q_1\gamma_1 + \cdots + q_{n_2-m}\gamma_{n_2-m} = \mathbf{0}$$

于是:

$$k_1\alpha_1 + \cdots + k_m\alpha_m + l_1\beta_1 + \cdots + l_{n_1-m}\beta_{n_1-m} = -(q_1\gamma_1 + \cdots + q_{n_2-m}\gamma_{n_2-m})$$

左边属于 X_1 , 右边属于 X_2 , 所以它们属于 $X_1 \cap X_2$ 。于是右边的向量也可以由 $\alpha_1, \alpha_2, \dots, \alpha_m$ 线性表出, 即:

$$(q_1\gamma_1 + \cdots + q_{n_2-m}\gamma_{n_2-m}) = p_1\alpha_1 + \cdots + p_m\alpha_m$$

因为 $\alpha_1, \alpha_2, \dots, \alpha_m, \gamma_1, \gamma_2, \dots, \gamma_{n_2-m}$ 是 X_2 的一组基, 所以:

$$q_1 = q_2 = \cdots = q_{n_2-m} = p_1 = p_2 = \cdots = p_m = 0$$

又因为 $\alpha_1, \alpha_2, \dots, \alpha_m, \beta_1, \beta_2, \dots, \beta_{n_1-m}$ 是 X_1 的一组基, 所以:

$$k_1 = k_2 = \dots = k_m = l_1 = l_2 = \dots = l_{n_1-m} = 0$$

于是 $\alpha_1, \alpha_2, \dots, \alpha_m, \beta_1, \beta_2, \dots, \beta_{n_1-m}, \gamma_1, \gamma_2, \dots, \gamma_{n_2-m}$ 线性无关。由基的定义, 它们是 X 的一组基, 于是有:

$$\begin{aligned} \dim(X_1 + X_2) &= m + n_1 - m + n_2 - m \\ &= \dim(X_1 \cap X_2) + \dim(X_1) - \dim(X_1 \cap X_2) + \dim(X_2) - \dim(X_1 \cap X_2) \\ &= \dim(X_1) + \dim(X_2) - \dim(X_1 \cap X_2) \end{aligned}$$

即:

$$\dim(X_1) + \dim(X_2) = \dim(X_1 + X_2) + \dim(X_1 \cap X_2) \quad \square$$

子空间的直和

Definition 1.20. 设 X 是域 F 上的线性空间, X_1, X_2, \dots, X_n 是 X 的子空间。如果 $X_1 + X_2 + \dots + X_n$ 中的每个向量 α 都能唯一地表示为:

$$\alpha = \sum_{i=1}^n \alpha_i, \alpha_i \in X_i$$

则称和 $X_1 + X_2 + \dots + X_n$ 为直和 (*direct sum*), 记为 $X_1 \oplus X_2 \oplus \dots \oplus X_n$, 也可以写作 $\oplus_{i=1}^n X_i$ 。若 $X = X_1 \oplus X_2$, 则称 X_2 为 X_1 的补空间 (*complement*)。

Theorem 1.12. 设 X 是域 F 上的线性空间, X_1, X_2, \dots, X_n 是 X 的有限维子空间, 则下列命题等价:

1. 和 $X_1 + X_2 + \dots + X_n$ 是直和;
2. 和 $X_1 + X_2 + \dots + X_n$ 中零向量的表示方法唯一;
3. $X_i \cap \left(\sum_{j \neq i} X_j \right) = \mathbf{0}, i, j = 1, 2, \dots, n$;
4. $\dim(X_1 + X_2 + \dots + X_n) = \dim(X_1) + \dim(X_2) + \dots + \dim(X_n)$;
5. $X_i, i = 1, 2, \dots, n$ 的基合起来是和 $X_1 + X_2 + \dots + X_n$ 的一组基。

前三条在 X_1, X_2, \dots, X_n 是无限维子空间时也成立。

Proof. $1 \Rightarrow 2$: 由直和的定义是显然的。

$2 \Rightarrow 3$: 任取 $\alpha \in X_i \cap \left(\sum_{j \neq i} X_j \right)$ 。因为 $\alpha \in X_i$, X_i 是一个子空间, 所以 $-\alpha \in X_i$ 。因为 $\alpha \in \left(\sum_{j \neq i} X_j \right)$, 所以 α 可表示为:

$$\alpha = \sum_{j \neq i} \alpha_j, j = 1, 2, \dots, i-1, i+1, \dots, n$$

于是:

$$\mathbf{0} = \alpha + (-\alpha) = -\alpha + \sum_{j \neq i} \alpha_j$$

因为 $\mathbf{0} = \mathbf{0} + \mathbf{0} + \cdots + \mathbf{0}$, 所以 $-\alpha = \mathbf{0}$, $\alpha = \mathbf{0}$ 。由 α 的任意性, $X_i \cap \left(\sum_{j \neq i} X_j \right) = \mathbf{0}$, $i, j = 1, 2, \dots, n$ 。

$3 \Rightarrow 1$: 假设和 $X_1 + X_2 + \cdots + X_n$ 不是直和, 则存在 $\alpha \in X_1 + X_2 + \cdots + X_n$ 有两种表示方式, 设:

$$\alpha = \sum_{i=1}^n \alpha_i = \sum_{i=1}^n \beta_i, \quad \alpha_i, \beta_i \in X_i, \alpha_i \neq \beta_i$$

则:

$$\alpha_i - \beta_i = \sum_{j \neq i} (\beta_j - \alpha_j), \quad i, j = 1, 2, \dots, n$$

注意到左边属于 X_i , 右边属于 $\sum_{j \neq i} X_j$, 所以:

$$\alpha_i - \beta_i \in X_i \cap \left(\sum_{j \neq i} X_j \right) = \mathbf{0}$$

于是 $\alpha_i = \beta_i$, 矛盾, 因此和 $X_1 + X_2 + \cdots + X_n$ 是直和。

$1 \Rightarrow 4$: 因为和 $X_1 + X_2 + \cdots + X_n$ 是直和, 由 (3) 可得 $X_i \cap \left(\sum_{j \neq i} X_j \right) = \mathbf{0}$, $i, j = 1, 2, \dots, n$ 。由定理 1.11 可得:

$$\begin{aligned} \dim \left(\sum_{i=1}^n X_i \right) &= \dim \left(X_1 + \sum_{i=2}^n X_i \right) = \dim(X_1) + \dim \left(\sum_{i=2}^n X_i \right) - \dim \left(X_1 \cap \sum_{i=2}^n X_i \right) \\ &= \dim(X_1) + \dim \left(\sum_{i=2}^n X_i \right) \end{aligned}$$

注意到:

$$X_2 \cap \sum_{i=3}^n X_i \subset X_2 \cap \left(X_1 + \sum_{i=3}^n X_i \right) = \mathbf{0}$$

所以:

$$\dim \left(\sum_{i=2}^n X_i \right) = \dim(X_2) + \dim \left(\sum_{i=3}^n X_i \right)$$

由数学归纳法可得:

$$\dim(X_1 + X_2 + \cdots + X_n) = \dim(X_1) + \dim(X_2) + \cdots + \dim(X_n)$$

$4 \Rightarrow 5$: 在 X_i 中取一组基 $\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ir_i}$, $i = 1, 2, \dots, n$, 由引理 1.1 和性质 1.1.7(6) 可得:

$$\begin{aligned} \sum_{i=1}^n X_i &= \langle \alpha_{11}, \alpha_{12}, \dots, \alpha_{1r_1} \rangle + \langle \alpha_{21}, \alpha_{22}, \dots, \alpha_{2r_2} \rangle + \cdots + \langle \alpha_{n1}, \alpha_{n2}, \dots, \alpha_{nr_n} \rangle \\ &= \langle \alpha_{11}, \alpha_{12}, \dots, \alpha_{1r_1}, \alpha_{21}, \alpha_{22}, \dots, \alpha_{2r_2}, \dots, \alpha_{n1}, \alpha_{n2}, \dots, \alpha_{nr_n} \rangle \end{aligned}$$

因为:

$$\dim(X_1 + X_2 + \cdots + X_n) = \dim(X_1) + \dim(X_2) + \cdots + \dim(X_n) = \sum_{i=1}^n r_i$$

所以 $\alpha_{ij}, i = 1, 2, \dots, n, j = 1, 2, \dots, r_i$ 线性无关, 否则的话, 由性质 1.1.7(4) 可得 $\dim(X_1 + X_2 + \cdots + X_n) < \sum_{i=1}^n r_i$, 矛盾。由性质 1.1.6(2) 可知 $\alpha_{ij}, i = 1, 2, \dots, n, j = 1, 2, \dots, r_i$ 是 $X_1 + X_2 + \cdots + X_n$ 的一组基, 即 $X_i, i = 1, 2, \dots, n$ 的基合起来是和 $X_1 + X_2 + \cdots + X_n$ 的一组基。

5 \Rightarrow 1: 在 X_i 中取一组基 $\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ir_i}, i = 1, 2, \dots, n$, 则 $\alpha_{ij}, i = 1, 2, \dots, n, j = 1, 2, \dots, r_i$ 是 $X_1 + X_2 + \cdots + X_n$ 的一组基。设:

$$\mathbf{0} = \alpha_1 + \alpha_2 + \cdots + \alpha_n, \alpha_i \in X_i$$

则有:

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n = \sum_{i=1}^n \sum_{j=1}^{r_i} k_{ij} \alpha_{ij} = \mathbf{0}$$

$k_{ij} \in F$ 。因为 $\alpha_{ij}, i = 1, 2, \dots, n, j = 1, 2, \dots, r_i$ 是 $X_1 + X_2 + \cdots + X_n$ 的一组基, 所以 $k_{ij} = 0, i = 1, 2, \dots, n, j = 1, 2, \dots, r_i$, 于是 $\alpha_i = \mathbf{0}$, 即和 $X_1 + X_2 + \cdots + X_n$ 中零向量的表示方法唯一, 由 (2) 可得和 $X_1 + X_2 + \cdots + X_n$ 是直和。□

Theorem 1.13. 设 X 是域 F 上的线性空间, 则 X 的任一子空间 E 都有补空间。

Proof. (1) $\dim(X) = n < +\infty$: 取 E 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_m$, 把它扩充为 X 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_m, \beta_1, \beta_2, \dots, \beta_{n-m}$, 由引理 1.1 和性质 1.1.7(6) 可知:

$$\begin{aligned} X &= \langle \alpha_1, \alpha_2, \dots, \alpha_m, \beta_1, \beta_2, \dots, \beta_{n-m} \rangle \\ &= \langle \alpha_1, \alpha_2, \dots, \alpha_m \rangle + \langle \beta_1, \beta_2, \dots, \beta_{n-m} \rangle = E + W \end{aligned}$$

其中 $W = \langle \beta_1, \beta_2, \dots, \beta_{n-m} \rangle$ 。因为 $\alpha_1, \alpha_2, \dots, \alpha_m, \beta_1, \beta_2, \dots, \beta_{n-m}$ 线性无关, 由性质 1.1.2(1) 可知 $\beta_1, \beta_2, \dots, \beta_{n-m}$ 线性无关, 所以 $\beta_1, \beta_2, \dots, \beta_{n-m}$ 是 W 的一组基。于是 E 的一组基和 W 的一组基合起来就是 X 的一组基。由定理 1.12(5) 可知 $X = E \oplus W$, 于是 W 是 E 的补空间。

(2) $\dim(X) = +\infty$: 考虑商空间 X/E , 设其一组基为 $\alpha_i + E, i \in I$, 其中 I 是一个指标集。由推论 1.2 可知 $\alpha_i, i \in I$ 线性无关。令:

$$W = \left\{ \sum_{i=1}^n k_i \alpha_i : k_i \in F, n \in \mathbb{N}^+ \right\}$$

显然 $\alpha_i, i \in I$ 是 W 的一组基。下面证明 W 是 E 的一个补空间。

任取 $\alpha \in X$, 因为 $\alpha_i + E, i \in I$ 是 X/E 的一组基, 所以:

$$\alpha + E = \sum_{i=1}^m l_i \alpha_i + E, l_i \in F$$

即:

$$\alpha - \sum_{i=1}^m l_i \alpha_i \in E$$

于是 α 可以表示为 E 和 W 中两个元素的和。由 α 的任意性, $X = E + W$ 。

任取 $\beta \in W \cap E$, 因为 $\beta \in W$, 所以:

$$\beta = \sum_{i=1}^t p_i a_i, p_i \in F$$

又因为 $\beta \in E$, 所以:

$$E = \beta + E = \sum_{i=1}^t p_i a_i + E = \sum_{i=1}^t p_i (a_i + E)$$

因为 $\alpha_i + E, i \in I$ 线性无关, 所以 $p_i = 0, i = 1, 2, \dots, t$ (E 是 X/E 的零元), 于是 $\beta = \mathbf{0}$ 。
由定理 1.12(3) 可知 $X = E \oplus W$, W 是 E 的补空间。

综上, X 的任一子空间 E 都有补空间。 □

1.1.4 线性空间的同构

Definition 1.21. 设 X, Y 为域 F 上的线性空间。如果存在 X 到 Y 的一个双射 σ , 使得对于任意的 $\alpha, \beta \in X, k_1, k_2 \in F$, 有:

$$\sigma(k_1 \alpha + k_2 \beta) = k_1 \sigma(\alpha) + k_2 \sigma(\beta)$$

则称 σ 是 X 到 Y 的一个同构映射 (*isomorphism*), 此时称 X 与 Y 同构 (*isomorphic*), 记作 $X \cong Y$ 。

Property 1.1.8. 设 X, Y 为域 F 上的线性空间, 且 X, Y 同构, σ 是 X 到 Y 的同构映射, 则:

1. $\sigma(\mathbf{0}_X) = \mathbf{0}_Y$;
2. 对于任意的 $\alpha \in X$, 有 $\sigma(-\alpha) = -\sigma(\alpha)$;
3. 对于任意的 $\alpha_1, \alpha_2, \dots, \alpha_n \in X, k_1, k_2, \dots, k_n \in F$, 有:

$$\sigma\left(\sum_{i=1}^n k_i \alpha_i\right) = \sum_{i=1}^n k_i \sigma(\alpha_i)$$

4. $\alpha_1, \alpha_2, \dots, \alpha_n \in X$ 线性相关当且仅当 $\sigma(\alpha_1), \sigma(\alpha_2), \dots, \sigma(\alpha_n)$ 线性相关;
5. 如果 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 X 的一组基, 则 $\sigma(\alpha_1), \sigma(\alpha_2), \dots, \sigma(\alpha_n)$ 是 Y 的一组基;
6. 若 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 X 的一组基, 则 $\alpha \in X$ 在基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的坐标和 $\sigma(\alpha) \in Y$ 在基 $\sigma(\alpha_1), \sigma(\alpha_2), \dots, \sigma(\alpha_n)$ 下的坐标相同。

7. 若 E 是 X 的一个子空间, 则 $\sigma(E)$ 是 Y 的一个子空间。若 $\dim(E) = n < +\infty$, 则 $\dim \sigma(E) = n$;

8. 线性空间的同构是一个等价关系, 其等价类被称为同构类。

Proof. (1) 任取 $\alpha \in X$, 由性质 1.1.1(3) 和同构映射的定义可得 $\sigma(\mathbf{0}_X) = \sigma(0\alpha) = 0\sigma(\alpha) = \mathbf{0}_Y$ 。

(2) 由性质 1.1.1(6) 可知 $\sigma(-\alpha) = \sigma[(-1)\alpha] = (-1)\sigma(\alpha) = -\sigma(\alpha)$ 。

(3) 由同构映射的定义直接可得。

(4) 设:

$$\sum_{i=1}^n k_i \alpha_i = \mathbf{0}_X, \quad k_i \in F$$

由 (3) 和 (1) 可得:

$$\sum_{i=1}^n k_i \sigma(\alpha_i) = \sigma(\mathbf{0}_X) = \mathbf{0}_Y$$

若 k_i 不全为 0, 即 $\alpha_1, \alpha_2, \dots, \alpha_n \in X$ 或 $\sigma(\alpha_1), \sigma(\alpha_2), \dots, \sigma(\alpha_n)$ 线性相关, 显然此时另一个也线性相关。

(5) 由 (4) 可得 $\sigma(\alpha_1), \sigma(\alpha_2), \dots, \sigma(\alpha_n)$ 线性无关。任取 $\beta \in Y$, 因为 σ 是一个满射, 则存在 $\alpha \in X$ 使得 $\sigma(\alpha) = \beta$ 。因为 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 X 的一组基, 所以存在 $k_1, k_2, \dots, k_n \in F$ 使得:

$$\alpha = \sum_{i=1}^n k_i \alpha_i$$

由 (3) 可得:

$$\beta = \sigma(\alpha) = \sigma\left(\sum_{i=1}^n k_i \alpha_i\right) = \sum_{i=1}^n k_i \sigma(\alpha_i)$$

于是 β 可由 $\sigma(\alpha_i), i = 1, 2, \dots, n$ 线性表出。由 β 的任意性, $\sigma(\alpha_i), i = 1, 2, \dots, n$ 是 Y 的一组基。

(6) 由 (3)(5) 直接得到。

(7) 任取 $\alpha, \beta \in \sigma(E), k_1, k_2 \in F$, 考虑 $k_1\alpha + k_2\beta$ 。因为 σ 是 X 到 Y 的一个双射, 所以对于 α, β , 存在 $a, b \in X$ 满足 $\sigma(a) = \alpha, \sigma(b) = \beta$ 。因为 E 是 X 的一个子空间, 所以 $k_1a + k_2b \in E$, 于是 $\sigma(k_1a + k_2b) = k_1\alpha + k_2\beta \in \sigma(E)$, 所以 $\sigma(E)$ 是 Y 的一个子空间。由 (5) 可直接得到有限维情况下 E 与 $\sigma(E)$ 之间的维数关系。

(8) 反身性由恒等映射保证, 对称性由双射保证, 传递性由复合映射可直接得到。□

Theorem 1.14. 设 X, Y 为域 F 上的有限维线性空间, 则 X 与 Y 同构的充分必要条件为它们的维数相同, 于是维数是有限维线性空间同构类的完全不变量。

Proof. (1) **必要性:** 由性质 1.1.8(5) 直接得到。

(2) **充分性:** 设二者维数都是 n , 取 X 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 和 Y 的一组基 $\beta_1, \beta_2, \dots, \beta_n$ 。令:

$$\sigma: \sum_{i=1}^n k_i \alpha_i \longrightarrow \sum_{i=1}^n k_i \beta_i$$

显然它是一个线性映射并且是一个双射, 于是 X 与 Y 同构。 \square

1.1.5 商空间

商空间的定义

Theorem 1.15. 设 X 是域 F 上的一个线性空间, L 是 E 的子空间。对于 $\forall x, y \in X$, 若 $x - y \in L$, 则称 x, y 是等价的, 记为 $x \sim y$ 。该关系是一个等价关系。对于该关系的任意等价类 α , 任取 $x \in \alpha$, α 可由 $\{x + y : y \in L\}$ 来表示, 简记为 $x + L$ 。

Proof. 任取 $x, y, z \in X$ 。

(1) 因为 $x - x = \mathbf{0} \in L$, 所以该关系满足自反性。

(2) 若 $x \sim y$, 即 $x - y \in L$, 因为 L 是一个子空间, 所以 $y - x = (-1)(x - y) \in L$, 于是 $y \sim x$, 该关系满足对称性。

(3) 若 $x \sim y, y \sim z$, 则有 $x - y \in L, y - z \in L$, 因为 L 是一个线性空间, 所以 $x - z = (x - y) + (y - z) \in L$, 即 $x \sim z$ 。该关系满足传递性。

综上, 该关系是一个等价关系。

下证明表示方法的正确性。

对任意的 $a \in \alpha$, 因为 $x \in \alpha$, 所以 $x - a \in L$, 于是存在 $z \in L$ 使得 $x - a = -z$, 即 $a = x + z$ 。 \square

Theorem 1.16. 设 X 是域 F 上的一个线性空间, L 是 X 的子空间, \hat{X} 为 X 中所有等价类构成的集合, 确定等价类的关系为定理 1.15 中的关系。对任意的 $\alpha, \beta \in \hat{X}, k \in F$, 在 \hat{X} 中定义线性运算如下:

$$\alpha + \beta = x + y + L, x \in \alpha, y \in \beta$$

$$k\alpha = kx + L, x \in \alpha$$

则 \hat{X} 成为一个线性空间, 称其为 X 关于 L 的商空间 (quotient space), 记作 X/L 。

Proof. 先证明上述线性运算与 x, y 的选择无关。

(1) 任取 $x' \in \alpha, y' \in \beta$, 满足 $x \neq x', y \neq y'$ 。于是有:

$$x' + y' + L = x + y + (x' - x) + (y' - y) + L$$

因为 $x, x' \in \alpha, y, y' \in \beta$, 所以 $x' - x, y' - y \in L$, 于是 $x' + y' + L = x + y + L$ 。

(2) 任取 $x' \in \alpha$, 满足 $x \neq x'$ 。于是有:

$$kx' + L = kx + k(x' - x) + L$$

因为 $x, x' \in \alpha$, 所以 $x' - x \in L$ 。因为 L 是线性空间, 所以 $k(x' - x) \in L$, 于是 $kx' + L = kx + L$ 。

下证明 \hat{X} 是一个线性空间。

1. 任取 \hat{X} 中的两个元素 $\alpha = x + L, \beta = y + L$ 。因为 $x, y \in X$, X 是一个线性空间, 所以 $x + y = y + x$, 于是 $\alpha + \beta = x + y + L = y + x + L = \beta + \alpha$ 。由 α, β 的任意性, \hat{X} 上的加法满足线性空间运算法则 (1);
2. 任取 \hat{X} 中的三个元素 $\alpha = x + L, \beta = y + L, \gamma = z + L$ 。因为 $x, y, z \in X$, X 是一个线性空间, 所以 $(x + y) + z = x + (y + z)$, 于是 $(\alpha + \beta) + \gamma = (x + y) + z + L = x + (y + z) + L = \alpha + (\beta + \gamma)$ 。由 α, β, γ 的任意性, \hat{X} 上的加法满足线性空间运算法则 (2);
3. 任取 $\alpha = x + L \in \hat{X}$, 则 $\alpha + \mathbf{0} + L = x + \mathbf{0} + L = x + L = \alpha$, 于是 $\mathbf{0} + L$ 是 \hat{X} 中的零元。 \hat{X} 上的加法满足线性空间运算法则 (3);
4. 任取 $\alpha = x + L \in \hat{X}$, 则 $\alpha + -x + L = x + (-x) + L = \mathbf{0} + L$, 于是 $-x + L$ 是 α 的负元。由 α 的任意性, \hat{X} 上的加法满足线性空间运算法则 (4);
5. 任取 $\alpha = x + L \in \hat{X}$ 。因为 $x \in X$, X 是一个线性空间, 所以 $1x = x$, 于是 $1\alpha = 1x + L = x + L = \alpha$ 。由 α 的任意性, \hat{X} 上的纯量乘法满足线性空间运算法则 (5);
6. 任取 $\alpha = x + L \in \hat{X}, k, l \in F$ 。因为 $x \in X$, X 是一个线性空间, 所以 $(kl)x = k(lx)$, 于是 $(kl)\alpha = (kl)x + L = k(lx) + L = k(l\alpha)$ 。由 α, k, l 的任意性, \hat{X} 上的纯量乘法满足线性空间运算法则 (6);
7. 任取 $\alpha = x + L \in \hat{X}, k, l \in F$ 。因为 $x \in X$, X 是一个线性空间, 所以 $(k+l)x = kx + lx$, 于是 $(k+l)\alpha = (k+l)x + L = kx + lx + L = k\alpha + l\alpha$ 。由 α, k, l 的任意性, \hat{X} 上的纯量乘法满足线性空间运算法则 (7);
8. 任取 \hat{X} 中的两个元素 $\alpha = x + L, \beta = y + L$ 。因为 $x, y \in X$, X 是一个线性空间, 所以 $k(x + y) = ky + kx$, 于是 $k(\alpha + \beta) = k(x + y) + L = kx + ky + L = k\beta + k\alpha$ 。由 α, β, k 的任意性, \hat{X} 上的加法满足线性空间运算法则 (8)。

综上, \hat{X} 是一个线性空间。

□

商空间的性质

Theorem 1.17. 设 X 是域 F 上的一个有限维线性空间, E 是 X 的一个子空间, 则:

$$\dim(X/E) = \dim(X) - \dim(E)$$

Proof. 设 $\dim(X) = n, \dim(E) = m$, 取 E 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_m$, 把它扩充为 X 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 。任取 $\beta + W \in X/E$, 因为 $\beta \in X$, 于是:

$$\beta + E = \sum_{i=1}^n k_i \alpha_i + E = \sum_{i=1}^m k_i \alpha_i + \sum_{i=m+1}^n k_i \alpha_i + E = \sum_{i=m+1}^n k_i \alpha_i + E$$

其中 $k_i \in F, i = 1, 2, \dots, n$ 。上式表明, 对任意的 $\beta + W \in X/E$, 都可以用 $\alpha_i + W, i = m+1, m+2, \dots, n$ 的线性组合表示。下证明它们线性无关。

设:

$$\sum_{i=m+1}^n k_i \alpha_i + E = E$$

则 $\sum_{i=m+1}^n k_i \alpha_i \in E$, 于是存在 $l_j \in F, j = 1, 2, \dots, m$ 使得:

$$\sum_{j=1}^m l_j \alpha_j = \sum_{i=m+1}^n k_i \alpha_i$$

因为 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关, 所以 $k_i = l_j = 0, i = m+1, m+2, \dots, n, j = 1, 2, \dots, m$, 因此 $\alpha_i + W, i = m+1, m+2, \dots, n$ 线性无关, 即它们是 X/E 的一组基, $\dim(X/E) = n - m = \dim(X) - \dim(E)$ 。□

Definition 1.22. 设 X 是域 F 上的一个线性空间, E 是 X 的一个子空间。若 X/E 是有限维的, 则称 $\dim(X/E)$ 是 E 在 X 中的余维数 (codimension), 记作 $\text{codim } E$ 。

标准映射

Definition 1.23. 设 X 是域 F 上的一个线性空间, E 是 X 的一个子空间, 定义映射:

$$\pi: \alpha \longrightarrow \alpha + E, \forall \alpha \in X$$

称之为标准映射 (canonical mapping)。

Property 1.1.9. 设 X 是域 F 上的一个线性空间, E 是 X 的一个子空间, π 是 X 到 X/E 的标准映射, 则 π 具有如下性质:

1. X/E 中的一个元素 $\alpha + E$ 在 π 下的原像是:

$$\{\alpha + \beta: \beta \in E\}$$

2. π 是一个满射。当且仅当 E 是零空间时, π 是一个双射;

3. π 是一个线性映射;

4. 若 $\alpha_1, \alpha_2, \dots, \alpha_n \in X$ 线性相关, 则 $\pi(\alpha_1), \pi(\alpha_2), \dots, \pi(\alpha_n)$ 线性相关。

Proof. (1) 显然:

$$\begin{aligned} \gamma \in \pi^{-1}(\alpha + E) &\iff \gamma + E = \alpha + E \iff \gamma - \alpha \in E \\ &\iff \exists \beta \in E, \gamma = \alpha + \beta \iff \gamma \in \{\alpha + \beta: \beta \in E\} \end{aligned}$$

(2) 满射是显然的结论。当 E 是零空间时, π 是 X 上的恒等变换, 恒等变换显然是双射。当 π 是双射时, π 是一个单射, 即对任何的 $\alpha + E \in X/E$, 有且仅有 $\alpha \in X$ 使得 $\pi(\alpha) = \alpha + E$, 也即不存在 $\beta \in X, \beta \neq \alpha$, 使得 $\beta - \alpha \in E$, 此时 E 只能为零空间。

(3) $\forall \alpha, \beta \in X, k_1, k_2 \in F$, 有:

$$\pi(k_1\alpha + k_2\beta) = k_1\alpha + k_2\beta + E = k_1\alpha + E + k_2\alpha + E = k_1\pi(\alpha) + k_2\pi(\beta)$$

于是 π 是一个线性映射。

(4) 因为 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性相关, 所以存在不全为零的 $k_1, k_2, \dots, k_n \in F$ 使得:

$$\sum_{i=1}^n k_i \alpha_i = \mathbf{0}$$

由 (3) 可得:

$$\pi\left(\sum_{i=1}^n k_i \alpha_i\right) = \sum_{i=1}^n k_i \pi(\alpha_i) = \sum_{i=1}^n k_i \alpha_i + E = E$$

于是 $\pi(\alpha_i), i = 1, 2, \dots, n$ 线性相关。 \square

Corollary 1.2. 设 X 是域 F 上的一个线性空间, E 是 X 的一个子空间。若 $\{\alpha_i + W : i \in I\}$ 是 X/E 的一组基, I 是一个指标集, 则 $\alpha_i, i \in I$ 线性无关。

Proof. 就 X/E 是无限维的给出证明, 有限维情况类似。

若此时 $\alpha_i, i \in I$ 线性相关, 则存在 $\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_n}$ 线性相关, 其中 $n \in \mathbb{N}^+$ 。由性质 1.1.9(4) 可得 $\alpha_{i_j} + W, j = 1, 2, \dots, n$ 线性相关, 于是 $\{\alpha_i + W : i \in I\}$ 线性相关, 矛盾。 \square

1.2 线性映射

1.2.1 线性映射的定义与基本性质

Definition 1.24. 设 X, Y 是域 F 上的线性空间, X 到 Y 上的一个映射 \mathcal{T} 如果对任意的 $\alpha, \beta \in X$ 和任意的 $k_1, k_2 \in F$, 有:

$$\mathcal{T}(k_1\alpha + k_2\beta) = k_1\mathcal{T}\alpha + k_2\mathcal{T}\beta$$

则称 \mathcal{T} 是 X 到 Y 的一个线性映射 (*linear mapping*)。若 $Y = X$, 则称 \mathcal{T} 为 X 上的线性变换 (*linear transformation*); 若 $Y = F$, 则称 \mathcal{T} 为 X 上的线性函数 (*linear function*)。

线性映射空间与线性映射的运算

Definition 1.25. 设 X, Y 是域 F 上的线性空间, 将 X 到 Y 的所有线性映射组成的集合记为 $\text{Hom}(X, Y)$ 。设 $\mathcal{T}_1, \mathcal{T}_2 \in \text{Hom}(X, Y), k \in F$, 定义线性映射的加法与纯量乘法如下:

$$(\mathcal{T}_1 + \mathcal{T}_2)\alpha = \mathcal{T}_1\alpha + \mathcal{T}_2\alpha, \forall \alpha \in X$$

$$(k\mathcal{T}_1)\alpha = k\mathcal{T}_1\alpha, \forall \alpha \in X$$

容易验证 $\text{Hom}(X, Y)$ 成为域 F 上的一个线性空间。 $X = Y$ 时将 $\text{Hom}(X, Y)$ 简记为 $\text{Hom}(X)$ 。

Definition 1.26. 设 X, Y 是域 F 上的线性空间, $\mathcal{T}_1, \mathcal{T}_2 \in \text{Hom}(X, Y)$, 定义线性映射的减法如下:

$$\mathcal{T}_1 - \mathcal{T}_2 = \mathcal{T}_1 + (-\mathcal{T}_2)$$

Definition 1.27. 设 X, Y 是域 F 上的线性空间, $\mathcal{T} \in \text{Hom}(X, Y)$, 若对任意的 $\alpha \in X$, 有 $\mathcal{T}\alpha = \mathbf{0}_Y$, 则称 \mathcal{T} 为 X 到 Y 的零映射 (zero mapping), 记作 \mathcal{O} 。

Definition 1.28. 设 X, Y, Z 是域 F 上的线性空间, $\mathcal{T}_1 \in \text{Hom}(X, Y), \mathcal{T}_2 \in \text{Hom}(Z, Y)$, 定义线性映射乘法如下:

$$\forall \alpha \in X, (\mathcal{T}_2 \mathcal{T}_1)\alpha = \mathcal{T}_2(\mathcal{T}_1\alpha)$$

Definition 1.29. 设 X, Y 是域 F 上的线性空间, $\mathcal{T}_1 \in \text{Hom}(X, Y)$, 若存在 $\mathcal{T}_2 \in \text{Hom}(Y, X)$ 使得 $\mathcal{T}_1 \mathcal{T}_2 = \mathcal{I}_Y, \mathcal{T}_2 \mathcal{T}_1 = \mathcal{I}_X$, 则称 \mathcal{T}_1 是可逆的, \mathcal{T}_2 是 \mathcal{T}_1 的逆映射。

Theorem 1.18. 设 X, Y, Z 是域 F 上的线性空间, $\mathcal{T}_1 \in \text{Hom}(X, Y), \mathcal{T}_2 \in \text{Hom}(Z, Y)$, 则 $\mathcal{T}_2 \mathcal{T}_1 \in \text{Hom}(X, Z)$ 。

Proof. 只需注意到对任意的 $\alpha, \beta \in X, k_1, k_2 \in F$, 有:

$$\begin{aligned} (\mathcal{T}_2 \mathcal{T}_1)(k_1\alpha + k_2\beta) &= \mathcal{T}_2[\mathcal{T}_1(k_1\alpha + k_2\beta)] = \mathcal{T}_2(k_1\mathcal{T}_1\alpha + k_2\mathcal{T}_1\beta) \\ &= k_1\mathcal{T}_2(\mathcal{T}_1\alpha) + k_2\mathcal{T}_2(\mathcal{T}_1\beta) = k_1(\mathcal{T}_2 \mathcal{T}_1)\alpha + k_2(\mathcal{T}_2 \mathcal{T}_1)\beta \quad \square \end{aligned}$$

Definition 1.30. 设 X, Y 是域 F 上的线性空间, \mathcal{T} 是 X 到 Y 上的一个线性映射, 分别称:

$$\{\alpha \in X : \mathcal{T}\alpha = \mathbf{0}_Y\}, \quad \{\mathcal{T}\alpha : \alpha \in X\}$$

为 \mathcal{T} 的核 (kernel) 与象 (image), 将它们分别记作 $\text{Ker } \mathcal{T}$ 和 $\text{Im } \mathcal{T}$ 。

线性空间的性质

Property 1.2.1. 设 X, Y 是域 F 上的线性空间, \mathcal{T} 是 X 到 Y 上的线性映射, 则:

1. 若 \mathcal{T} 可逆, 则 \mathcal{T} 是 X 到 Y 上的同构映射;
2. $\mathcal{T}\mathbf{0}_X = \mathbf{0}_Y$;
3. 对于任意的 $\alpha \in X$, 有 $\mathcal{T}(-\alpha) = -\mathcal{T}\alpha$;
4. 对于任意的 $\alpha_1, \alpha_2, \dots, \alpha_n \in X, k_1, k_2, \dots, k_n \in F$, 有:

$$\mathcal{T}\left(\sum_{i=1}^n k_i \alpha_i\right) = \sum_{i=1}^n k_i \mathcal{T}\alpha_i$$

这表明, 如果 X 是有限维的, 那么只要知道 X 的一组基在 \mathcal{T} 下的象, 那么 X 中所有向量在 \mathcal{T} 下的象就都确定了;

5. 若 $\alpha_1, \alpha_2, \dots, \alpha_n \in X$ 线性相关, 则 $\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n$ 线性相关;

6. $\text{Ker } \mathcal{T}$ 和 $\text{Im } \mathcal{T}$ 分别是 X 和 Y 的子空间;

7. \mathcal{T} 是单射当且仅当 $\text{Ker } \mathcal{T} = \mathbf{0}_X$;

8. \mathcal{T} 是满射当且仅当 $\text{Im } \mathcal{T} = Y$ 。

9. $X/\text{Ker } \mathcal{T}$ 与 $\text{Im } \mathcal{T}$ 在映射:

$$\sigma: \alpha + \text{Ker } \mathcal{T} \longrightarrow \mathcal{T}\alpha$$

下同构;

10. 若 X 是有限维的, 则 $\text{Ker } \mathcal{T}$ 和 $\text{Im } \mathcal{T}$ 都是有限维的, 且有:

$$\dim(X) = \dim(\text{Ker } \mathcal{T}) + \dim(\text{Im } \mathcal{T})$$

11. 若 $\dim(X) = \dim Y = n < +\infty$, 则 \mathcal{T} 是单射当且仅当 \mathcal{T} 是满射;

12. 设 $\alpha_1, \alpha_2, \dots, \alpha_n \in X$, 则有 $\mathcal{T} \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle = \langle \mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n \rangle$;

Proof. (1)(2)(3)(4)(5)(8) 证明都是显然的, 只需参考性质 1.1.8 即可, 这是因为线性映射只比同构映射少了双射这一条件, 所以同构映射不涉及双射条件的性质对于线性映射也成立。

(6) 任取 $\alpha, \beta \in \text{Ker } \mathcal{T}$ 和 $k_1, k_2 \in F$, 则有:

$$\mathcal{T}(k_1\alpha + k_2\beta) = k_1\mathcal{T}\alpha + k_2\mathcal{T}\beta = \mathbf{0}$$

于是 $k_1\alpha + k_2\beta \in \text{Ker } \mathcal{T}$, 所以 $\text{Ker } \mathcal{T}$ 是 X 的子空间。

任取 $\mathcal{T}\alpha, \mathcal{T}\beta \in \text{Im } \mathcal{T}$ 和 $k_1, k_2 \in F$, 则有:

$$k_1\mathcal{T}\alpha + k_2\mathcal{T}\beta = \mathcal{T}(k_1\alpha + k_2\beta)$$

因为 X 是一个线性空间, 所以 $k_1\alpha + k_2\beta \in X$, 于是 $\mathcal{T}(k_1\alpha + k_2\beta) \in \text{Im } \mathcal{T}$, 因此 $\text{Im } \mathcal{T}$ 是 Y 的子空间。

(7) **充分性:** 假设此时 \mathcal{T} 不是单射, 则存在 $\mathcal{T}\alpha, \mathcal{T}\beta \in \mathcal{T}$ 使得 $\mathcal{T}\alpha = \mathcal{T}\beta$ 且 $\alpha \neq \beta$, 而此时 $\mathcal{T}\alpha - \mathcal{T}\beta = \mathcal{T}(\alpha - \beta) = \mathbf{0}_Y$, 由已知条件可得 $\alpha - \beta = \mathbf{0}_X$, 即 $\alpha = \beta$, 矛盾。

必要性: 由 (2) 可知 $\mathcal{T}\mathbf{0}_X = \mathbf{0}_Y$, 因为 \mathcal{T} 是一个单射, 所以 $\text{Ker } \mathcal{T} = \mathbf{0}_X$ 。

(9) 先证明 σ 是一个映射。若 $\alpha + \text{Ker } \mathcal{T} = \beta + \text{Ker } \mathcal{T}$, 则 $\alpha - \beta \in \text{Ker } \mathcal{T}$, 即 $\mathcal{T}(\alpha - \beta) = \mathcal{T}\alpha - \mathcal{T}\beta = \mathbf{0}_Y$, 于是 $\mathcal{T}\alpha = \mathcal{T}\beta$, 所以 σ 是一个映射。

任取 $\alpha + \text{Ker } \mathcal{T}, \beta + \text{Ker } \mathcal{T} \in X/\text{Ker } \mathcal{T}$ 和 $k_1, k_2 \in F$, 则有:

$$\begin{aligned} \sigma[k_1(\alpha + \text{Ker } \mathcal{T}) + k_2(\beta + \text{Ker } \mathcal{T})] &= \sigma(k_1\alpha + k_2\beta + \text{Ker } \mathcal{T}) = \mathcal{T}(k_1\alpha + k_2\beta) \\ &= k_1\mathcal{T}\alpha + k_2\mathcal{T}\beta = k_1\sigma(\alpha + \text{Ker } \mathcal{T}) + k_2\sigma(\beta + \text{Ker } \mathcal{T}) \end{aligned}$$

所以 σ 是一个线性映射。

显然 σ 是一个满射。

若存在 $\alpha + \text{Ker } \mathcal{T}, \beta + \text{Ker } \mathcal{T} \in X$ 满足 $\alpha + \text{Ker } \mathcal{T} \neq \beta + \text{Ker } \mathcal{T}$ 且 $\mathcal{T}\alpha = \mathcal{T}\beta$, 则此时有 $\mathcal{T}(\alpha - \beta) = \mathcal{T}\alpha - \mathcal{T}\beta = \mathbf{0}_Y$, 所以 $\alpha - \beta \in \text{Ker } \mathcal{T}$, 即 $\alpha + \text{Ker } \mathcal{T} = \beta + \text{Ker } \mathcal{T}$, 矛盾, 因此 σ 是个单射。

综上, σ 是一个双射且是一个线性映射, 于是 $X/\text{Ker } \mathcal{T}$ 与 $\text{Im } \mathcal{T}$ 在 σ 下同构。

(10) 因为 X 是有限维的, 由定理 1.17 可知 $X/\text{Ker } \mathcal{T}$ 和 $\text{Ker } \mathcal{T}$ 都是有限维的。由定理 1.14 和 (9) 可知 $\dim(X/\text{Ker } \mathcal{T}) = \dim(\text{Im } \mathcal{T})$, 于是 $\text{Im } \mathcal{T}$ 也是有限维的。由定理 1.17 可得:

$$\dim(\text{Im } \mathcal{T}) = \dim(X/\text{Ker } \mathcal{T}) = \dim(X) - \dim(\text{Ker } \mathcal{T})$$

(11) 由 (7)(10)(6) 和定理 1.8 可得:

$$\begin{aligned} \mathcal{T} \text{ 是单射} &\iff \text{Ker } \mathcal{T} = \mathbf{0}_X \iff \dim(\text{Ker } \mathcal{T}) = 0 \\ &\iff \dim Y = \dim(X) = \dim(\text{Im } \mathcal{T}) \iff \mathcal{T} \text{ 是满射} \end{aligned}$$

(12) 由 (4) 可得:

$$\begin{aligned} \langle \mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n \rangle &= \left\langle \sum_{i=1}^n c_i \mathcal{T}\alpha_i : c_i \in F \right\rangle = \left\langle \mathcal{T} \left(\sum_{i=1}^n c_i \alpha_i \right) : c_i \in F \right\rangle \\ &= \mathcal{T} \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle \end{aligned} \quad \square$$

Definition 1.31. 设 X, Y 是域 F 上的线性空间, X 是有限维的, $\mathcal{T} \in \text{Hom}(X, Y)$, 由性质 1.2.1(10) 可知 $\text{Ker}(\mathcal{T}), \text{Im}(\mathcal{T})$ 都是有限维的, 称 $\dim(\text{Ker } \mathcal{T})$ 为 \mathcal{T} 的零度 (nullity), 称 $\dim(\text{Im } \mathcal{T})$ 为 \mathcal{T} 的秩, 记为 $\text{rank}(\mathcal{T})$ 。

Definition 1.32. 设 X, Y 是域 F 上的线性空间, $\mathcal{T} \in \text{Hom}(X, Y)$, 称 $Y/\text{Im } \mathcal{T}$ 为 \mathcal{T} 的余核 (cokernel)。

Theorem 1.19. 设 X, Y 是域 F 上的线性空间, $\mathcal{T} \in \text{Hom}(X, Y)$, 则 \mathcal{T} 是满射当且仅当 $\text{Coker } \mathcal{T} = \mathbf{0}$ 。

Proof. \mathcal{T} 是满射 $\iff \text{Im } \mathcal{T} = Y \iff Y/\text{Im } \mathcal{T} = \mathbf{0}$ 。这里的 $\mathbf{0}$ 实际上是商空间的零元, 也即 Y 。 \square

1.2.2 线性映射的矩阵表示

Definition 1.33. 设 X, Y 分别为域 F 上的 m 维、 n 维线性空间, \mathcal{T} 是 X 到 Y 的一个线性映射。由性质 1.2.1(4) 可知 \mathcal{T} 被它在 X 的一组基上的作用所决定。取 X 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_m$ 和 Y 的一组基 $\beta_1, \beta_2, \dots, \beta_n$, 则:

$$(\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_m) = (\beta_1, \beta_2, \dots, \beta_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$

将 $(\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_m)$ 记作 $\mathcal{T}(\alpha_1, \alpha_2, \dots, \alpha_m)$, 将上式右端矩阵记为 A , 称 A 为 \mathcal{T} 在 X 的基 $\alpha_1, \alpha_2, \dots, \alpha_m$ 和 Y 的基 $\beta_1, \beta_2, \dots, \beta_n$ 下的矩阵。若 $X = Y$, 取 $(\beta_1, \beta_2, \dots, \beta_n)$ 为 $(\alpha_1, \alpha_2, \dots, \alpha_m)$, 称 A 为 \mathcal{T} 在基 $\alpha_1, \alpha_2, \dots, \alpha_m$ 下的矩阵。

Theorem 1.20. 设 X, Y 分别为域 F 上的 m 维、 n 维线性空间, 则映射 $\sigma: \mathcal{T} \in \text{Hom}(X, Y) \longrightarrow \mathcal{T}$ 在 X 的一组基和 Y 的一组基下的矩阵 A 是 $\text{Hom}(X, Y)$ 到 $M_{n \times m}(F)$ 的一个同构映射, 于是有:

$$\text{Hom}(X, Y) \cong M_{n \times m}(F), \quad \dim[\text{Hom}(X, Y)] = \dim(X) \dim(Y) = mn$$

Proof. 任取 X 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_m$ 和 Y 的一组基 $\beta_1, \beta_2, \dots, \beta_n$ 。

任意的 $\mathcal{T} \in \text{Hom}(X, Y)$ 都存在对应的 $A \in M_{n \times m}(F)$ 使得:

$$(\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_m) = (\beta_1, \beta_2, \dots, \beta_n)A$$

因为 $\beta_1, \beta_2, \dots, \beta_n$ 是 Y 的一组基, 由性质 1.1.3(1) 可知 $\mathcal{T}\alpha_i$ 由 $\beta_1, \beta_2, \dots, \beta_n$ 表出的方式唯一, 所以 A 是唯一的。

任取 $B = (b_{ij}) \in M_{n \times m}(F)$, 令:

$$\mathcal{T}: \alpha = \sum_{i=1}^m a_i \alpha_i \in X \longrightarrow \sum_{i=1}^m a_i \sum_{j=1}^n b_{ji} \beta_j \in Y$$

因为 $\alpha_1, \alpha_2, \dots, \alpha_m$ 是 X 的一组基, 由性质 1.1.3(1) 可知 α 由 $\alpha_1, \alpha_2, \dots, \alpha_m$ 表出的方式唯一, 所以 \mathcal{T} 是一个映射。显然 \mathcal{T} 是一个线性映射。因为:

$$(\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_m) = (\beta_1, \beta_2, \dots, \beta_n)B$$

所以 B 是 \mathcal{T} 在 X 的基 $\alpha_1, \alpha_2, \dots, \alpha_m$ 和 Y 的基 $\beta_1, \beta_2, \dots, \beta_n$ 下的矩阵。因为 \mathcal{T} 满足:

$$\mathcal{T}\alpha_i = \sum_{j=1}^n b_{ji} \beta_j, \quad i = 1, 2, \dots, m$$

且 $\alpha_1, \alpha_2, \dots, \alpha_m$ 是 X 的一组基, 由性质 1.2.1(4) 可知 B 所对应的 \mathcal{T} 是唯一的。

综上, 存在 $\text{Hom}(X, Y)$ 到 $M_{n \times m}(K)$ 上的一个双射 σ 。

任取 $\mathcal{T}_1, \mathcal{T}_2 \in \text{Hom}(X, Y)$, $\sigma(\mathcal{T}_1) = C$, $\sigma(\mathcal{T}_2) = D$, 则对任意的 $k_1, k_2 \in F$ 有:

$$\begin{aligned} & [(k_1 \mathcal{T}_1 + k_2 \mathcal{T}_2)\alpha_1, (k_1 \mathcal{T}_1 + k_2 \mathcal{T}_2)\alpha_2, \dots, (k_1 \mathcal{T}_1 + k_2 \mathcal{T}_2)\alpha_m] \\ &= (k_1 \mathcal{T}_1 \alpha_1 + k_2 \mathcal{T}_2 \alpha_1, k_1 \mathcal{T}_1 \alpha_2 + k_2 \mathcal{T}_2 \alpha_2, \dots, k_1 \mathcal{T}_1 \alpha_m + k_2 \mathcal{T}_2 \alpha_m) \\ &= (k_1 \mathcal{T}_1 \alpha_1, k_1 \mathcal{T}_1 \alpha_2, \dots, k_1 \mathcal{T}_1 \alpha_m) + (k_2 \mathcal{T}_2 \alpha_1, k_2 \mathcal{T}_2 \alpha_2, \dots, k_2 \mathcal{T}_2 \alpha_m) \\ &= k_1 (\mathcal{T}_1 \alpha_1, \mathcal{T}_1 \alpha_2, \dots, \mathcal{T}_1 \alpha_m) + k_2 (\mathcal{T}_2 \alpha_1, \mathcal{T}_2 \alpha_2, \dots, \mathcal{T}_2 \alpha_m) \\ &= k_1 (\beta_1, \beta_2, \dots, \beta_n)C + k_2 (\beta_1, \beta_2, \dots, \beta_n)D = (\beta_1, \beta_2, \dots, \beta_n)(k_1 C + k_2 D) \end{aligned}$$

即:

$$\sigma(k_1 \mathcal{T}_1 + k_2 \mathcal{T}_2) = k_1 C + k_2 D = k_1 \sigma(\mathcal{T}_1) + k_2 \sigma(\mathcal{T}_2)$$

所以 σ 是一个线性映射。

综上, 映射 $\sigma: \mathcal{T} \in \text{Hom}(X, Y) \rightarrow \mathcal{T}$ 在 X 的一组基和 Y 的一组基下的矩阵 A 是 $\text{Hom}(X, Y)$ 到 $M_{n \times m}(F)$ 的一个同构映射, $\text{Hom}(X, Y) \cong M_{n \times m}(F)$, 由定理 1.14 可得 $\dim[\text{Hom}(X, Y)] = mn = \dim(X) \dim(Y)$ 。 \square

向量在线性映射下象的坐标

Theorem 1.21. 设 X, Y 分别为域 F 上的 m 维、 n 维线性空间, \mathcal{T} 是 X 到 Y 的一个线性映射, \mathcal{T} 在 X 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_m$ 和 Y 的一组基 $\beta_1, \beta_2, \dots, \beta_n$ 下的矩阵为 A , 向量 α 在 $\alpha_1, \alpha_2, \dots, \alpha_m$ 下的坐标为 x , 则 $\mathcal{T}\alpha$ 在 $\beta_1, \beta_2, \dots, \beta_n$ 下的坐标为 Ax 。

Proof. 显然:

$$\mathcal{T}\alpha = \mathcal{T}[(\alpha_1, \alpha_2, \dots, \alpha_m)x] = [\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_m]x = (\beta_1, \beta_2, \dots, \beta_n)Ax \quad \square$$

1.3 线性变换

Definition 1.34. 设 X 是域 F 上的线性空间, $\mathcal{T} \in \text{Hom}(X)$, 定义 \mathcal{T} 的正整数指数幂如下:

$$\mathcal{T}^n = \underbrace{\mathcal{T} \cdot \mathcal{T} \cdots \mathcal{T}}_{n \text{ 个 } \mathcal{T}}, \quad n \in \mathbb{N}^+$$

若 \mathcal{T} 可逆, 还可以定义 \mathcal{T} 的负整数指数幂如下:

$$\mathcal{T}^{-n} = (\mathcal{T}^{-1})^n, \quad n \in \mathbb{N}^+$$

Definition 1.35. 设 X 是域 F 上的线性空间, $\mathcal{T} \in \text{Hom}(X)$ 。

1. 若对任意的 $\alpha \in X$, 有 $\mathcal{T}\alpha = \alpha$, 则称 \mathcal{T} 为 X 到 Y 的恒等变换 (*identity transformation*), 记作 \mathcal{I} ;
2. 给定 $k \in F$, 若对任意的 $\alpha \in X$, 有 $\mathcal{T}\alpha = k\alpha$, 则称 \mathcal{T} 为 X 上的数乘变换 (*scalar transformation*), 记作 \mathcal{K} ;
3. 若 E 和 W 是 X 的子空间, 且有 $X = E \oplus W$, 若对任意的 $\alpha = \alpha_1 + \alpha_2 \in X$, $\alpha_1 \in E$, $\alpha_2 \in W$, 有:

$$\mathcal{T}\alpha = \alpha_1$$

则称 \mathcal{T} 为平行于 W 在 E 上的投影变换 (*projection transformation*), 记为 \mathcal{P}_E 。

4. 若 $\mathcal{T}^2 = \mathcal{I}$, 则称 \mathcal{T} 为对合变换 (*involution*);
5. 若 $\mathcal{T}^2 = \mathcal{T}$, 则称 \mathcal{T} 为幂等变换 (*idempotent transformation*);
6. 若存在 $n \in \mathbb{N}^+$ 使得 $\mathcal{T}^n = \mathcal{O}$, 则称 \mathcal{T} 为幂零变换 (*nilpotent transformation*), 使得 $\mathcal{T}^n = \mathcal{O}$ 成立的最小正整数 n 被称为是 \mathcal{T} 的幂零指数 (*nilpotent index*)。

Definition 1.36. 设 X 是域 F 上的线性空间, $\mathcal{T}_1, \mathcal{T}_2 \in \text{Hom}(X)$ 。若 $\mathcal{T}_1\mathcal{T}_2 = \mathcal{T}_2\mathcal{T}_1 = \mathcal{O}$, 则称 \mathcal{T}_1 和 \mathcal{T}_2 正交。

1.3.1 线性变换的性质

线性变换保持的矩阵运算

Theorem 1.22. 线性变换对其对应的矩阵保持加法、纯量乘法和乘法的运算。

Proof. 设 X 分别为域 F 上的 n 维线性空间, $\sigma: \mathcal{T} \in \text{Hom}(X) \rightarrow \mathcal{T}$ 在 X 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的矩阵。加法和纯量乘法的结论由定理 1.20 可得。取 $\mathcal{T}_1, \mathcal{T}_2 \in \text{Hom}(X)$, $\sigma(\mathcal{T}_1) = A$, $\sigma(\mathcal{T}_2) = B$, 于是有:

$$\begin{aligned} [(\mathcal{T}_1\mathcal{T}_2)\alpha_1, (\mathcal{T}_1\mathcal{T}_2)\alpha_2, \dots, (\mathcal{T}_1\mathcal{T}_2)\alpha_n] &= [\mathcal{T}_1(\mathcal{T}_2\alpha_1), \mathcal{T}_1(\mathcal{T}_2\alpha_2), \dots, \mathcal{T}_1(\mathcal{T}_2\alpha_n)] \\ &= \mathcal{T}_1(\mathcal{T}_2\alpha_1, \mathcal{T}_2\alpha_2, \dots, \mathcal{T}_2\alpha_n) = \mathcal{T}_1[(\alpha_1, \alpha_2, \dots, \alpha_n)B] = (\mathcal{T}_1\alpha_1, \mathcal{T}_1\alpha_2, \dots, \mathcal{T}_1\alpha_n)B \\ &= (\alpha_1, \alpha_2, \dots, \alpha_n)AB \end{aligned}$$

于是 $\sigma(\mathcal{T}_1\mathcal{T}_2) = AB$, 即线性变换对其对应的矩阵保持乘法运算。 \square

线性变换在不同基下的矩阵的关系

Theorem 1.23. 设 X 是域 F 上的 n 维线性空间, $\alpha_1, \alpha_2, \dots, \alpha_n$ 和 $\beta_1, \beta_2, \dots, \beta_n$ 是 X 的两组基, P 为 $\alpha_1, \alpha_2, \dots, \alpha_n$ 到 $\beta_1, \beta_2, \dots, \beta_n$ 的过渡矩阵, $\mathcal{T} \in \text{Hom}(X)$, 则 \mathcal{T} 在 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的矩阵 A 与在 $\beta_1, \beta_2, \dots, \beta_n$ 下的矩阵 B 满足:

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

Proof. 由已知条件可得:

$$\begin{aligned} (\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n) &= (\alpha_1, \alpha_2, \dots, \alpha_n)A \\ (\mathcal{T}\beta_1, \mathcal{T}\beta_2, \dots, \mathcal{T}\beta_n) &= (\beta_1, \beta_2, \dots, \beta_n)B \\ (\beta_1, \beta_2, \dots, \beta_n) &= (\alpha_1, \alpha_2, \dots, \alpha_n)P \end{aligned}$$

设 P 的列向量为 P_1, P_2, \dots, P_n , 所以:

$$\begin{aligned} &(\mathcal{T}\beta_1, \mathcal{T}\beta_2, \dots, \mathcal{T}\beta_n) \\ &= \{\mathcal{T}[(\alpha_1, \alpha_2, \dots, \alpha_n)P_1], \mathcal{T}[(\alpha_1, \alpha_2, \dots, \alpha_n)P_2], \dots, \mathcal{T}[(\alpha_1, \alpha_2, \dots, \alpha_n)P_n]\} \\ &= \{[\mathcal{T}(\alpha_1, \alpha_2, \dots, \alpha_n)]P_1, [\mathcal{T}(\alpha_1, \alpha_2, \dots, \alpha_n)]P_2, \dots, [\mathcal{T}(\alpha_1, \alpha_2, \dots, \alpha_n)]P_n\} \\ &= (\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n)P = (\alpha_1, \alpha_2, \dots, \alpha_n)AP = (\beta_1, \beta_2, \dots, \beta_n)P^{-1}AP \end{aligned}$$

由定理 1.20 可知 \mathcal{T} 在 $\beta_1, \beta_2, \dots, \beta_n$ 下的矩阵是唯一的, 所以 $B = P^{-1}AP$ 。 \square

Theorem 1.24. 设 X 是域 F 上的一个 n 维线性空间, $\mathcal{T} \in \text{Hom}(X)$ 在 X 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的矩阵为 A , 则 $\text{rank}(A) = \text{rank}(\mathcal{T})$ 。

Proof. 由性质 1.2.1(12) 可得:

$$\begin{aligned}\operatorname{rank}(\mathcal{T}) &= \dim(\operatorname{Im} \mathcal{T}) = \dim(\mathcal{T}X) = \dim(\mathcal{T} \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle) \\ &= \dim \langle \mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n \rangle\end{aligned}$$

因为 $\dim(X) = n$, 由定理 1.14 可知 $X \cong F^n$, 映射:

$$\sigma: \sum_{i=1}^n b_i \beta_i \longrightarrow (b_1, b_2, \dots, b_n)^T$$

是 X 到 F^n 的一个同构映射。注意到:

$$(\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n) = (\alpha_1, \alpha_2, \dots, \alpha_n)A$$

$\mathcal{T}\alpha_i$ 在基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的坐标就是 A 的第 i 列, 所以 $\sigma(\mathcal{T}\alpha_i) = A_i$, 其中 A_i 表示 A 的第 i 列, 于是由性质 1.2.1(12) 可得 $\sigma \langle \mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n \rangle = \langle A_1, A_2, \dots, A_n \rangle$, 所以由性质 1.1.8(7) 可得:

$$\begin{aligned}\operatorname{rank}(\mathcal{T}) &= \dim \langle \mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n \rangle = \dim(\sigma \langle \mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n \rangle) \\ &= \dim \langle A_1, A_2, \dots, A_n \rangle = \operatorname{rank}(A)\end{aligned}\quad \square$$

由上述两个定理, 我们可以给出如下定义:

Definition 1.37. 设 X 是域 F 上的 n 维线性空间, $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 X 的一组基, $\mathcal{T} \in \operatorname{Hom}(X)$, A 是 \mathcal{T} 在 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的矩阵 A 。定义:

$$\operatorname{rank}(\mathcal{T}) = \operatorname{rank}(A), \quad \det(\mathcal{T}) = \det(A), \quad \operatorname{tr}(\mathcal{T}) = \operatorname{tr}(A)$$

称 A 的特征多项式为 \mathcal{T} 的特征多项式。

1.3.2 线性变换的特征值与特征向量

Definition 1.38. 设 X 是域 F 上的一个线性空间, $\mathcal{T} \in \operatorname{Hom}(X)$ 。若 X 中存在非零向量 ξ , F 中存在元素 λ , 使得:

$$\mathcal{T}\xi = \lambda\xi$$

则称 λ 是 \mathcal{T} 的一个特征值, ξ 是 \mathcal{T} 属于特征值 λ 的一个特征向量。

Theorem 1.25. 设 X 是域 F 上的一个 n 维线性空间, $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 X 的一组基, $\mathcal{T} \in \operatorname{Hom}(X)$, A 是 \mathcal{T} 在 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的矩阵, 则:

1. λ 是 \mathcal{T} 的特征值 $\iff \lambda$ 是 A 的特征值;
2. ξ 是 \mathcal{T} 属于 λ 的特征向量 $\iff \xi$ 在 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的坐标是 A 属于 λ 的特征向量。

Proof. 设 ξ 在 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的坐标为 x , 由定理 1.21 可知若 $\mathcal{T}\xi = \lambda\xi$ 则有 $Ax = \lambda x$ 。当 $Ax = \lambda x$ 时, $\mathcal{T}\xi$ 与 $\lambda\xi$ 坐标相同, 即二者相等。于是有

$$\mathcal{T}\xi = \lambda\xi \iff Ax = \lambda x \quad \square$$

Property 1.3.1. 设 X 是域 F 上的一个线性空间, $\mathcal{T} \in \text{Hom}(X)$, λ 是 \mathcal{T} 的一个特征值, 则:

1. \mathcal{T} 属于特征值 λ 的特征向量构成 X 的一个子空间, 称该子空间为 \mathcal{T} 属于特征值 λ 的特征子空间, 记为 X_λ ;
2. $X_\lambda = \text{Ker}(\lambda\mathcal{I} - \mathcal{T})$;
3. 若 $\dim(X) = n$, 则 $\dim(X_\lambda) = n - \text{rank}(\lambda\mathcal{I} - \mathcal{T})$;
4. \mathcal{T} 属于不同特征值的特征向量是线性无关的。

Proof. (1) 任取 \mathcal{T} 属于 λ 的两个特征向量 ξ_1, ξ_2 以及域 F 上的任意两个元素 k_1, k_2 , 则:

$$\mathcal{T}(k_1\xi_1 + k_2\xi_2) = k_1\mathcal{T}\xi_1 + k_2\mathcal{T}\xi_2 = k_1\lambda\xi_1 + k_2\lambda\xi_2 = \lambda(k_1\xi_1 + k_2\xi_2)$$

所以 $k_1\xi_1 + k_2\xi_2$ 也是 \mathcal{T} 属于 λ 的特征向量, 由定理 1.7 可知 \mathcal{T} 属于 λ 的特征向量构成 X 的一个子空间。

$$(2) \xi \in X_\lambda \iff \mathcal{T}\xi = \lambda\xi \iff (\mathcal{T} - \lambda\mathcal{I})\xi = \mathbf{0} \iff \xi \in \text{Ker}(\lambda\mathcal{I} - \mathcal{T}).$$

(3) 由定理 1.14, 设 σ 是 X 到 F^n 的同构映射, 它将 X 中的元素映射到在基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的坐标, 记 \mathcal{T} 在 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的矩阵为 A 。由定理 1.25 可知 $\sigma(X_\lambda)$ 等于齐次线性方程组 $(\lambda I_n - A)x = \mathbf{0}$ 的解空间 W , 根据性质 2.3.1(3)、性质 1.1.8(7) 和定理 1.20 可得:

$$\dim(X_\lambda) = \dim[\sigma(X_\lambda)] = \dim(W) = n - \text{rank}(\lambda I_n - A) = n - \text{rank}(\lambda\mathcal{I} - \mathcal{T})$$

(4) 设 σ 是 X 中的向量到它在基 $\xi_1, \xi_2, \dots, \xi_n$ 下的坐标的同构映射。由性质 2.6.1(2)、定理 1.25 和性质 1.1.8(4) 立即可得。 \square

Definition 1.39. 设 X 是域 F 上的一个线性空间, $\mathcal{T} \in \text{Hom}(X)$, λ 是 \mathcal{T} 的一个特征值, 称 λ 作为 \mathcal{T} 的特征多项式的根的重数称为 λ 的代数重数, 将 X_λ 的维数称为 λ 的几何重数。

线性变换的对角化

Definition 1.40. 设 X 是域 F 上的一个 n 维线性空间, $\mathcal{T} \in \text{Hom}(X)$, 若 X 中存在一组基使得 \mathcal{T} 在这组基下的矩阵是对角矩阵, 则称 \mathcal{T} 可对角化。

Theorem 1.26. 设 X 是域 F 上的一个 n 维线性空间, $\mathcal{T} \in \text{Hom}(X)$, \mathcal{T} 可对角化当且仅当:

1. X 中存在一组基使得 \mathcal{T} 在这组基下的矩阵可对角化;
2. \mathcal{T} 有 n 个线性无关的特征向量 $\xi_1, \xi_2, \dots, \xi_n$;
3. X 中存在由 \mathcal{T} 的特征向量构成的一组基;

4. \mathcal{T} 属于不同特征值的特征子空间的维数之和为 n ;
5. X 可表为 \mathcal{T} 属于不同特征值的特征子空间的直和;
6. \mathcal{T} 的特征多项式在 $F[\lambda]$ 中可分解为:

$$(\lambda - \lambda_1)^{r_1}(\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_m)^{r_m}$$

其中 $\lambda_1, \lambda_2, \dots, \lambda_m$ 两两不等且 \mathcal{T} 的每个特征值 λ_i 的几何重数等于它的代数重数, $i = 1, 2, \dots, m$ 。

此时 $(\mathcal{T}\xi_1, \mathcal{T}\xi_2, \dots, \mathcal{T}\xi_n) = (\xi_1, \xi_2, \dots, \xi_n) \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, λ_i 为 ξ_i 所属的特征值, $A = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 被称为 \mathcal{T} 的标准形。除了主对角线上元素的排列顺序外, A 是由 \mathcal{T} 唯一决定的。

Proof. (1) 必要性显然。设 \mathcal{T} 在基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的矩阵 A 可对角化, 即存在可逆矩阵 P 与对角矩阵 Λ 使得 $A = P\Lambda P^{-1}$, 设 P 的列为 P_1, P_2, \dots, P_n , 于是有:

$$\begin{aligned} (\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n) &= (\alpha_1, \alpha_2, \dots, \alpha_n)A \\ (\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n) &= (\alpha_1, \alpha_2, \dots, \alpha_n)P\Lambda P^{-1} \\ (\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n)P &= (\alpha_1, \alpha_2, \dots, \alpha_n)P\Lambda \\ &= [(\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n)P_1, (\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n)P_2, \dots, (\mathcal{T}\alpha_1, \mathcal{T}\alpha_2, \dots, \mathcal{T}\alpha_n)P_n] \Lambda \\ &= [(\alpha_1, \alpha_2, \dots, \alpha_n)P_1, (\alpha_1, \alpha_2, \dots, \alpha_n)P_2, \dots, (\alpha_1, \alpha_2, \dots, \alpha_n)P_n] \Lambda \\ &\quad \{ \mathcal{T}[(\alpha_1, \alpha_2, \dots, \alpha_n)P_1], \mathcal{T}[(\alpha_1, \alpha_2, \dots, \alpha_n)P_2], \dots, \mathcal{T}[(\alpha_1, \alpha_2, \dots, \alpha_n)P_n] \} \\ &= [(\alpha_1, \alpha_2, \dots, \alpha_n)P_1, (\alpha_1, \alpha_2, \dots, \alpha_n)P_2, \dots, (\alpha_1, \alpha_2, \dots, \alpha_n)P_n] \Lambda \end{aligned}$$

由定理 1.5 可知 $(\alpha_1, \alpha_2, \dots, \alpha_n)P_1, (\alpha_1, \alpha_2, \dots, \alpha_n)P_2, \dots, (\alpha_1, \alpha_2, \dots, \alpha_n)P_n$ 还是 X 的一组基, 于是 \mathcal{T} 在这组基下的矩阵是对角矩阵 Λ , \mathcal{T} 可对角化, 充分性得证。

(2) 由 \mathcal{T} 可对角化的定义可得:

$$\begin{aligned} \mathcal{T} \text{ 可对角化} &\iff \mathcal{T} \text{ 在 } X \text{ 的基 } \xi_1, \xi_2, \dots, \xi_n \text{ 下的矩阵为 } \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} \\ &\iff (\mathcal{T}\xi_1, \mathcal{T}\xi_2, \dots, \mathcal{T}\xi_n) = (\xi_1, \xi_2, \dots, \xi_n) \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} \\ &\iff (\mathcal{T}\xi_1, \mathcal{T}\xi_2, \dots, \mathcal{T}\xi_n) = (\lambda_1\xi_1, \lambda_2\xi_2, \dots, \lambda_n\xi_n) \\ &\iff \mathcal{T}\xi_i = \lambda_i\xi_i \iff \mathcal{T} \text{ 有 } n \text{ 个线性无关的特征向量 } \xi_1, \xi_2, \dots, \xi_n \end{aligned}$$

(3) 由 (2) 和性质 1.1.6(2) 立即可得。

(4) 由 (1) 可得 \mathcal{T} 可对角化当且仅当存在 X 的一组基 $\xi_1, \xi_2, \dots, \xi_n$ 使得 \mathcal{T} 在这组基下的矩阵 A 可对角化, 设 σ 是 X 中的向量到它在基 $\xi_1, \xi_2, \dots, \xi_n$ 下的坐标的同构映射。由定理 2.27(2) 可知 A 的特征子空间维数之和为 n , 根据性质 1.1.8(7) 可知 \mathcal{T} 属于不同特征值的特征子空间的维数之和为 n 。

(5) 设 \mathcal{T} 全部的不同的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_m$ 。由 (4)、性质 1.3.1(4) 和性质 1.1.6(2) 可得:

$$\begin{aligned} \mathcal{T} \text{ 可对角化} &\iff \sum_{i=1}^m \dim(X_{\lambda_i}) = n \\ &\iff X_{\lambda_i}, i = 1, 2, \dots, m \text{ 的基合起来是 } n \text{ 个线性无关的向量} \\ &\iff X_{\lambda_i}, i = 1, 2, \dots, m \text{ 的基合起来是 } X \text{ 的一组基} \\ &\iff X = X_{\lambda_1} \oplus X_{\lambda_2} \oplus \dots \oplus X_{\lambda_m} \end{aligned}$$

(6) 由 (1) 和定理 2.27(3) 立即可得。 \square

1.3.3 不变子空间

Definition 1.41. 设 X 是域 F 上的线性空间, \mathcal{T} 是 X 上的一个线性变换, E 是 X 的一个子空间。若对于任意的 $\alpha \in E$ 都有 $\mathcal{T}\alpha \in E$, 则称 E 是 X 的一个不变子空间 (*invariant subspace*), 简称为 \mathcal{T} -子空间。

Property 1.3.2. 设 X 是域 F 上的线性空间, \mathcal{T} 是 X 上的一个线性变换, 则:

1. X 、 X 的零子空间是 \mathcal{T} -子空间, 称二者为平凡的 \mathcal{T} -子空间;
2. $\text{Ker}(\mathcal{T}), \text{Im}(\mathcal{T})$ 和 \mathcal{T} 的特征子空间是 \mathcal{T} -子空间;
- 3.

Proof. (1)(2) 显然 \square

1.3.4 投影变换

Property 1.3.3. 设 X 是域 F 上的线性空间, E 和 W 是 X 的子空间, 且有 $X = E \oplus W$, \mathcal{P}_E 是平行于 W 在 E 上的投影变换, \mathcal{P}_W 是平行于 E 在 W 上的投影变换, 则:

1. \mathcal{P}_E 是线性变换;
2. \mathcal{P}_E 是幂等变换;
3. 幂等变换 $\mathcal{T} \in \text{Hom}(X)$ 是平行于 $\text{Ker } \mathcal{T}$ 在 $\text{Im } \mathcal{T}$ 上的投影变换, 此时 $X = \text{Im } \mathcal{T} \oplus \text{Ker } \mathcal{T}$;
4. \mathcal{P}_E 与 \mathcal{P}_W 正交;
5. $\mathcal{P}_E + \mathcal{P}_W = \mathcal{I}$;
6. 若 $\mathcal{T}_1, \mathcal{T}_2 \in \text{Hom}(X)$, \mathcal{T}_1 和 \mathcal{T}_2 是正交的幂等变换, 且 $\mathcal{T}_1 + \mathcal{T}_2 = \mathcal{I}$, 则 $X = \text{Im } \mathcal{T}_1 \oplus \text{Im } \mathcal{T}_2$, 并且有 \mathcal{T}_1 是平行于 $\text{Im } \mathcal{T}_2$ 在 $\text{Im } \mathcal{T}_1$ 上的投影, \mathcal{T}_2 是平行于 $\text{Im } \mathcal{T}_1$ 在 $\text{Im } \mathcal{T}_2$ 上的投影;
7. 平行于 W 在 E 上的投影变换是唯一的。

Proof. (1) 任取 $\alpha, \beta \in X$ 和 $k_1, k_2 \in F$, 其中 $\alpha = \alpha_1 + \alpha_2$, $\beta = \beta_1 + \beta_2$, $\alpha_1, \beta_1 \in E$, $\alpha_2, \beta_2 \in W$, 于是有:

$$\mathcal{P}_E(k_1\alpha + k_2\beta) = \mathcal{P}_E[(k_1\alpha_1 + k_2\beta_1) + (k_1\alpha_2 + k_2\beta_2)] = k_1\alpha_1 + k_2\beta_1 = k_1\mathcal{P}_E\alpha + k_2\mathcal{P}_E\beta$$

所以 \mathcal{P}_E 是线性变换。

(2) 任取 $\alpha = \alpha_1 + \alpha_2 \in X$, $\alpha_1 \in E$, $\alpha_2 \in W$, 则:

$$\mathcal{P}_E(\mathcal{P}_E\alpha) = \mathcal{P}_E(\alpha_1) = \alpha_1 = \mathcal{P}_E\alpha$$

所以 \mathcal{P}_E 是幂等变换。

(3) 任取 $\alpha \in X$, 则 $\mathcal{T}\alpha \in \text{Im } \mathcal{T}$ 。因为:

$$\mathcal{T}(\alpha - \mathcal{T}\alpha) = \mathcal{T}\alpha - \mathcal{T}^2\alpha = \mathcal{T}\alpha - \mathcal{T}\alpha = \mathbf{0}$$

所以 $\alpha - \mathcal{T}\alpha \in \text{Ker } \mathcal{T}$ 。因为 $\alpha = \mathcal{T}\alpha + \alpha - \mathcal{T}\alpha$, 所以 $X = \text{Im } \mathcal{T} + \text{Ker } \mathcal{T}$ 。

任取 $\beta \in \text{Im } \mathcal{T} \cap \text{Ker } \mathcal{T}$, 则存在 $\gamma \in X$ 使得 $\mathcal{T}\gamma = \beta$, 且有 $\mathcal{T}\beta = \mathbf{0}$, 于是:

$$\mathbf{0} = \mathcal{T}\beta = \mathcal{T}(\mathcal{T}\gamma) = \mathcal{T}^2\gamma = \mathcal{T}\gamma = \beta$$

所以 $\text{Im } \mathcal{T} \cap \text{Ker } \mathcal{T} = \mathbf{0}$, 由定理 1.12(3) 可知 $X = \text{Im } \mathcal{T} \oplus \text{Ker } \mathcal{T}$ 。

记 $\text{Im } \mathcal{T} = E$, $\text{Ker } \mathcal{T} = W$, 对上述 $\alpha = \mathcal{T}\alpha + \alpha - \mathcal{T}\alpha$, 有 $\mathcal{P}_E\alpha = \mathcal{T}\alpha$ 。由 α 的任意性, $\mathcal{T} = \mathcal{P}_E$ 。

(4) 任取 $\alpha = \alpha_1 + \alpha_2 \in X$, $\alpha_1 \in E$, $\alpha_2 \in W$, 则:

$$\mathcal{P}_E(\mathcal{P}_W\alpha) = \mathcal{P}_E\alpha_2 = \mathbf{0}, \mathcal{P}_W(\mathcal{P}_E\alpha) = \mathcal{P}_W\alpha_2 = \mathbf{0}$$

所以 $\mathcal{P}_E\mathcal{P}_W = \mathcal{P}_W\mathcal{P}_E = \mathcal{O}$, 即 \mathcal{P}_E 与 \mathcal{P}_W 正交。

(5) 任取 $\alpha = \alpha_1 + \alpha_2 \in X$, $\alpha_1 \in E$, $\alpha_2 \in W$, 则:

$$(\mathcal{P}_E + \mathcal{P}_W)\alpha = \mathcal{P}_E\alpha + \mathcal{P}_W\alpha = \alpha_1 + \alpha_2 = \alpha$$

于是 $\mathcal{P}_E + \mathcal{P}_W = \mathcal{I}$ 。

(6) 任取 $\alpha \in X$, 则 $\alpha = (\mathcal{T}_1 + \mathcal{T}_2)\alpha = \mathcal{T}_1\alpha + \mathcal{T}_2\alpha$, 所以 $X = \text{Im } \mathcal{T}_1 + \text{Im } \mathcal{T}_2$ 。

任取 $\beta \in \text{Im } \mathcal{T}_1 \cap \text{Im } \mathcal{T}_2$, 则存在 $\gamma, \delta \in X$ 使得 $\mathcal{T}_1\gamma = \beta$, $\mathcal{T}_2\delta = \beta$, 于是有 $\beta = \mathcal{T}_1\gamma = \mathcal{T}_1^2\gamma = \mathcal{T}_1(\mathcal{T}_1\gamma) = \mathcal{T}_1\beta$ 。因为 \mathcal{T}_1 与 \mathcal{T}_2 正交, 所以:

$$\mathcal{T}_1\beta = \mathcal{T}_1(\mathcal{T}_2\delta) = (\mathcal{T}_1\mathcal{T}_2)\delta = \mathbf{0}$$

于是 $\beta = \mathbf{0}$, 即 $\text{Im } \mathcal{T}_1 \cap \text{Im } \mathcal{T}_2 = \mathbf{0}$ 。由定理 1.12 可得 $X = \text{Im } \mathcal{T}_1 \oplus \text{Im } \mathcal{T}_2$ 。

任取 $\varepsilon = \mathcal{T}_1\varepsilon_1 + \mathcal{T}_2\varepsilon_2 \in X$, 其中 $\varepsilon_1, \varepsilon_2 \in X$ 。因为 \mathcal{T}_1 与 \mathcal{T}_2 正交、 \mathcal{T}_1 是幂等变换, 所以显然有:

$$\mathcal{T}_1\varepsilon = \mathcal{T}_1(\mathcal{T}_1\varepsilon_1 + \mathcal{T}_2\varepsilon_2) = \mathcal{T}_1^2\varepsilon_1 + (\mathcal{T}_1\mathcal{T}_2)\varepsilon_2 = \mathcal{T}_1\varepsilon_1$$

于是 \mathcal{T}_1 是平行于 $\text{Im } \mathcal{T}_2$ 在 $\text{Im } \mathcal{T}_1$ 上的投影。 \mathcal{T}_2 同理。

(7) 若存在另一平行于 W 在 E 上的投影变换, 则它与 \mathcal{P}_E 的作用完全相同, 于是二者相等, 唯一性得证。□

Corollary 1.3. 设 X 是域 F 上的线性空间, 由性质 1.3.3(3) 可得到如下推论:

1. 若 $X = E \oplus W$, \mathcal{P}_E 为平行于 W 在 E 上的投影变换, 则:

$$E = \text{Im } \mathcal{P}_E, W = \text{Ker } \mathcal{P}_E$$

2. X 的任一子空间 E 是平行于 E 的一个补空间在 E 上的投影变换的象;

3. X 的任一子空间 E 是平行于 E 在 E 的一个补空间上的投影变换的核。

Proof. (1) 任取 $\alpha = \alpha_1 + \alpha_2 \in X$, $\alpha_1 \in E$, $\alpha_2 \in W$, 则 $\mathcal{P}_E \alpha = \alpha_1 \in E$, 所以 $\text{Im } \mathcal{P}_E \subset E$ 。任取 $\beta \in E$, 有 $\mathcal{P}_E \beta = \beta \in \text{Im } \mathcal{P}_E$, 所以 $E \subset \text{Im } \mathcal{P}_E$ 。因此 $E = \text{Im } \mathcal{P}_E$ 。

任取 $\gamma \in W$, 有 $\mathcal{P}_E \gamma = \mathbf{0}$, 所以 $\gamma \in \text{Ker } \mathcal{P}_E$, 于是 $W \subset \text{Ker } \mathcal{P}_E$ 。任取 $\delta = \delta_1 + \delta_2 \in \text{Ker } \mathcal{P}_E$, $\delta_1 \in E$, $\delta_2 \in W$, 则 $\mathcal{P}_E \delta = \delta_1 = \mathbf{0}$, 所以 $\delta = \delta_2 \in W$, 于是 $\text{Ker } \mathcal{P}_E \subset W$ 。因此 $W = \text{Ker } \mathcal{P}_E$ 。

(2) 由定理 1.13 可知 E 必定存在一个补空间 W , 于是 $X = E \oplus W$, 由 (1) 即可得到 $E = \text{Im } \mathcal{P}_E$ 。

(3) 与 (2) 类似可得。 □

1.4 内积空间

内积空间的定义

Definition 1.42. 设 X 为实 (复) 数域 K 上的线性空间。若 X 中任意一对元素 x, y 都对应于 K 中的一个数, 记为 (x, y) , 满足:

1. 线性性: $(\alpha x + \beta y, z) = \alpha(x, z) + \beta(y, z)$, 这里 $z \in X$ 。
2. 对称性: 当 K 为实数域时, $(x, y) = (y, x)$; 当 K 为复数域时, $(x, y) = \overline{(y, x)}$ 。
3. 非负性: $(x, x) \geq 0$, 等号成立当且仅当 $x = \mathbf{0}$ 。

那么就称 X 为实 (复) 内积空间 (*inner product space*), 称 (x, y) 为元素 x, y 的内积 (*inner product*)。

Chapter 2

矩阵

2.1 矩阵空间

Definition 2.1. 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 $A = B$ 。

2.1.1 矩阵的运算

加减法与数量乘法

Definition 2.2. 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 2.3. 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 2.4. 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 2.5. 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。

2.1.2 矩阵的行列式

2.1.3 矩阵的秩

Definition 2.6. 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 2.1. 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 2.2. 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, 由性质 1.1.5(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 2.3. 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 2.1. 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 2.2 可知则 A 的行秩等于 J 的行秩, 由引理 2.1 可知 J 的行秩等于 J 的列秩, 由引理 2.3(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 2.7. 矩阵 A 的行秩和列秩统称为矩阵 A 的秩, 记为 $\text{rank}(A)$ 。

Corollary 2.1. 矩阵的初等变换不改变矩阵的秩。

Proof. 由引理 2.2 和定理 2.1 立即得到。□

¹与引理 2.1 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。

2.2 矩阵的向量空间

Definition 2.8. 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 2.2. 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 2.3. 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 2.3.1(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:

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

□

2.3 线性方程组

Definition 2.9. 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 2.10. 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 2.11. 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 2.4. 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 2.12. 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*)。

2.3.1 初等方法

Theorem 2.5. 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。

2.3.2 秩与子空间

Theorem 2.6. 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) 由性质 1.1.7(4) 可得 $Ax = b$ 有解 $\iff b \in \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle \iff \langle \alpha_1, \alpha_2, \dots, \alpha_n, \beta \rangle = \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle \iff \dim \langle \alpha_1, \alpha_2, \dots, \alpha_n, \beta \rangle = \dim \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle \iff \text{rank}(A) = \text{rank}([A|b])$ 。

(3) 若 $\text{rank}(A) = n$, 则阶梯形矩阵的非零行数 $r = n$, 由定理 2.5 可得此时有唯一解。

(4) 与 (3) 类似。 □

2.3.3 解的结构

齐次线性方程组

Property 2.3.1. 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:

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

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

所以由性质 1.1.7(4) 可得:

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

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

由性质 1.2.1(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 2.13. 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 2.3.2. 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。 □

2.4 矩阵的等价关系

2.4.1 相抵

Definition 2.14. $A, B \in M_{s \times m}(K)$, 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 2.7. 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 2.8. 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 2.9 (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。 □

2.4.2 相似

Definition 2.15. $A, B \in M_n(K)$ 。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 2.10. 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 2.4.1 (相似的不变量). 相似的矩阵具有相同的行列式值、秩、迹、特征多项式、特征值 (包括重数相同)。

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

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

(2) 初等行变换与初等列变换不改变矩阵的秩。

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

(4)(5) 参考定理 2.25。 □

2.4.3 合同

Definition 2.16. $A, B \in M_n(K)$ 。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 2.11. 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 2.17. 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 2.4. $A, B \in M_n(K)$, 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:

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

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

$$C = P_1P_2 \cdots P_t$$

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

□

Theorem 2.12. 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:

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

于是有:

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

因为 A 是一个对称矩阵, 所以 A_2 是一个对称矩阵, 于是:

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

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

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

于是有:

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

因为:

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

并且:

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

是一个可逆矩阵, 所以 A 合同于对角矩阵:

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

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

把 A 的第 $1, i$ 行呼唤, 再把所得矩阵的第 $1, i$ 列呼唤, 得到的矩阵 B 的 $(1, 1)$ 元即为 $a_{ii} \neq 0$ 。根据情形一的讨论, B 合同于一个对角矩阵。因为 B 是由 A 作成对初等行、列变换得到的, 由引理 2.4 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 2.4 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 2.13. 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 2.14. 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 2.12 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 2.4 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 2.15. 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 2.12 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 2.4 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

2.5 相抵的应用

2.5.1 广义逆

Definition 2.18. 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 2.16. 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 2.5.1. 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. $\text{rank}(A^-) \geq \text{rank}(A) = \text{rank}(AA^-) = \text{rank}(A^-A)$;

证明有问题

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:

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

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

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) 由性质 2.8.2(2) 可知 $A(A^T A)^- A^T$ 是向 $\mathcal{M}(A)$ 的正交投影阵, 由性质 1.3.3(7) 和定理 1.20 可得 $A(A^T A)^- A$ 是唯一的, 与 $(A^T A)^-$ 的选择无关。(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:

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

(7) 由 (5) 可得:

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

2.5.2 Moore-Penrose 广义逆

Definition 2.19. 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 2.17. 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 2.3 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:

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

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 2.18. 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。 \square

Moore-Penrose 广义逆的性质

Property 2.5.2. 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 2.18) 可得:

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

将其代入 A^H 的 Penrose 方程组可得:

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

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 2.18) 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 2.6.2(3) 可知存在正交矩阵 Q 使得:

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

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 2.18) 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 2.43(3) 的第三条可知 $I - A^+ A \geq \mathbf{0}$ 。

(8) 由 (3) 可得:

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

由 A 的奇异值分解 (定理 2.47) 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 2.18) 可得:

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

2.5.3 线性方程组的解

Theorem 2.19. 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 2.20. 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 2.21 (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 2.19 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 2.20 可知齐次线性方程组 $Ax = \mathbf{0}$ 的通解为 $(I_n - A^-A)y$, 由性质 2.3.2(3) 可得 $Ax = \beta$ 的通解为 $x = A^-\beta + (I_n - A^-A)y$ 。 \square

Theorem 2.22 (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 2.19 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 2.21 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 2.23. 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 2.21 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 2.5.2(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

2.6 相似的应用

2.6.1 特征值与特征向量

Definition 2.20. $A \in M_n(K)$ 。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 2.21. $A \in M_n(K)$, 称 $|\lambda I - A|$ 为 A 的特征多项式 (characteristic polynomial)。

Theorem 2.24. $A \in M_n(K)$, 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:

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

□

特征向量的性质

Property 2.6.1. $A \in M_n(K)$, 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), 记为 W_λ ;
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。由定理 1.7 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 2.25. 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 2.22. $A \in M_n(K)$, 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 2.26. $A \in M_n(K)$, 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 中取一组基 $\alpha_1, \alpha_2, \dots, \alpha_r$, 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 2.25 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

2.6.2 矩阵的对角化

Definition 2.23. 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 2.27. $A \in M_n(K)$ 可对角化的充分必要条件为

1. A 有 n 个线性无关的特征向量 $\alpha_1, \alpha_2, \dots, \alpha_n$;
2. A 的属于不同特征值的特征子空间的维数之和等于 n ;
3. 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;

4. 设 $\lambda_1, \lambda_2, \dots, \lambda_m$ 是 A 全部的不同的特征值, 则:

$$K^n = W_{\lambda_1} \oplus W_{\lambda_2} \oplus \dots \oplus W_{\lambda_m}$$

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. (1) 显然:

A 与对角矩阵 $D = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 相似, 其中 $\lambda_i \in K, i = 1, 2, \dots, n$

\iff 如果存在可逆矩阵 $P \in M_n(K)$, 使得 $P^{-1}AP = D$

即 $AP = PD$

即 $A(\alpha_1, \alpha_2, \dots, \alpha_n) = (\alpha_1, \alpha_2, \dots, \alpha_n)D$

即 $(A\alpha_1, A\alpha_2, \dots, A\alpha_n) = (\lambda_1\alpha_1, \lambda_2\alpha_2, \dots, \lambda_n\alpha_n)$

$\iff K^n$ 中有 n 个线性无关的列向量 $\alpha_1, \alpha_2, \dots, \alpha_n$ 使得 $A\alpha_i = \lambda_i\alpha_i, i = 1, 2, \dots, n$ \square

(2) 充分性: 由性质 2.6.1(2) 和 (1) 的充分性可直接得出。

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 2.26 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 (1) 的必要性可得 A 不可以对角化。

(3) 充分性: 由 (2) 的充分性可直接得到。

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。

(4) 由 (2)、性质 2.6.1(2)、性质 1.1.6 和定理 1.12(5) 可得:

$$\begin{aligned}
 A \text{ 可对角化} &\iff \sum_{i=1}^m \dim(W_{\lambda_i}) = n \\
 &\iff W_{\lambda_i}, i = 1, 2, \dots, m \text{ 的基合起来是 } n \text{ 个线性无关的向量} \\
 &\iff W_{\lambda_i}, i = 1, 2, \dots, m \text{ 的基合起来是 } K^n \text{ 的一组基} \\
 &\iff K^n = W_{\lambda_1} \oplus W_{\lambda_2} \oplus \dots \oplus W_{\lambda_m}
 \end{aligned}$$

2.6.3 Hermitian 矩阵的对角化

Definition 2.24. 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 2.28. 正交相似是 $M_n(\mathbb{C})$ 上的一个等价关系。

Theorem 2.29. $A \in M_n(\mathbb{C})$ 。若 A 正交相似与一个对角矩阵 D , 则 A 一定是 Hermitian 矩阵。

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

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

所以 A 是一个 Hermitian 矩阵。 □

Corollary 2.2. 正交相似一定相似, 相似不一定正交相似。

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

Property 2.6.2. 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 $\lambda(\alpha, \alpha) = \overline{\lambda}(\alpha, \alpha)$ 。因为 α 是特征向量, 所以 $\alpha \neq \mathbf{0}$, 于是 $\lambda = \overline{\lambda}$, 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:

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

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 的一组基并进行 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 2.25 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 2.30. 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 2.6.2(2) 可知:

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

实对称矩阵特征值的极值性质

Theorem 2.31. 设 $A \in M_n(\mathbb{R})$, A 的特征值从大到小记作 $\lambda_1, \lambda_2, \dots, \lambda_n$, $\varphi_1, \varphi_2, \dots, \varphi_n$ 为对应的标准正交化特征向量, 则:

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

Proof. 由性质 2.6.2(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:

$$\frac{x^T Ax}{x^T x} = \frac{y^T Q^T A Q y}{y^T Q^T Q y} = \frac{y^T \Lambda y}{y^T y} = \frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2} \leq \lambda_1 \frac{\sum_{i=1}^n y_i^2}{\sum_{i=1}^n y_i^2} = \lambda_1$$

$$\frac{x^T Ax}{x^T x} = \frac{y^T Q^T A Q y}{y^T Q^T Q y} = \frac{y^T \Lambda y}{y^T y} = \frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2} \geq \lambda_n \frac{\sum_{i=1}^n y_i^2}{\sum_{i=1}^n y_i^2} = \lambda_n$$

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

2.7 合同的应用——二次型

Definition 2.25. 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 2.26. 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 2.27. 对于数域 K 上的两个 n 元二次型 $x^T A x$ 与 $y^T B y$, 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 2.32. 数域 K 上两个 n 元二次型 $x^T Ax$ 与 $y^T By$ 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 $C^T AC = B$ 。作非退化线性变换 $x = Cy$, 可得到 $(Cy)^T A(Cy) = y^T C^T ACy = y^T By$, 所以 $x^T Ax \cong y^T By$ 。

(2) 必要性: 因为 $x^T Ax \cong y^T By$, 所以存在非退化线性变换 $x = Cy$, C 是一个可逆矩阵, 把 $x^T Ax$ 变为 $y^T By$, 即 $(Cy)^T A(Cy) = y^T C^T ACy = y^T By$, 所以 $C^T AC = B$, 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 2.33. 数域 K 上任一 n 元二次型都等价于一个只含平方项的二次型。

Proof. 当二次型的矩阵是对角矩阵时该二次型只含平方项, 由定理 2.12 与定理 2.32 可立即得出结论。 \square

Theorem 2.34. 设 n 元二次型 $x^T Ax$ 的矩阵 A 合同于对角矩阵 $D = \text{diag}\{d_1, d_2, \dots, d_n\}$, 即存在可逆矩阵 C 使得 $C^T AC = D$ 。令 $x = Cy$, 则可以得到 $x^T Ax$ 的一个标准形:

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

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

Proof. 设 n 元二次型 $x^T Ax$ 经过非退化线性变换 $x = Cy$ 化成标准形 $d_1 y_1^2 + d_2 y_2^2 + \dots + d_r y_r^2$, 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 2.13 可得 $\text{rank}(A) = r$ 。 \square

Definition 2.28. 称二次型 $x^T Ax$ 的矩阵 A 的秩为二次型 $x^T Ax$ 的秩。

2.7.1 二次型的规范形

实二次型的规范形

Definition 2.29. 实数域上的二次型称为实二次型。由定理 2.14 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 2.32 可知经过一个适当的非退化线性变换可以将 $x^T Ax$ 化作:

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

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

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

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

要证规范形唯一, 即证 $p = q$ 。

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

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

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

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

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

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

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

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

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

Definition 2.30. 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 2.37. 两个 n 元实二次型等价

\iff 它们的规范形相同

\iff 它们的秩相等, 并且正惯性指数也相等。

Proof. 第一条由定理 2.36 以及二次型等价的传递性、对称性可直接得到（必要性的证明中需要考虑规范形的定义，然后使用定理 2.36），第二条是显然的。□

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

Proof. 由定理 2.32 可得矩阵合同等价于各自对应的二次型等价，再由定理 2.37 可得两个二次型的秩与正惯性指数都相等。因为矩阵的秩与正惯性指数等于对应的二次型的秩与正惯性指数，所以结论成立。□

复二次型的规范形

Definition 2.31. 复数域上的二次型称为复二次型。由定理 2.15 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 2.32 可知经过一个适当的非退化线性变换可以将 $x^T Ax$ 化作：

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

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

Theorem 2.39. 复二次型 $x^T Ax$ 的规范形是唯一的。

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

Theorem 2.40. 两个 n 元复二次型等价

\iff 它们的规范形相同

\iff 它们的秩相等。

Proof. 第一条由定理 2.39 以及二次型的传递性、对称性可直接得到（必要性的证明中需要考虑规范形的定义，然后使用定理 2.39），第二条是显然的。□

2.7.2 正定二次型与正定矩阵

Definition 2.32. 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 2.33. 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 2.41. n 元实二次型 $x^T Ax$ 是正定的当且仅当它的正惯性指数等于 n 。

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

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

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:

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

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。□

Theorem 2.42. 由上述定理可得到如下推论:

1. 对于 n 元实二次型 $x^T Ax$, 下述说法等价:

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

2. 与正定二次型等价的实二次型也是正定的;

3. 对于 n 阶实对称矩阵 A , 下述说法等价:

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

4. 与正定矩阵合同的实对称矩阵也是正定矩阵。

5. 正定矩阵的行列式大于 0;

Proof. (1) $1 \iff 2$: 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 2.32 和正定矩阵的定义可直接得到。

(3) $1 \Rightarrow 2$: 因为 A 是正定的, 所以 n 元二次型 $x^T A x$ 是正定的, 由上一定理可得 $x^T A x$ 的正惯性指数为 n 。因为 A 的正惯性指数等于 $x^T A x$ 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 2.6.2(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 2.32、上一定理和矩阵正定的定义可直接得到。

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

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

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

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

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

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

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

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

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

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

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

注意到:

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

可逆矩阵行列式链接

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:

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

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 2.34. 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 2.35. 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 2.43. 由上述定理可得到如下推论:

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

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

2. 与半正定二次型等价的实二次型也是半正定的;

3. 对于 n 阶实对称矩阵 A , $\text{rank}(A) = r$, 下述说法等价:

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

4. 与半正定矩阵合同的实对称矩阵也是半正定矩阵。

5. 半正定矩阵的行列式为 0;

Proof. (1) $1 \Rightarrow 3$: 作非退化线性变换 $x = Cy$ 把 $x^T A x$ 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 $x^T A x$ 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 2.32 和半正定矩阵的定义可直接得到。

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

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

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

$4 \Rightarrow 5$: 由性质 2.6.2(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 2.32、(1) 的第二条和矩阵半正定的定义可直接得到。

$1 \Rightarrow 6$:

$6 \Rightarrow 5$:

有空证明

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

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

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

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

负定矩阵

Definition 2.36. 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 $x^T Ax$ 是负定 (negative definite) 的。

Definition 2.37. 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 2.44. 对称矩阵 $A \in M_n(\mathbb{R})$ 负定的充分必要条件为: 它的奇数阶顺序主子式都小于 0, 偶数阶顺序主子式都大于 0。

Proof. 设 $|A_k|$ 为 A 的 k 阶顺序主子式, 由定理 2.42(3) 的第六条:

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

2.8 特殊矩阵

2.8.1 幂等阵

Definition 2.38. 若矩阵 $A \in M_n(K)$ 满足 $A^2 = A$, 则称 A 为幂等矩阵 (*idempotent matrix*)。

Theorem 2.45. $A^+ A, A A^+, I - A^+ A, I - A A^+$ 都是幂等阵。

Proof. 代入广义逆方程即可得出结论。 \square

Property 2.8.1. 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 $\iff \text{rank}(A) + \text{rank}(I_n - A) = n$;
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:

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

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??(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??(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 2.3.1(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 2.6.2(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 2.5.2(5) 和 (1) 可得:

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

令 $X = A^+$, 将其代入关于 A^+ 的 Penrose 方程组可知相容。由性质 2.5.2(2)(1) 即可得 $A = A^+$ 。□

2.8.2 正交投影阵

Definition 2.39. 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 2.8.2. 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 $\iff P_1 P_2 = P_2 P_1 = \mathbf{0}$;
7. 当 $P_1 P_2 = P_2 P_1 = \mathbf{0}$ 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. $P = P_1 P_2$ 为正交投影阵 $\iff P_1 P_2 = P_2 P_1$;
9. 当 $P_1 P_2 = P_2 P_1$ 时, $P = P_1 P_2$ 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 $\iff P_1 P_2 = P_2 P_1 = P_2$;
11. 当 $P_1 P_2 = P_2 P_1 = P_2$ 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 $P_A x \in \mathcal{M}(A)$, 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 $P_A x = y$ (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 $\mathcal{M}(A) = \mathcal{M}(P_A)$ 。由性质 1.1.7(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, 根据定理 1.13 和性质 1.1.7(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:

存在性证明

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

$$P_A x = P_A A\alpha + P_A B\beta = A\alpha, \quad \forall \alpha \in K^{\text{rank}(A)}, \beta \in K^{m-\text{rank}(A)}$$

所以:

$$\begin{cases} P_A A = A \\ P_A B = 0 \end{cases}$$

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 $A^T A C = A^T$, 因为该方程是相容的, 由定理 2.22 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 2.5.1(6) 可得 $P_A = A(A^T A)^- A^T$ 。

(3) 充分性: 由 (2) 和性质 2.8.1(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 2.5.1(6) 得出。由性质 2.5.1(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。

□

K^n 子空间
为闭集

2.9 矩阵的分解

2.9.1 SVD 分解

Theorem 2.46. 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 2.43(3) 的第五条可知 AA^T 是半正定矩阵。□

Theorem 2.47. 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:

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

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 2.3 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 2.6.2 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 $B = A Q$, 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 $B = A Q$, 所以:

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

Definition 2.40. 设 $A \in M_{m \times n}(\mathbb{C})$, $\text{rank}(A) = r$, $A^H A$ 的正特征值为 λ_i , $i = 1, 2, \dots, r$, 称 $\delta_i = \sqrt{\lambda_i}$ 为矩阵 A 的奇异值 (*singular value*)。

Chapter 3

随机变量的数字特征

3.1 期望

Definition 3.1. 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*)。

3.2 方差

Property 3.2.1. 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:

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

□

3.3 矩

3.3.1 原点矩

Definition 3.2. 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 3.3. 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 3.1. 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。

□

3.3.2 中心矩

Definition 3.4. 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 3.5. 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 3.2. 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:

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

3.4 协方差

Definition 3.6. 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 3.7. 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 3.4.1. 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;

期望的线性
性

(5) 由可得:

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

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

(6) 显然:

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

3.5 二次型

Definition 3.8. \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 3.3. \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??(3)。 \square

独立随机变量二次型的方差

Theorem 3.4. 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 3.2.1(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 3.3 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 3.3 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 3.3 和 ??(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:

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

3.6 矩母函数

Definition 3.9. 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 3.10. 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 3.6.1. 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) $M_{\mathbf{X}}(\mathbf{0}) = E(e^0) = 1$ 。

(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:

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

Jensen 不等式链接

(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:

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

期望的线性性质,
Lebesgue
积分

(6) 由可得:

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

□

3.7 累积量生成函数

Definition 3.11. 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 3.12. 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 3.13. 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 3.7.1.

3.8 特征函数

Definition 3.14. 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 3.15. 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 3.16. 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 3.8.1. 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:

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

链 接
Lebesgue
积分性质

所以 $|e^{itX}| = 1$, 于是:

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

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:

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

由可得 $X_k, k = 1, 2, \dots, n$ 相互独立。

链接独立性
条件

(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 3.1 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:

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

需要证明对
小于的都存
在

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

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

存在, 于是 $E(X^n)$ 存在。由定理 3.1 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:

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

□

3.9 Fisher 信息量

Definition 3.17. 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 3.1. 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 3.9.1. 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 4

正态分布与三大抽样分布

4.1 多元正态分布

4.1.1 多元正态分布的定义

Definition 4.1. 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 4.1. 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 3.4.1 可得:

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

Definition 4.2. 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??:

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

多元正态分布的等价定义

Definition 4.3. \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 4.2. \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 4.1 的充分必要条件是满足定义 4.3, 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 4.3, 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:

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

期望的性质 因为 $\mathbf{X} = A\mathbf{U} + \boldsymbol{\mu}$, 由和性质 3.4.1(3)(4)(5) 可得:

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

写完矩阵的秩做链接

因为 $\text{Cov}(\mathbf{X}) > 0$, 由定理 2.42 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

写完变量变换法链接过来

变量函数的分布中的变量变换法可知:

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

即 \mathbf{X} 满足定义 4.4。

(2) 必要性：设 \mathbf{X} 满足定义 4.1，此时只要选择 $A = \Sigma^{\frac{1}{2}}$ 即可得到 \mathbf{X} 满足定义 4.3。□

4.1.2 多元正态分布的性质

Theorem 4.3. 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 4.3 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 4.3 可得 $\mathbf{Y} \sim N(B\boldsymbol{\mu} + c, B\Sigma B^T)$ 。□

Corollary 4.1. 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??(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:

$$Q\sigma^2IQ^T = \sigma^2QQ^T = \sigma^2I_n$$

(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 4.4. 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 4.3 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 3.8.1(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 3.8.1(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 4.5. 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 4.1(3) 直接得到。

(2) 充分性：由定理 4.4 可知此时 $\alpha^T \mathbf{X}$ 的特征函数为：

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

其中 μ 和 σ^2 分别为 $\alpha^T \mathbf{X}$ 的均值与方差。由和性质 3.4.1(3) 可得：

期望的性质

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

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 3.8.1(6) 和定理 4.4 可知 \mathbf{X} 服从多元正态分布。 □

Lemma 4.1. 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 4.1 即可得到。 □

Theorem 4.6. 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 4.1 和定理 4.3 即可得到结论。 □

Theorem 4.7. 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 4.1(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 4.4 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 4.4 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 3.8.1(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 4.8. 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 3.3 得到。记 $\mathbf{Y} = \Sigma^{-\frac{1}{2}} \mathbf{X}$, 由推论 4.1(1) 可知 $\mathbf{Y} \sim N_n(\Sigma^{-\frac{1}{2}} \mu, I_n)$, 根据定理 4.7, \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 3.4、??(3) 和 ??(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 4.9. 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 4.1(1) 可得:

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

于是根据??(1) 和 (3) 可得:

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

Theorem 4.10. 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 2.8.1(1) 可知 A 的特征值只能为 0 或 1。根据性质 2.6.2(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 4.1(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 4.1(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 2.6.2(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 4.1(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 4.7 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 4.1(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 3.8.1(4) 和性质 4.2.1(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 3.8.1(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 4.2. 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 4.3 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??(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 4.10 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 4.11. 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 2.6.2(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:

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

记 $\mathbf{Y} = Q^T \mathbf{X}$, 由推论 4.1(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 4.7 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 4.3. 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 4.12. 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

4.1.3 矩阵正态分布的定义

密度函数定义

Definition 4.4. 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 4.5. 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 4.13. 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 4.4 的充分必要条件为满足定义 4.5。

Proof. 由??(3)、??(1)(2)、??(2)(3) 可得:

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

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 4.4. 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 $M = \mathbf{0}$, $V \otimes U = I_{mn}$ 。

需要补充证明, 但这里涉及到了 Jordan 标准形, 学完再来补。

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:

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

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:

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

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 4.6. \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 4.14. \mathbf{X} 是一个 $m \times n$ 随机矩阵, 则 \mathbf{X} 满足定义 4.5 的充分必要条件是满足定义 4.6。

Proof. (1) 必要性: 设 \mathbf{X} 满足定义 4.6, 由 \mathbf{Y} 的定义可知:

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

由 ??(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 3.4.1(3) 和 ??(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 4.3 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 4.5。

(2) 充分性: 设 \mathbf{X} 满足定义 4.5, 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 ??(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 4.6。 □

4.1.4 矩阵正态分布的性质

Theorem 4.15. 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 4.6 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 4.6 即可得到结论。 \square

4.2 χ^2 分布, t 分布和 F 分布

4.2.1 χ^2 分布

Definition 4.7. 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 4.2.1. 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 3.2.1(1) 可知:

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

因为 X_i 相互独立, 由可知:

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

链接独立方差等于和

由性质 3.2.1(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 4.7 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 3.8.1(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 4.1(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:

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

□

4.2.2 t 分布

Definition 4.8. 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 $T \sim t_n$ 。

4.2.3 F 分布

Definition 4.9. 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 4.2.2. 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 5

something

5.1 抽样分布

Theorem 5.1. 设 X_1, X_2, \dots, X_m i.i.d. $\sim N(\mu, \sigma^2)$, Y_1, Y_2, \dots, Y_n i.i.d. $\sim N(\nu, \sigma^2)$, X_1, X_2, \dots, X_m 和 Y_1, Y_2, \dots, Y_n 相互独立, \bar{X}, \bar{Y} 为样本均值, S_X^2, S_Y^2 为样本方差, 则:

1. $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{m}\right)$;
2. $\frac{(m-1)S_X^2}{\sigma^2} \sim \chi_{m-1}^2$;
3. \bar{X} 与 S_X^2 独立;
4. $\frac{\sqrt{m}(\bar{X} - \mu)}{S_X} \sim t_{m-1}$;
5. $\frac{\bar{X} - \bar{Y} - (\mu - \nu)}{S_w} \sqrt{\frac{mn}{m+n}} \sim t_{m+n-2}$, 其中:

$$(m+n-2)S_w^2 = (m-1)S_X^2 + (n-1)S_Y^2$$

6. 若 X_1, X_2, \dots, X_m i.i.d. $\sim N(\mu, \sigma_1^2)$, Y_1, Y_2, \dots, Y_n i.i.d. $\sim N(\nu, \sigma_2^2)$, 其它条件不变, 则:

$$\frac{S_X^2 \sigma_2^2}{S_Y^2 \sigma_1^2} \sim F_{m-1, n-1}$$

Proof. 令 $\mathbf{X} = (X_1, X_2, \dots, X_m)^T$, $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$. 因为 X_1, X_2, \dots, X_m i.i.d. $\sim N(\mu, \sigma^2)$, Y_1, Y_2, \dots, Y_n i.i.d. $\sim N(\nu, \sigma^2)$, 所以 $\mathbf{X} \sim N_m(\boldsymbol{\mu}, \Sigma_m)$, $\mathbf{Y} \sim N_n(\boldsymbol{\nu}, \Sigma_n)$, 其中:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix}, \Sigma_m = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}, \boldsymbol{\nu} = \begin{pmatrix} \nu \\ \nu \\ \vdots \\ \nu \end{pmatrix}, \Sigma_n = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$$

(1) 令 m 维行向量 $c = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$, 由定理 4.3 可知:

$$\bar{X} = c\mathbf{X} \sim N(c\boldsymbol{\mu}, c\Sigma c^T)$$

而:

$$c\boldsymbol{\mu} = \sum_{i=1}^m \frac{\mu}{m} = \mu, \quad c\Sigma c^T = \sum_{i=1}^m \frac{\sigma^2}{m^2} = \frac{\sigma^2}{m}$$

所以 $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{m}\right)$ 。

考虑链接什
么过来

(2) 由 Schmidt 正交化 可知存在正交矩阵:

$$A = \begin{pmatrix} \frac{1}{\sqrt{m}} & \frac{1}{\sqrt{m}} & \cdots & \frac{1}{\sqrt{m}} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix}$$

令 $\mathbf{Z} = A\mathbf{X}$, 由推论 4.1(2) 可知 $\mathbf{Z} \sim N_m(A\boldsymbol{\mu}, \sigma^2 I_m)$ 。由推论 4.1(4) 可知 $\mathbf{Z}_i \sim N(\mu_i, \sigma^2)$, 其中:

$$\mu_i = \mu \sum_{j=1}^m a_{ij}$$

因为 A 是一个正交矩阵, 所以:

$$\mu_i = \sqrt{m}\mu \sum_{j=1}^m \frac{1}{\sqrt{m}} a_{ij} = \sqrt{m}\mu \left(\frac{1}{\sqrt{m}}, \frac{1}{\sqrt{m}}, \dots, \frac{1}{\sqrt{m}}\right) (a_{i1}, a_{i2}, \dots, a_{im})^T = 0$$

由定理 4.3 即可得:

$$\frac{\mathbf{Z}_i}{\sigma} \sim N(0, 1), \quad \forall i = 1, 2, \dots, m$$

因为 $\mathbf{Z} = A\mathbf{X}$, 所以:

$$\mathbf{Z}_1 = \frac{1}{\sqrt{m}} \sum_{i=1}^m X_i = \sqrt{m}\bar{X}$$

因为:

$$\mathbf{Z}^T \mathbf{Z} = \sum_{i=1}^m \mathbf{Z}_i^2 = \mathbf{X}^T A^T A \mathbf{X} = \mathbf{X}^T \mathbf{X} = \sum_{i=1}^m X_i^2$$

于是:

$$(m-1)S_X^2 = \sum_{i=1}^m (X_i - \bar{X})^2 = \sum_{i=1}^m X_i^2 - m\bar{X}^2 = \sum_{i=1}^m \mathbf{Z}_i^2 - \mathbf{Z}_1^2 = \sum_{i=2}^m \mathbf{Z}_i^2$$

所以:

$$\frac{(m-1)S_X^2}{\sigma^2} = \sum_{i=2}^m \left(\frac{\mathbf{Z}_i}{\sigma}\right)^2 \sim \chi_{m-1}^2$$

(3) 由 (2) 的证明过程可得 $\mathbf{Z} \sim N_m(A\boldsymbol{\mu}, \sigma^2 I_m)$, 根据定理 4.7 可知 $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_m$ 相互独立。而:

$$S_X^2 = \frac{\sum_{i=2}^m \mathbf{Z}_i^2}{(m-1)}, \quad \bar{X} = \frac{\mathbf{Z}_1}{\sqrt{m}}$$

所以 S_X^2 与 \bar{X} 独立。

(4) 对 \bar{X} 进行标准化可得:

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{m}}} = \frac{\sqrt{m}(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$$

由 (2) 和 (3) 进一步可得:

$$\frac{\frac{\sqrt{m}(\bar{X} - \mu)}{\sigma}}{\sqrt{\frac{(m-1)S_X^2}{\sigma^2(m-1)}}} = \frac{\sqrt{m}(\bar{X} - \mu)}{S_X} \sim t_{m-1}$$

(5) 由 (1) (得到 \bar{X} 和 \bar{Y} 的分布)、 X_1, X_2, \dots, X_m 与 Y_1, Y_2, \dots, Y_n 相互独立 (得到二维随机向量 (\bar{X}, \bar{Y}) 的分布) 和定理 4.3 (对 (\bar{X}, \bar{Y}) 用二维行向量 $(1, -1)$ 做线性变换) 可得:

$$\bar{X} - \bar{Y} \sim N\left(\mu - \nu, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right)$$

于是:

$$\frac{\bar{X} - \bar{Y} - (\mu - \nu)}{\sqrt{\frac{m+n}{mn}\sigma^2}} \sim N(0, 1)$$

由 (2) 可得:

$$\frac{(m-1)S_X^2}{\sigma^2} \sim \chi_{m-1}^2, \quad \frac{(n-1)S_Y^2}{\sigma^2} \sim \chi_{n-1}^2$$

由性质 4.2.1(1) 可得:

$$\frac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2} \sim \chi_{m+n-2}^2$$

于是:

$$\frac{(m+n-2)S_w^2}{\sigma^2} \sim \chi_{m+n-2}^2$$

由 (3) 可得 \bar{X} 与 S_X^2 独立、 \bar{Y} 与 S_Y^2 独立, 所以:

$$\frac{\frac{\bar{X} - \bar{Y} - (\mu - \nu)}{\sqrt{\frac{m+n}{mn}\sigma^2}}}{\sqrt{\frac{(m+n-2)S_w^2}{\sigma^2(m+n-2)}}} = \frac{\bar{X} - \bar{Y} - (\mu - \nu)}{S_w} \sqrt{\frac{mn}{m+n}} \sim t_{m+n-2}$$

(6) 由 (2) 可知:

$$\frac{(m-1)S_X^2}{\sigma_1^2} \sim \chi_{m-1}^2, \quad \frac{(n-1)S_Y^2}{\sigma_2^2} \sim \chi_{n-1}^2$$

因为 X_1, X_2, \dots, X_m 和 Y_1, Y_2, \dots, Y_n 相互独立, 所以上两式也相互独立。由 F 分布的定义即可得:

$$\frac{\frac{(m-1)S_X^2}{\sigma_1^2(m-1)}}{\frac{(n-1)S_Y^2}{\sigma_2^2(n-1)}} = \frac{S_X^2 \sigma_2^2}{S_Y^2 \sigma_1^2} \sim F_{m-1, n-1}$$

□

5.2 次序统计量

Definition 5.1. 设 X_1, X_2, \dots, X_n 为从总体中抽取的样本, 将其按大小排列为 $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, 称 $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ 为样本 X_1, X_2, \dots, X_n 的次序统计量 (order statistics)。

次序统计量的联合分布

Theorem 5.2. 设总体的密度函数为 $f(x)$, $x \in \mathbb{R}$, X_1, X_2, \dots, X_n 为从总体中抽取的简单样本。令 $Y_i = X_{(i)}$, $i = 1, 2, \dots, n$, 则次序统计量 (Y_1, Y_2, \dots, Y_n) 的联合密度为:

$$g(y_1, y_2, \dots, y_n) = \begin{cases} n! f(y_1) f(y_2) \cdots f(y_n), & y_1 < y_2 < \cdots < y_n \\ 0, & \text{其它} \end{cases}$$

Proof. 在 \mathbb{R}^n 中划分 $n!$ 个区域, 每个区域分别对应着一个 i_1, i_2, \dots, i_n 使得 $x_{i_1} < x_{i_2} < \cdots < x_{i_n}$, 因为 x_1, x_2, \dots, x_n 的排列一共有 $n!$ 种, 所以这 $n!$ 个区域加上包括等于号的一些零测集就构成了整个 \mathbb{R}^n 。因为次序统计量的密度函数也是在 \mathbb{R}^n 上的一个概率测度, 则可以对每个划分的区域求 (Y_1, Y_2, \dots, Y_n) 的概率测度, 再对所有区域求和, 即可得到次序统计量的联合密度。这个过程类似于全概率公式。

任取一个上述区域 A 作变换:

$$y_j = x_{i_j}, \quad j = 1, 2, \dots, n, \quad x_{i_1} < x_{i_2} < \cdots < x_{i_n}$$

则该变换的 Jacobi 行列式为 $|\mathbf{J}| = |I_n| = 1$, 因为 X_1, X_2, \dots, X_n 是简单样本, 所以在该区域上的:

$$g(y_1, y_2, \dots, y_n | A) = \begin{cases} \prod_{i=1}^n f(x_i) = \prod_{i=1}^n f(y_i), & y_1 < y_2 < \cdots < y_n \\ 0, & \text{其它} \end{cases}$$

由区域的任意性可得在整个 \mathbb{R}^n 上:

$$g(y_1, y_2, \dots, y_n) = \begin{cases} n! f(y_1) f(y_2) \cdots f(y_n), & y_1 < y_2 < \cdots < y_n \\ 0, & \text{其它} \end{cases} \quad \square$$

部分次序统计量的联合分布

Lemma 5.1. 设总体的分布函数为 $F(x)$, 密度函数为 $f(x)$, X_1, X_2, \dots, X_n 为从总体中抽取的简单样本, 则有:

$$\int_a^b \cdots \int_b^b f(x_1) \cdots f(x_n) dx_1 \cdots dx_n = \frac{1}{n!} [F(b) - F(a)]^n$$

其中 $a, b \in \overline{\mathbb{R}}$ 。

Proof. 因为 X_1, X_2, \dots, X_n 独立同分布, 所以:

$$\int_a^b \cdots \int_a^b f(x_1) \cdots f(x_n) dx_1 \cdots dx_n = \left[\int_a^b f(x_1) dx_1 \right]^n = [F(b) - F(a)]^n$$

在这个区域上对 x_1, x_2, \dots, x_n 的排序结果一共有 $n!$ 种 (不考虑等于的情况, 测度为 0, 不影响积分结果), 每种排序都是等可能的, 于是结论成立。□

Theorem 5.3. 设 X_1, X_2, \dots, X_n 为从总体中抽取的简单样本, 总体的分布函数和密度函数分别为 $F(x), f(x)$, 则样本次序统计量中任意 m 个分量 $Y_{(i_1)}, Y_{(i_2)}, \dots, Y_{(i_m)}$, $i_1 < i_2 < \cdots < i_m$ 的联合密度为:

$$g(y_{i_1}, y_{i_2}, \dots, y_{i_m}) = n! \prod_{j=1}^m f(y_{i_j}) \frac{1}{(i_1 - 1)!} F^{i_1 - 1}(y_{i_1}) \frac{1}{(n - i_m)!} [1 - F(y_{i_m})]^{n - i_m} \\ \left\{ \prod_{k=2}^m \frac{1}{(i_k - i_{k-1} - 1)!} [F(y_{i_k}) - F(y_{i_{k-1}})]^{i_k - i_{k-1} - 1} \right\}$$

Proof. 注意到 $Y_{(i_1)}, Y_{(i_2)}, \dots, Y_{(i_m)}$ 的联合密度是次序统计量的边缘密度, 所以由引理 5.1 可得:

$$g(y_{i_1}, y_{i_2}, \dots, y_{i_m}) = \int_{-\infty < y_1 < \cdots < y_n < +\infty} n! f(y_1) f(y_2) \cdots f(y_n) dy_1 \cdots dy_{i_1-1} \\ dy_{i_1+1} \cdots dy_{i_2-1} dy_{i_2+1} \cdots dy_{i_m-1} dy_{i_m+1} \cdots dy_n \\ = n! \prod_{j=1}^m f(y_{i_j}) \int_{-\infty < y_1 < \cdots < y_{i_1}} f(y_1) \cdots f(y_{i_1-1}) dy_1 \cdots dy_{i_1-1} \\ \times \int_{y_{i_1} < y_{i_1+1} < y_{i_1+2} < \cdots < y_{i_2}} f(y_{i_1+1}) \cdots f(y_{i_2-1}) dy_{i_1+1} \cdots dy_{i_2-1} \\ \cdots \cdots \times \int_{y_{i_m} < y_{i_m+1} < y_{i_m+2} < \cdots < +\infty} f(y_{i_m+1}) \cdots f(y_n) dy_{i_m+1} \cdots dy_n \\ = n! \prod_{j=1}^m f(y_{i_j}) \frac{1}{(i_1 - 1)!} F^{i_1 - 1}(y_{i_1}) \frac{1}{(i_2 - i_1 - 1)!} [F(y_{i_2}) - F(y_{i_1})]^{i_2 - i_1 - 1} \\ \cdots \frac{1}{(n - i_m)!} [1 - F(y_{i_m})]^{n - i_m} \\ = n! \prod_{j=1}^m f(y_{i_j}) \frac{1}{(i_1 - 1)!} F^{i_1 - 1}(y_{i_1}) \frac{1}{(n - i_m)!} [1 - F(y_{i_m})]^{n - i_m} \\ \left\{ \prod_{k=2}^m \frac{1}{(i_k - i_{k-1} - 1)!} [F(y_{i_k}) - F(y_{i_{k-1}})]^{i_k - i_{k-1} - 1} \right\} \quad \square$$

极差的分布

Theorem 5.4. 设 X_1, X_2, \dots, X_n 为从总体中抽取的简单样本, 总体的分布函数和密度函数分别为 $F(x), f(x)$, 样本的次序统计量为 (Y_1, Y_2, \dots, Y_n) , 则对于任意的 $i, j = 1, 2, \dots, n$ 满足 $i < j$, 令 $V = Y_j - Y_i$, 则有:

$$g(v) = \int_{-\infty}^{+\infty} g(u, v) du$$

其中:

$$g(u, v) = \begin{cases} \frac{n!f(u)f(u+v)}{(i-1)!(j-i-1)!(n-j)!} F^{i-1}(u)[F(u+v) - F(u)]^{j-i-1} \\ \quad [1 - F(u+v)]^{n-j}, & v > 0 \\ 0, & \text{其它} \end{cases}$$

链接随机变量函数的分布中的增补变量法

Proof. 使用增补变量法, 做变换:

$$\begin{cases} U = Y_i \\ V = Y_j - Y_i \end{cases} \Leftrightarrow \begin{cases} Y_i = U \\ Y_j = V + U \end{cases}$$

该变换的 Jacobi 行列式为:

$$|\mathbf{J}| = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$$

由定理 5.3 可得 (Y_i, Y_j) 的联合密度:

$$g(y_i, y_j) = \begin{cases} \frac{n!f(y_i)f(y_j)}{(i-1)!(j-i-1)!(n-j)!} F^{i-1}(y_i)[F(y_j) - F(y_i)]^{j-i-1} \\ \quad [1 - F(y_j)]^{n-j}, & y_i < y_j \\ 0, & \text{其它} \end{cases}$$

于是:

$$g(u, v) = \begin{cases} \frac{n!f(u)f(u+v)}{(i-1)!(j-i-1)!(n-j)!} F^{i-1}(u)[F(u+v) - F(u)]^{j-i-1} \\ \quad [1 - F(u+v)]^{n-j}, & v > 0 \\ 0, & \text{其它} \end{cases}$$

所以:

$$g(v) = \int_{-\infty}^{+\infty} g(u, v) du$$

□

Corollary 5.1. 次序统计量 (Y_1, Y_2, \dots, Y_n) 极差的分布为:

$$g(v) = \int_{-\infty}^{+\infty} g(u, v) du$$

其中:

$$g(u, v) = \begin{cases} n(n-1)f(u)f(u+v)[F(u+v) - F(u)]^{n-2}, & v > 0 \\ 0, & \text{其它} \end{cases}$$

5.3 充分统计量

Definition 5.2. 设 X 的分布族为 $\{f(x, \theta) : \theta \in \Theta\}$, Θ 是参数空间。令 $T = T(\mathbf{X})$ 为一个统计量。若 $P(\mathbf{X}|T(\mathbf{X}))$ 与 θ 无关, 则称 $T(\mathbf{X})$ 为 θ 的充分统计量 (*sufficient statistics*)。

5.4 Delta method

Delta method 可以给出随机变量函数的近似方差。

Theorem 5.5. 设随机向量 \mathbf{X} 的均值为 $E(\mathbf{X})$, 方差为 $Var(\mathbf{X})$, 现有另一随机变量 $g(\mathbf{X})$, 则该随机变量有如下近似方差:

$$Var[g(\mathbf{X})] \approx \nabla g[E(\mathbf{X})]^\top Cov(\mathbf{X}) \nabla g[E(\mathbf{X})]$$

Proof. 将 $g(\mathbf{X})$ 在 $g[E(\mathbf{X})]$ 处进行泰勒展开:

$$g(\mathbf{X}) \approx g[E(\mathbf{X})] + \nabla g[E(\mathbf{X})]^\top [\mathbf{X} - E(\mathbf{X})]$$

对此式求方差:

$$Var[g(\mathbf{X})] \approx Var[\nabla g[E(\mathbf{X})]^\top \mathbf{X}] = \nabla g[E(\mathbf{X})]^\top Cov(\mathbf{X}) \nabla g[E(\mathbf{X})] \quad \square$$

Corollary 5.2. 设随机变量 X 的均值为 $E(X)$, 方差为 $Var(X)$, 现有另一随机变量 $g(X)$, 则该随机变量有如下近似方差:

$$Var[g(X)] \approx g'[E(X)]^2 Var(X)$$

Theorem 5.6. 设 x_1, x_2, \dots, x_n 独立同分布于 $N(\mu, \sigma^2)$, 其中 μ 未知 σ^2 已知, 若 $\mu \sim N(\theta, \tau^2)$, 求 μ 的后验分布。

Proof. 由 □

5.5 主成分分析

主成分分析 (*principal component analysis, PCA*) 的目的是: 对数据进行一个线性变换, 在最大程度保留原始信息的前提下去除数据中彼此相关的信息。反映在变量上就是说, 对所有的变量进行一个线性变换, 使得变换后得到的变量彼此之间不相关, 并且是所有可能的线性变换中方差最大的一些变量 (我们认为方差体现了信息量的大小)。

5.5.1 总体主成分分析

Definition 5.3. 设 \mathbf{X} 是一个 n 维随机向量, 其均值向量为 $\mu = (\mu_1, \mu_2, \dots, \mu_n)$ 、协方差矩阵为 $\Sigma = (\sigma_{ij})$, $i, j = 1, 2, \dots, n$ 。对 \mathbf{X} 进行一个线性变换 \mathcal{T} 得到一个 n 维随机向量

$\mathbf{Y} = \mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$, \mathcal{T} 的矩阵为:

$$A = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$$

若:

1. $\text{Cov}(\mathbf{Y})$ 是一个对角矩阵, 即 $\text{Cov}(\mathbf{Y}_i, \mathbf{Y}_j) = 0, i \neq j$;
2. \mathbf{Y}_1 是所有对 \mathbf{X} 进行线性变换后得到的随机变量中方差最大的随机变量, \mathbf{Y}_2 是与 \mathbf{Y}_1 不相关的所有对 \mathbf{X} 进行线性变换后得到的随机变量中方差第二大的随机变量, 以此类推。

则分别称 $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$ 是第一、第二、……、第 n 主成分。

这一定义是否足够?

Theorem 5.7. 若不对 \mathcal{T} 的矩阵 A 作出相应的限制, 对 \mathbf{X} 进行线性变换后得到的 $\mathbf{Y}_i, i = 1, 2, \dots, n$ 的方差可以任意大。

Proof. 由性质 3.4.1(3) 可知:

$$\text{Var}(\mathbf{Y}_i) = \text{Cov}(\mathbf{Y})_{(i,i)} = \text{Cov}(A\mathbf{X})_{(i,i)} = (A\Sigma A^T)_{i,i} = \alpha_i \Sigma \alpha_i^T$$

若 $\text{Var}(\mathbf{Y}_i) > 0$, 取矩阵 $B = kA$, $\mathbf{Z} = kA\mathbf{X}$, 则:

$$\text{Var}(\mathbf{Y}_i) = (k\alpha_i)\Sigma(k\alpha_i)^T = k^2\alpha_i\Sigma\alpha_i$$

改变 k 的值, 即可对 $\mathbf{Y}_i, i = 1, 2, \dots, n$ 的方差进行任意的放缩。 □

因此, 我们需要对 A 进行相应的限制, 在这里我们人为地选择要求 A 是一个正交矩阵, 也就是让 $\alpha_i \alpha_i^T = 1$ 。

Definition 5.4. 设 \mathbf{X} 是一个 n 维随机向量, 其均值向量为 $\mu = (\mu_1, \mu_2, \dots, \mu_n)$ 、协方差矩阵为 $\Sigma = (\sigma_{ij}), i, j = 1, 2, \dots, n$ 。对 \mathbf{X} 进行一个线性变换 \mathcal{T} 得到一个 n 维随机向量 $\mathbf{Y} = \mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$, \mathcal{T} 的矩阵为:

$$A = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$$

若:

1. $AA^T = I$;
2. $\text{Cov}(\mathbf{Y})$ 是一个对角矩阵, 即 $\text{Cov}(\mathbf{Y}_i, \mathbf{Y}_j) = 0, i \neq j$;
3. \mathbf{Y}_1 是所有对 \mathbf{X} 进行线性变换后得到的随机变量中方差最大的随机变量, \mathbf{Y}_2 是与 \mathbf{Y}_1 不相关的所有对 \mathbf{X} 进行线性变换后得到的随机变量中方差第二大的随机变量, 以此类推。

则分别称 $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$ 是第一、第二、……、第 n 主成分。

Theorem 5.8. 设 \mathbf{X} 是一个 n 维随机向量, Σ 是其协方差矩阵, Σ 的特征值¹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 \mathbf{X} 的第 i 个主成分以及其方差为:

$$\mathbf{Y}_i = \varphi_i^T \mathbf{X}, \text{Var}(\mathbf{Y}_i) = \varphi_i^T \Sigma \varphi_i = \lambda_i$$

Proof. 考虑到:

$$\text{Var}(\mathbf{Y}_i) = \alpha_i \Sigma \alpha_i^T, \quad \text{Cov}(\mathbf{Y}_i, \mathbf{Y}_j) = \alpha_i \Sigma \alpha_j^T$$

求解主成分的过程即为求解:

$$\begin{aligned} \alpha_i &= \arg \max \alpha_i \Sigma \alpha_i^T \\ \text{s. t. } &\begin{cases} \|\alpha_i\| = 1, & i = 1, 2, \dots, n \\ \alpha_i \Sigma \alpha_j = 0, & j < i \end{cases} \end{aligned}$$

由定理 2.31 可知上述结论成立。 □

Definition 5.5. 将第 i 个主成分 \mathbf{Y}_i 与变量 \mathbf{X}_j 的相关系数 $\rho(\mathbf{Y}_i, \mathbf{X}_j)$ 称为因子负荷量 (*factor loading*)。可推得:

$$\rho(\mathbf{Y}_i, \mathbf{X}_j) = \frac{\sqrt{\lambda_i} \alpha_{ij}}{\sqrt{\sigma_{jj}}}, \quad i, j = 1, 2, \dots, n$$

推导 5.1. 由相关系数的定义:

$$\begin{aligned} \rho(\mathbf{Y}_i, \mathbf{X}_j) &= \frac{\text{Cov}(\mathbf{Y}_i, \mathbf{X}_j)}{\sqrt{\text{Var}(\mathbf{Y}_i) \text{Var}(\mathbf{X}_j)}} = \frac{\text{Cov}(\alpha_i \mathbf{X}, e_j^T \mathbf{X})}{\sqrt{\lambda_i \sigma_{jj}}} \\ &= \frac{\alpha_i \Sigma e_j}{\sqrt{\lambda_i \sigma_{jj}}} = \frac{e_j^T \Sigma \alpha_i}{\sqrt{\lambda_i \sigma_{jj}}} = \frac{e_j^T \lambda_i \alpha_i}{\sqrt{\lambda_i \sigma_{jj}}} = \frac{\sqrt{\lambda_i} \alpha_{ij}}{\sqrt{\sigma_{jj}}} \end{aligned}$$

Property 5.5.1. 总体主成分具有如下性质:

1. $\text{Cov}(\mathbf{Y}) = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$;
2. \mathbf{Y} 的方差之和等于 \mathbf{X} 的方差之和, 即 $\sum_{i=1}^n \lambda_i = \sum_{i=1}^n \sigma_{ii}$;

¹若特征多项式有重根, 则标准正交化特征向量组不唯一, 主成分也不唯一。

3. 第 i 个主成分与原变量的因子负荷量满足:

$$\sum_{j=1}^n \sigma_{jj} \rho^2(\mathbf{Y}_i, \mathbf{X}_j) = \lambda_i$$

4. 原变量的第 j 个分量与所有主成分的因子负荷量满足:

$$\sum_{i=1}^n \rho^2(\mathbf{Y}_i, \mathbf{X}_j) = 1$$

Proof. (1) 由定理 5.8 直接可得。

(2) 由性质 3.4.1(3) 和 ??(3) 可得:

$$\begin{aligned} \sum_{i=1}^n \text{Var}(\mathbf{Y}_i) &= \text{tr}[\text{Cov}(\mathbf{Y})] = \text{tr}[\text{Cov}(A\mathbf{X})] = \text{tr}(A\Sigma A^T) \\ &= \text{tr}(\Sigma A^T A) = \text{tr}(\Sigma) = \sum_{i=1}^n \text{Var}(\mathbf{X}_i) \end{aligned}$$

(3) 由推导 5.1 可得:

$$\sum_{j=1}^n \sigma_{jj} \rho^2(\mathbf{Y}_i, \mathbf{X}_j) = \sum_{j=1}^n \lambda_i \alpha_{ij}^2 = \lambda_i \alpha_i \alpha_i^T = \lambda_i$$

(4) 因为 A 是正交矩阵, 所以 A 可逆, 于是 \mathbf{X} 可以表示为 $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$ 的线性组合, 所以二者的复相关系数为 1。由可直接得出结论。□

复相关系数
性质

Definition 5.6. 称第 i 个主成分 \mathbf{Y}_i 的方差与所有主成分方差之和为 \mathbf{Y}_i 的方差贡献率, 记为 η_i , 即:

$$\eta_i = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j}$$

将:

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i}$$

称为主成分 $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_k$ 的累计方差贡献率。

Definition 5.7. 称主成分 $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_k$ 与变量 \mathbf{X}_j 之间的复相关系数的平方 R^2 为 $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_k$ 对 \mathbf{X}_j 的贡献率, 其计算公式为:

$$R^2 = \sum_{i=1}^k \frac{\lambda_i \alpha_{ij}^2}{\sigma_{ii}}$$

链接复相关
系数的性质

推导 5.2. 由推导 5.1 和直接可得。

由前述, 我们一般通过选择主成分的个数来实现对数据的降维, 即选择主成分的个数使它们的累计方差贡献率达到一定比例 (一般为 85%)。

5.5.2 样本主成分分析

假设对 n 维随机变量 \mathbf{X} 进行 m 次独立观测，得到 m 个 n 维样本 x_1, x_2, \dots, x_m 。在样本主成分分析中，我们使用样本来估计 \mathbf{X} 的协方差矩阵，即：

$$S = (s_{ij}), s_{ij} = \frac{1}{m-1} \sum_{k=1}^m (x_{ki} - \hat{\mathbf{X}}_i)(x_{kj} - \hat{\mathbf{X}}_j), i, j = 1, 2, \dots, n$$

其中：

$$\hat{\mathbf{X}}_i = \frac{1}{m} \sum_{j=1}^m x_{ji}, i = 1, 2, \dots, n$$

其余步骤与总体主成分分析一致。

5.5.3 注意事项

多重共线性问题

当原始变量出现多重共线性时，PCA 的效果会受到影响，这是因为重复的信息在方差占比中重复进行了计算。我们可以通过计算协方差矩阵的最小特征值来判断是否出现多重共线性的情况。若最小特征值趋于 0，则需要对纳入研究的变量进行考察与筛选。

写线性模型的时候再把多重共线性推导链接过来

相关矩阵导出主成分

上面我们都是对协方差矩阵的特征值分解进行计算，但在现实中，我们可能会对数据进行标准化处理来消除量纲带来的影响，注意到标准化后数据的协方差矩阵即为相关矩阵，此时将相关矩阵作对应的特征值分解即可。但需要注意：标准化后各变量方差相等均为 1，损失了部分信息，所以会使得标准化后的各变量在对主成分构成中的作用趋于相等。因此，取值范围在同量级的数据建议使用协方差矩阵直接求解主成分，若变量之间数量级差异较大，再使用相关矩阵求解主成分。

具体算法

Algorithm 1 主成分分析 (PCA)

```

1: Input: 原始数据矩阵  $\mathbf{X}$ 
2: Output: 主成分得分  $\mathbf{Y}$ , 选定的主成分个数  $k$ 
3: if 变量量纲差异明显 then
4:   标准化:  $\mathbf{Z} = \frac{\mathbf{X} - \mu}{\sigma}$ 
5: else
6:   保留原始数据  $\mathbf{Z} = \mathbf{X}$ 
7: end if
8: if 存在多重共线性 (最小特征值  $\approx 0$ ) then
9:   删除或合并相关变量
10: else
11:   保留原始变量
12: end if
13: 计算协方差矩阵:  $\Sigma = \text{cov}(\mathbf{Z})$ 
14: 对  $\Sigma$  进行特征值分解:  $\Sigma = \mathbf{A}\mathbf{\Lambda}\mathbf{A}^T$ 
15: 初始化  $k \leftarrow 1$ , 累计贡献率  $\leftarrow 0$ 
16: repeat
17:   选择第  $k$  个主成分
18:   更新累计贡献率
19:   if 累计贡献率 < 阈值 then
20:      $k \leftarrow k + 1$ 
21:   end if
22: until 累计贡献率  $\geq$  阈值
23: 计算主成分得分:  $\mathbf{Y} = \mathbf{Z}\mathbf{A}_k$  ( $\mathbf{A}_k$  维矩阵  $\mathbf{A}$  的前  $k$  行构成的矩阵)
24: 输出  $\mathbf{Y}$  和  $k$ 

```

5.6 因子分析

因子分析 (factor analysis) 的目的是从多个高度相关的观测变量中提取出少数几个潜在因子 (latent factor), 这些因子代表了变量背后的共通结构, 从而实现降维并提升可解释性。

假设对一组学生进行了以下六门课程的测试: 语文、英语、数学、物理、化学、生物, 发现语文和英语成绩之间高度相关, 数学、物理、化学、生物也彼此高度相关。此时可以猜测: 这些成绩可能是由两个更基本的“能力”决定的, 比如语言能力和理科能力。通过因子分析就可以提取出这两个潜在因子, 并发现语文和英语主要由“语言能力”因子决定, 理科四门主要由“理科能力”因子解释。这样就可以用两个因子有效地概括了六个变量的结构, 同时让模型更易解释、更简洁。

Definition 5.8. 设 \mathbf{X} 是一个可观测的 m 维随机向量, $E(\mathbf{X}) = \boldsymbol{\mu}$, $\text{Cov}(\mathbf{X}) = \Sigma = (\sigma_{ij})$ 。因子分析的数学模型为:

$$\begin{aligned} \mathbf{X} &= \boldsymbol{\mu} + A\mathbf{F} + \varepsilon \\ \begin{cases} E(\mathbf{F}) = \mathbf{0}, \text{Cov}(\mathbf{F}) = I_n \\ E(\varepsilon) = \mathbf{0}, \text{Cov}(\varepsilon) = D = \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2\} \\ \text{Cov}(\mathbf{F}, \varepsilon) = \mathbf{0} \end{cases} \end{aligned}$$

其中 $\mathbf{F} = (f_1, f_2, \dots, f_n)^T$ 是不可观测的 n 维随机变量, ε 是不可观测的 m 维随机变量, 分别称 \mathbf{F} 和 ε 为公共因子 (*common factor*) 和特殊因子 (*specific factor*)。 $A = (a_{ij})$ 是一个非随机矩阵, a_{ij} 表示公共因子 f_j 、随机变量 \mathbf{X}_i 的因子载荷。 $a_{1j}, a_{2j}, \dots, a_{mj}$ 中至少有两个不为 0, 否则可将 f_j 并入到 ε_j 中去; ε_i 也仅出现在 \mathbf{X}_i 的表达式中。

Property 5.6.1. 上述因子分析模型具有如下性质:

1. $\Sigma = AA^T + D$;
2. 模型不受单位影响。若 $\mathbf{X}^* = C\mathbf{X}$, 则有:

$$\mathbf{Y} = C\boldsymbol{\mu} + CAF + C\varepsilon = \boldsymbol{\mu}^* + A^*\mathbf{F} + \varepsilon^*$$

3. 因子载荷不唯一;
4. $\text{Cov}(\mathbf{X}, \mathbf{F}) = A$, 即 $\text{Cov}(\mathbf{X}_i, F_j) = a_{ij}$
5. 令 $h_i^2 = \sum_{j=1}^n a_{ij}^2$, 则有:

$$\text{Var}(\mathbf{X}_i) = \sigma_{ii} = \sum_{j=1}^n a_{ij}^2 + \sigma_i^2 = h_i^2 + \sigma_i^2, \quad i = 1, 2, \dots, m$$

6. 令 $g_j^2 = \sum_{i=1}^m a_{ij}^2$, 则有:

$$\sum_{i=1}^m \text{Var}(\mathbf{X}_i) = \sum_{j=1}^n g_j^2 + \sum_{i=1}^m \sigma_i^2$$

Proof. (1) 由性质 3.4.1(3)(4)(5) 可得:

$$\begin{aligned} \Sigma &= \text{Cov}(\mathbf{X}) = \text{Cov}(\boldsymbol{\mu} + A\mathbf{F} + \varepsilon, \boldsymbol{\mu} + A\mathbf{F} + \varepsilon) \\ &= \text{Cov}(\boldsymbol{\mu}, \boldsymbol{\mu} + A\mathbf{F} + \varepsilon) + \text{Cov}(A\mathbf{F}, \boldsymbol{\mu} + A\mathbf{F} + \varepsilon) + \text{Cov}(\varepsilon, \boldsymbol{\mu} + A\mathbf{F} + \varepsilon) \\ &= \text{Cov}(A\mathbf{F}, \boldsymbol{\mu}) + \text{Cov}(A\mathbf{F}) + \text{Cov}(A\mathbf{F}, \varepsilon) + \text{Cov}(\varepsilon, \boldsymbol{\mu}) + \text{Cov}(\varepsilon, A\mathbf{F}) + \text{Cov}(\varepsilon) \\ &= A \text{Cov}(\mathbf{F}) A^T + A \text{Cov}(\mathbf{F}, \varepsilon) + \text{Cov}(\varepsilon, \mathbf{F}) A^T + D \\ &= AA^T + D \end{aligned}$$

(2) 显然。

(3) 取正交矩阵 Q , 令 $A^* = AQ$, $F^* = Q^T F$, 由性质 3.4.1(3) 则依然有:

期望的性质

$$E(F^*) = Q^T E(F) = \mathbf{0}, \text{Cov}(F^*) = Q^T \text{Cov}(F)Q = I_n, \mathbf{X} = \boldsymbol{\mu} + A^* F^* + \varepsilon$$

(4) 由性质 3.4.1(3)(4)(5) 可得:

$$\text{Cov}(\mathbf{X}, F) = \text{Cov}(\boldsymbol{\mu} + AF + \varepsilon, F) = \text{Cov}(\boldsymbol{\mu}, F) + \text{Cov}(AF, F) + \text{Cov}(\varepsilon, F) = A$$

(5) 由 (1) 即可得到结论。

(6) 由性质 3.4.1(1)、(1) 和 ??(1) 可得:

$$\begin{aligned} \sum_{i=1}^m \text{Var}(\mathbf{X}_i) &= \text{tr}[\text{Cov}(\mathbf{X})] = \text{tr}(AA^T + D) = \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 + \sum_{i=1}^n \sigma_i^2 \\ &= \sum_{j=1}^n \sum_{i=1}^m a_{ij}^2 + \sum_{i=1}^n \sigma_i^2 = \sum_{j=1}^n g_j^2 + \sum_{i=1}^n \sigma_i^2 \end{aligned} \quad \square$$

Definition 5.9. 称 h_i^2 为变量 \mathbf{X}_i 的公共方差 (common variance), 它反映了公共因子对 \mathbf{X}_i 的方差贡献度。称 σ_i^2 为 \mathbf{X}_i 的特殊方差 (specific variance), 它反映了特殊因子 ε_i 对 \mathbf{X}_i 的方差贡献度。 g_j^2 可视为公共因子 f_j 对 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$ 的总方差贡献度。

5.6.1 参数估计方法

主成分法

方法 5.1. 设观测变量 \mathbf{X} 的协方差矩阵 Σ , 它的特征值从大到小依次为 $\lambda_1, \lambda_2, \dots, \lambda_m$, 对应的单位正交特征向量分别为 l_1, l_2, \dots, l_m 。于是 Σ 有分解式:

$$\Sigma = \begin{pmatrix} l_1 & l_2 & \cdots & l_m \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{pmatrix} \begin{pmatrix} l_1^T \\ l_2^T \\ \vdots \\ l_m^T \end{pmatrix} = \sum_{i=1}^m \lambda_i l_i l_i^T$$

由性质 3.4.1(2) 和定理 2.43(3) 的第五条可知 $\lambda_m \geq 0$ 。当最后 $m - n$ 个特征值较小时, Σ 有如下近似:

$$\Sigma = \sum_{i=1}^m \lambda_i l_i l_i^T \approx \sum_{i=1}^n \lambda_i l_i l_i^T + \hat{D} = \hat{A} \hat{A}^T + \hat{D}$$

其中:

$$\hat{A} = \begin{pmatrix} \sqrt{\lambda_1} l_1 & \cdots & \sqrt{\lambda_n} l_n \end{pmatrix}, \hat{D} = \text{diag}(\Sigma - \hat{A} \hat{A}^T)$$

与 PCA 一样, 一般通过使 $\left(\sum_{i=1}^n \lambda_i \right) / \left(\sum_{i=1}^m \lambda_i \right)$ 大于一定比例来选择 n 的具体值。

主因子法

方法 5.2. 令 $AA^T = \Sigma - D$ 。取 $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_m^2$ 为特殊方差的合理初始估计 ((1) 全零, (2) 取 $\max_{j \neq i} \sigma_{ij}$), 则有:

$$\widehat{AA^T} = \begin{pmatrix} \sigma_{11} - \hat{\sigma}_1^2 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} - \hat{\sigma}_2^2 & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} - \hat{\sigma}_m^2 \end{pmatrix}$$

取 $\widehat{AA^T}$ 前 n 个大于 0 的特征值, 从大到小依次为 $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_n$, 对应的单位正交特征向量为 $\hat{l}_1, \hat{l}_2, \dots, \hat{l}_n$, 则有近似的:

$$\hat{A} = \begin{pmatrix} \sqrt{\hat{\lambda}_1} \hat{l}_1 & \cdots & \sqrt{\hat{\lambda}_n} \hat{l}_n \end{pmatrix}$$

令 $\hat{\sigma}_i^2 = \sigma_{ii} - \hat{h}_i^2$, 继续上面的迭代过程以得到稳定的近似解。

Algorithm 2 主因子法求解因子分析

-
- 1: **Input:** 协方差矩阵 Σ , 初始特殊方差估计 $\hat{\sigma}_1^2, \dots, \hat{\sigma}_m^2$, 目标因子数 n
 - 2: **Output:** 因子载荷矩阵估计 \hat{A} , 特殊方差估计 $\hat{\sigma}_i^2$
 - 3: 初始化 $\hat{\sigma}_i^2$ 为合理值
 - 4: **repeat**
 - 5: 构造矩阵 $\widehat{AA^T} = \Sigma - \text{diag}(\hat{\sigma}_1^2, \dots, \hat{\sigma}_m^2)$
 - 6: 对 $\widehat{AA^T}$ 做特征值分解, 得到部分特征值 $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_n$, 及对应单位正交特征向量 $\hat{l}_1, \dots, \hat{l}_n$
 - 7: 构造因子载荷矩阵估计: $\hat{A} = (\hat{a}_{ij}) = \begin{pmatrix} \sqrt{\hat{\lambda}_1} \hat{l}_1 & \cdots & \sqrt{\hat{\lambda}_n} \hat{l}_n \end{pmatrix}$
 - 8: 令 $\hat{h}_i^2 = \sum_{j=1}^n \hat{a}_{ij}^2$, 更新 $\hat{\sigma}_i^2 = \sigma_{ii} - \hat{h}_i^2$, $i = 1, 2, \dots, m$
 - 9: **until** 特殊方差估计 $\hat{\sigma}_i^2$ 收敛或达到最大迭代次数
-

正态分布假设下的极大似然估计法

推导 5.3. 若假设 $F \sim N_n(\mathbf{0}, I_n)$, $\varepsilon \sim N_m(\mathbf{0}, D)$, 因为 F 和 ε 不相关, 由定理 4.7 可知 F 和 ε 独立。由定理 4.15 和定理 4.6 可得 $\mathbf{X} \sim N_m(\boldsymbol{\mu}, AA^T + D)$ 。对 \mathbf{X} 进行简单抽样获得 s 个样本, 由定理 4.6 和定理 4.3 可得这 s 个样本的均值 $\bar{\mathbf{X}} \sim N_n\left(\boldsymbol{\mu}, \frac{1}{s}(AA^T + D)\right)$ 。若样本均值为 \bar{x} , 则似然函数为:

$$L(A, D) = (2\pi)^{-\frac{m}{2}} |\det[\frac{1}{s}(AA^T + D)]|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}(\bar{x} - \boldsymbol{\mu})^T \left[\frac{1}{s}(AA^T + D) \right]^{-1} (\bar{x} - \boldsymbol{\mu}) \right\}$$

对数似然函数省去常数项即为：

$$\begin{aligned}\ln L(A, D) &= -\frac{1}{2} \ln \left\{ \left| \det \left[\frac{1}{s} (AA^T + D) \right] \right| \right\} - \frac{1}{2} (\bar{x} - \boldsymbol{\mu})^T \left[\frac{1}{s} (AA^T + D) \right]^{-1} (\bar{x} - \boldsymbol{\mu}) \\ &= -\frac{1}{2} \ln \left\{ \frac{1}{s^m} |\det(AA^T + D)| \right\} - \frac{s}{2} (\bar{x} - \boldsymbol{\mu})^T (AA^T + D)^{-1} (\bar{x} - \boldsymbol{\mu})\end{aligned}$$

5.6.2 因子旋转

为了提高因子的可解释性，我们希望每个因子对观测变量的影响是集中且明显的，即一个因子主要对少数几个变量有显著影响，对其余变量几乎没有作用。这种结构反映在因子载荷矩阵 A 上即为 A 每一列的元素 a_{ij} , $i = 1, 2, \dots, m$ 不是均匀地分布在中间水平，而是趋于两极分化：其绝对值要么接近于 0，要么较大。这样可以使得每个因子更容易被识别和解释——因为它只与一小组变量高度相关。这种结构等价于希望载荷矩阵 A 的每一列具有稀疏性，从而便于赋予因子明确的语义标签。

推导 5.4. 由性质 5.6.1(3) 可知在初步求得因子载荷矩阵 A 后，可以使用一个正交矩阵右乘 A ，此时仍能得到一个因子模型。使用正交矩阵来右乘 A 相当于是对因子 F 进行旋转变换，我们可以通过不断旋转 F 来得到更加稀疏的因子载荷矩阵，从而提高因子的可解释性。

如何旋转？怎么衡量旋转后因子载荷矩阵的优良性？

令：

$$d_{ij}^2 = \frac{a_{ij}^2}{h_i^2}, \quad i = 1, 2, \dots, m, j = 1, 2, \dots, n$$

d_{ij}^2 衡量了因子 j 对观测变量 \mathbf{X}_i 的影响，且消除了 a_{ij} 的正负号带来的差异和各观测变量在因子载荷大小上的不同带来的差异。定义第 j 列 p 个数据 d_{ij}^2 , $i = 1, 2, \dots, m$ 的方差为：

$$\begin{aligned}V_j &= \frac{1}{m} \sum_{i=1}^m (d_{ij}^2 - \bar{d}_j)^2 = \frac{1}{m} \sum_{i=1}^m \left(d_{ij}^2 - \frac{1}{p} \sum_{i=1}^m d_{ij}^2 \right)^2 \\ &= \frac{1}{m} \left[\sum_{i=1}^m d_{ij}^4 - m \frac{1}{m^2} \left(\sum_{i=1}^m d_{ij}^2 \right)^2 \right] \\ &= \frac{1}{m^2} \left[m \sum_{i=1}^m d_{ij}^4 - \frac{1}{m} \left(\sum_{i=1}^m d_{ij}^2 \right)^2 \right] \\ &= \frac{1}{m^2} \left[m \sum_{i=1}^m \frac{a_{ij}^4}{h_i^4} - \frac{1}{m} \left(\sum_{i=1}^m \frac{a_{ij}^2}{h_i^2} \right)^2 \right]\end{aligned}$$

若 V_j 越大，则第 j 个因子对观测变量的影响越集中。定义因子载荷矩阵 A 的方差为：

$$V = \sum_{j=1}^n V_j = \frac{1}{m^2} \left\{ \sum_{j=1}^n \left[m \sum_{i=1}^m \frac{a_{ij}^4}{h_i^4} - \frac{1}{m} \left(\sum_{i=1}^m \frac{a_{ij}^2}{h_i^2} \right)^2 \right] \right\}$$

若 V 越大，则表明因子对观测变量的影响越集中。

综上，我们只需使得旋转后得到的因子载荷矩阵 A 的方差 V 达到最大即可。

5.6.3 模型检验

由上面的讨论可以看出, 潜在因子的数目是一个超参数, 也是一个非常重要的参数, 我们该如何选择呢? 有没有什么办法能够确定这一超参数的值?

推导 5.5. 在正态性假设下 (仍需假设 \mathbf{X} 是 n 维正态随机向量), 我们可以对求解后的因子分析模型进行似然比检验。

设样本数为 p , 分别为 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p$, 都独立同分布于 $N_m(\boldsymbol{\mu}, \Sigma)$ 。构建似然比检验假设:

$$H_0: \Sigma = AA^T + D, \quad H_1: \Sigma \text{ 为其它任一正定矩阵}$$

由定义 4.2 可得此时备择假设下的对数似然函数为:

$$\begin{aligned} L_1 &= \sum_{i=1}^p \ln \left\{ \frac{1}{(2\pi)^{\frac{m}{2}} (\det \Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2} \text{tr}[(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})^T \Sigma^{-1}]} \right\} \\ &= \sum_{i=1}^p \left\{ -\frac{m}{2} \ln(2\pi) - \frac{1}{2} \ln(\det \Sigma) - \frac{1}{2} \text{tr}[(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})^T \Sigma^{-1}] \right\} \\ &= -\frac{p}{2} \left\{ m \ln(2\pi) + \ln(\det \Sigma) + \frac{1}{p} \sum_{i=1}^p \text{tr}[(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})^T \Sigma^{-1}] \right\} \end{aligned}$$

上式可以化为²:

$$L_1(\Sigma) = -\frac{p}{2} [m \ln(2\pi) + \ln(\det \Sigma) + \text{tr}(S\Sigma^{-1})]$$

其中 S 为样本协方差阵。这个似然函数在 $\Sigma = S$ 时取最大值, 于是:

证明

$$L_1 = -\frac{p}{2} [m \ln(2\pi) + \ln(\det S) + p]$$

同理, 此时原假设下的似然函数值为:

$$L_2 = -\frac{p}{2} [m \ln(2\pi) + \ln(\det \hat{\Sigma}) + \text{tr}(S\hat{\Sigma})]$$

其中 $\hat{\Sigma} = \hat{A}\hat{A}^T - \hat{D}$ 。

由似然比检验原理可知:

似然比检验

$$-2[L_2(\Sigma) - L_1(\Sigma)] \sim \chi_{df}^2$$

当

$$p[\ln(\det \hat{\Sigma}) + \text{tr}(S\hat{\Sigma}) - \ln(\det S) - p] > \chi_{0.95}^2(df)$$

时应拒绝原假设, 即 n 个因子不足以解释数据, 应增大因子个数。其中 $\chi_{0.95}^2(df)$ 为分布的 0.95 分位数。

自由度的计算

²样本因子分析时需要注意使用协方差矩阵的无偏估计。

5.6.4 因子得分

在拟合得到因子载荷矩阵后，我们可以反过来求解各样本因子的取值，这样一来就可以根据因子值去进行进一步的分析。例如在开头的例子里，我们可以得到每个学生语言能力与理科能力的值，进而可以进行分类或选择。因子得分有两种计算方式。

加权最小二乘法

推导 5.6. 考虑加权最小二乘函数：

$$\varphi(F) = (\mathbf{X} - \boldsymbol{\mu} - AF)^T D^{-1} (\mathbf{X} - \boldsymbol{\mu} - AF)$$

求：

$$\hat{F} = \arg \min \varphi(F)$$

由极值的必要条件得到：

$$\begin{aligned} \frac{\partial \varphi(F)}{\partial F} &= -2A^T D^{-1} (\mathbf{X} - \boldsymbol{\mu} - AF) = 0 \\ A^T D^{-1} (\mathbf{X} - \boldsymbol{\mu}) &= A^T D^{-1} AF \\ F &= (A^T D^{-1} A)^{-1} A^T D^{-1} (\mathbf{X} - \boldsymbol{\mu}) \end{aligned}$$

需要注意 $A^T D^{-1} A$ 的可逆性。

若认为 $\mathbf{X} \sim N_m(\boldsymbol{\mu} + AF, D)$ ，则上述解得的 F 也是极大似然估计的结果。

Definition 5.10. 称加权最小二乘法得到的因子得分为 *Bartlett* 因子得分。

从求解过程可以看出，该方法实际上是对特殊方差更大的变量施以更宽容的残差值。

回归法

推导 5.7. 设：

$$f_j = \sum_{i=1}^m b_{ji} \mathbf{X}_i + \varepsilon_j, \quad \text{Cov}(\mathbf{X}_i, \varepsilon_j) = 0, \quad j = 1, 2, \dots, n$$

由性质 5.6.1(4) 和性质 3.4.1(3)(5) 可知：

$$a_{ij} = \text{Cov}(\mathbf{X}_i, f_j) = \text{Cov} \left(\mathbf{X}_i, \sum_{k=1}^m b_{jk} \mathbf{X}_k + \varepsilon_j \right) = \sum_{k=1}^m \sigma_{ik} b_{jk}$$

令 $B = (b_{ij})$ ，则有：

$$A = \Sigma B^T$$

于是 $B = A^T \Sigma^{-1}$ ，需要注意 Σ 的可逆性。回归法的因子得分即为：

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

Chapter 6

线性模型

6.1 一般线性模型

Definition 6.1. 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 6.2. 称方程 $X^T X \beta = X^T y$ 为正则方程 (*normal equation*)。

Theorem 6.1. 对于定义 6.1, $\hat{\beta} = (X^T X)^- X^T y$ 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 2.2 和定理 2.6(1) 可知方程 $X^T X\beta = X^T y$ 是相容的, 根据定理 2.22 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 $\hat{\beta} = (X^T X)^{-1} X^T y$ 。□

推导 6.1. 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 2.42(3) 的第五点和可得 $X^T X$ 可逆。此时 $\hat{\beta} = (X^T X)^{-1} X^T y$, 称 $\hat{\beta}$ 为 β 的最小二乘估计 (least squares estimate, LSE)。

行列式等于特征值的积, 行列式大于 0 矩阵可逆

6.1.1 参数估计

回归系数

Definition 6.3. 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 6.1.1. 对于定义 6.1, $c^T \beta$ 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 2.5.1(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 2.5.1(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 3.4.1(3) 和性质 2.5.1(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 2.5.1(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 2.5.1(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 2.5.1(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 2.5.2(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 6.4. 对于定义 6.1, 若 $c^T \beta$ 是可估函数, 称 $c^T \hat{\beta}$ 为 $c^T \beta$ 的 LSE, 其中 $\hat{\beta}$ 为正则方程的解。

残差

Definition 6.5. 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 6.1.2. 对于定义 6.1, $\hat{\beta}$ 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 2.8.2(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 2.8.2(4) 可知 $I_n - P_X$ 是对称幂等阵, 所以由性质 3.4.1(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 6.2. 对于定义 6.1, $\hat{\beta}$ 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 6.1.2(2) 和定理 3.3 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 2.8.2(4)、性质 2.8.1(2)(3) 和性质 2.8.2(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:

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

□

Definition 6.6. 称 $\hat{\sigma}^2$ 为 σ^2 的 LSE。

6.1.2 约束最小二乘估计

Theorem 6.3. 对于定义 6.1, 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 2.22 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：

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

$A(X^T X)^- A^T$ 由可知：
的可逆性

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

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 2.5.1(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 2.5.1(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 6.4. 在定理 6.3 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 6.3 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 6.2 可知:

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

由性质 2.5.1(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 6.1.1(1) 可知 $A\beta$ 的每一个元素都是可估函数, 于是由定理 3.3 和性质 6.1.1(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。 □

6.1.3 实际计算

Theorem 6.5. 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 2.5.1(5) 可知:

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

由定理 6.3 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 6.7. 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 6.1. 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。

6.1.4 预测

假设要预测 m 个点 $x_{0i} = (x_{0i1}, x_{0i2}, \dots, x_{0ip})$, $i = 1, 2, \dots, m$, 它们所对应的因变量为 $y_{01}, y_{02}, \dots, y_{0m}$ 。已知 y_{0i} 和历史数据服从同一个线性模型, 即:

$$y_0 = X_0 \beta + \varepsilon_0, \quad E(\varepsilon_0) = \mathbf{0}, \quad \text{Cov}(\varepsilon_0) = \sigma^2 I_m$$

$$y_0 = \begin{pmatrix} y_{01} \\ y_{02} \\ \vdots \\ y_{0m} \end{pmatrix}, \quad X_0 = \begin{pmatrix} x_{011} & x_{012} & \cdots & x_{01p} \\ x_{021} & x_{022} & \cdots & x_{02p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{0m1} & x_{0m2} & \cdots & x_{0mp} \end{pmatrix}, \quad \varepsilon_0 = \begin{pmatrix} \varepsilon_{01} \\ \varepsilon_{02} \\ \vdots \\ \varepsilon_{0m} \end{pmatrix}$$

假设 $\mathcal{M}(X_0^T) \subseteq \mathcal{M}(X^T)$ 。

被测量与历史数据无关

推导 6.2. 此时使用 $E(y_0) = X_0\beta$ 的估计 $X_0\hat{\beta} = X_0(X^T X)^- X^T y$ 去进行预测。

由定理 2.2 可知 $\mathcal{M}(X^T) = \mathcal{M}(X^T X)$, 于是 $\mathcal{M}(X_0^T) \subseteq \mathcal{M}(X^T) = \mathcal{M}(X^T X)$, 由性质 2.5.1(3) 可知 $X_0(X^T X)^- X^T y$ 与广义逆 $(X^T X)^-$ 的选择无关。

因为 $\mathcal{M}(X_0^T) \subseteq \mathcal{M}(X^T)$, 所以 $X_0\beta$ 是可估的, 它具有可估函数的性质。但需注意, 因为 y_0 也是一个随机变量, 所以这里的无偏性指的是 $E(\hat{y}_0 - y_0) = 0$ 。

被测量与历史数据相关

推导 6.3. 在某些情况下, y_0 和 y 确实具有一定的相关性, 用 $\text{Cov}(\varepsilon, \varepsilon_0) = \sigma^2 V^T$ 来度量它们之间的相关性, 此时有:

$$\text{Cov}[(y, y_0)^T] = \text{Cov}[(\varepsilon, \varepsilon_0)^T] = \sigma^2 \begin{pmatrix} I_n & V^T \\ V & I_m \end{pmatrix}$$

Definition 6.8. 记 $\hat{y} = Cy$ 是 y 的一个线性无偏估计, 称:

$$\text{PMSE}(\hat{y}) = E[(\hat{y} - y)^T A (\hat{y} - y)]$$

为广义预测均方误差 (*generalized prediction MSE, PMSE*), 其中 $A > 0$ 。

Theorem 6.6. y_0 在广义预测均方误差意义下的最优线性无偏估计为:

$$\hat{y}_0 = X_0\hat{\beta} + V(y - X\hat{\beta})$$

其中 $\hat{\beta}$ 为正则方程的解。

Proof. 令 $\hat{y}_0 = Cy$ 是一个无偏估计, 由 ??(6) 和性质 3.4.1(3) 可得:

$$\begin{aligned} \hat{y}_0 - y_0 &= Cy - X_0\beta - \varepsilon_0 = CX\beta + C\varepsilon - X_0\beta - \varepsilon_0 = (CX - X_0)\beta + C\varepsilon - \varepsilon_0 \\ E(\hat{y}_0 - y_0) &= (CX - X_0)\beta = \mathbf{0} \text{ 对一切 } \beta \text{ 成立} \iff CX = X_0 \\ \text{Cov}(C\varepsilon - \varepsilon_0) &= E[(C\varepsilon - \varepsilon_0)(C\varepsilon - \varepsilon_0)^T] = E(C\varepsilon\varepsilon^T C^T - C\varepsilon\varepsilon_0^T - \varepsilon_0\varepsilon^T C^T + \varepsilon_0\varepsilon_0^T) \\ &= E(C\varepsilon\varepsilon^T C^T) - E(C\varepsilon\varepsilon_0^T) - E(\varepsilon_0\varepsilon^T C^T) + E(\varepsilon_0\varepsilon_0^T) \\ &= C E(\varepsilon\varepsilon^T) C^T - \text{Cov}(C\varepsilon, \varepsilon_0) - \text{Cov}(\varepsilon_0, C\varepsilon) + \text{Cov}(\varepsilon_0) \\ &= C \text{Cov}(\varepsilon) C^T - 2 \text{Cov}(C\varepsilon, \varepsilon_0) + \sigma^2 I_m \\ &= \sigma^2 C C^T - 2C \text{Cov}(\varepsilon, \varepsilon_0) + \sigma^2 I_m \\ &= \sigma^2 C C^T - 2\sigma^2 C V^T + \sigma^2 I_m = \sigma^2 (C C^T - 2C V^T + I_m) \end{aligned}$$

由定理 3.3 和 ??(2) 可得:

$$\begin{aligned} \text{PMSE}(\hat{y}_0) &= E[(\hat{y}_0 - y)^T A (\hat{y}_0 - y)] = E[(C\varepsilon - \varepsilon_0)^T A (C\varepsilon - \varepsilon_0)] \\ &= \text{tr}[A \sigma^2 (C C^T - 2C V^T + I_m)] = \sigma^2 \text{tr}[A (C C^T - 2C V^T + I_m)] \end{aligned}$$

接下来的目标就是求解：

$$\min_C \text{PMSE}(\hat{y}_0), \quad \text{s. t. } CX = X_0$$

使用 Lagrange 乘子法，构造辅助函数（不引入 X_0 是因为它是常数，对结果没影响）：

$$F(C, \Lambda) = \sigma^2 \text{tr}[A(CC^T - 2CV^T + I_m)] - 2 \text{tr}(CX\Lambda)$$

其中 Λ 是 Lagrange 乘子。由矩阵求导和??(1) 可得：

$$\begin{aligned} \frac{\partial \text{PSME}(\hat{y}_0)}{\partial C} &= \sigma^2 \frac{\partial \text{tr}(ACC^T)}{\partial C} - 2\sigma^2 \frac{\partial \text{tr}(ACV^T)}{\partial C} - 2 \frac{\partial \text{tr}(CX\Lambda)}{\partial C} \\ &= \sigma^2 2AC - 2\sigma^2 AV - 2\Lambda^T X^T \end{aligned}$$

令上式为 0 可得：

$$\begin{aligned} \sigma^2 AC &= \sigma^2 AV + \Lambda^T X^T \\ C &= V + \frac{A^{-1}\Lambda^T X^T}{\sigma^2} \end{aligned}$$

代入 $CX = X_0$ 可得：

$$VX + \frac{A^{-1}\Lambda^T X^T X}{\sigma^2} = X_0 \Lambda^T X^T X = \sigma^2 A(X_0 - VX)X^T X \Lambda = \sigma^2 (X_0^T - X^T V^T) A^T$$

由定理 2.2 可知 $\mathcal{M}(X^T) = \mathcal{M}(X^T X)$ ，而 $\mathcal{M}(X_0^T) \subseteq \mathcal{M}(X^T)$ ，所以上式等式右边矩阵的每一列都在 $\mathcal{M}(X^T X)$ 中，即方程组是相容的。由定理 2.22 可得：

$$\Lambda = \sigma^2 (X^T X)^- (X_0^T - X^T V^T) A^T$$

于是由性质 2.5.1(6) 可得：

$$\begin{aligned} C &= V + A^{-1} A (X_0 - VX) (X^T X)^- X^T = V + (X_0 - VX) (X^T X)^- X^T \\ &= X_0 (X^T X)^- X^T + V [I_n - X (X^T X)^- X^T] \end{aligned}$$

由性质 2.5.1(3)(4) 可知 C 是唯一的，与 $(X^T X)^-$ 的选择无关。所以：

$$\hat{y}_0 = X_0 (X^T X)^- X^T y + V [I_n - X (X^T X)^- X^T] y = X_0 \hat{\beta} + V(y - X\hat{\beta})$$

为 y_0 在广义预测均方误差意义下的最优线性无偏估计。 □

6.2 正态线性模型

6.2.1 参数估计

Definition 6.9. 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 6.2.1. 对于定义 6.9, 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 6.3 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 6.9, 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 6.1.1(1) 可知存在 α 使得 $c = X^T \alpha$, 根据性质 2.5.1(7) 可知:

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

由性质 2.5.1(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 4.3 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 2.8.2(4) 可得 $I_n - P_X$ 是对称阵, 所以由性质 6.1.2(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 4.15 可知 $\frac{\varepsilon}{\sigma} \sim N_n(\mathbf{0}, I_n)$ 。根据性质 2.8.2(4) 可知 $I_n - P_X$ 是对称幂等阵, 由性质 2.8.1(3) 和性质 2.8.2(1) 可得 $\text{rank}(I_n - P_X) = n - r$, 于是根据定理 4.10 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 6.1.2(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 2.8.2(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 4.11 可知 $c^T \hat{\beta}$ 与 $\hat{\sigma}^2$ 独立。

(4)

□

6.2.2 假设检验

Theorem 6.7. 对于定义 6.9, 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:

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

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 6.2.1(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 2.5.1(6)(7) 和性质 2.8.2(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 6.2.1(2) 可知:

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

$A(X^T X)^{-1} A^T$ 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 4.1(1) 可知:

的正定性

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

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 6.1.2(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 2.8.2(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 4.11 和定理 4.12 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 6.2.1(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。

6.2.3 置信域

置信椭圆

Theorem 6.8. 对于定义 6.9, 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 6.2.1(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 4.2 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 6.2.1(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:

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

note 6.2. 这里其实是一个未知方差构造 F 分布的思想。

Scheffe 置信区间

Theorem 6.9. 对于定义 6.9, 对任何可估函数 $l^T \beta$, 其中 $l \in \mathcal{M}(A^T)$ 且 $l \neq \mathbf{0}$, 其置信度为 $1 - \alpha$ 的同时置信区间为:

$$l^T \hat{\beta} \pm [k\hat{\sigma}^2 F_{k, n-r}(\alpha) l^T (X^T X)^{-1} l]^{\frac{1}{2}}$$

$A(X^T X)^{-1} A^T$ 的正定性 *Proof.* 由定理 6.8、与??可知:

$$\begin{aligned} 1 - \alpha &= P \left[(A\beta - A\hat{\beta})^T [A(X^T X)^{-1} A^T]^{-1} (A\beta - A\hat{\beta}) \leq k\hat{\sigma}^2 F_{k, n-r}(\alpha) \right] \\ &= P \left\{ \sup_{b \neq \mathbf{0}} \frac{[(A\hat{\beta} - A\beta)^T b]^2}{b^T A(X^T X)^{-1} A^T b} \leq k\hat{\sigma}^2 F_{k, n-r}(\alpha) \right\} \\ &= P \left\{ \frac{[(A\hat{\beta} - A\beta)^T b]^2}{b^T A(X^T X)^{-1} A^T b} \leq k\hat{\sigma}^2 F_{k, n-r}(\alpha), \text{ 对任意的 } b \neq \mathbf{0} \right\} \\ &= P \left\{ |(A\hat{\beta} - A\beta)^T b| \leq [k\hat{\sigma}^2 F_{k, n-r}(\alpha) b^T A(X^T X)^{-1} A^T b]^{\frac{1}{2}}, \text{ 对任意的 } b \neq \mathbf{0} \right\} \\ &= P \left\{ |\hat{\beta}^T A^T b - \beta^T A^T b| \leq [k\hat{\sigma}^2 F_{k, n-r}(\alpha) b^T A(X^T X)^{-1} A^T b]^{\frac{1}{2}}, \text{ 对任意的 } b \neq \mathbf{0} \right\} \end{aligned}$$

记 $A^T b = l$, 因为 $\mathcal{M}(A^T) \subseteq \mathcal{M}(X^T)$, 由性质 6.1.1(1) 可知 $l^T \beta$ 也是一个可估函数, 于是有:

$$1 - \alpha = P \left\{ |l^T \hat{\beta} - l^T \beta| \leq [k\hat{\sigma}^2 F_{k, n-r}(\alpha) l^T (X^T X)^{-1} l]^{\frac{1}{2}}, \text{ 对任意的 } l \in \mathcal{M}(A^T) \text{ 且 } l \neq \mathbf{0} \right\} \quad \square$$

Bonferroni 置信区间

Theorem 6.10. 对于定义 6.9, 记 A 的行分别为 $a_1^T, a_2^T, \dots, a_k^T$, 则 $a_i^T \beta$ 置信度为 $1 - \alpha$ 的 Bonferroni 置信区间为:

$$a_i^T \hat{\beta} \pm t_{n-r} \left(\frac{\alpha}{2k} \right) [\hat{\sigma}^2 a_i^T (X^T X)^{-1} a_i]^{\frac{1}{2}}$$

Proof. 由定理 6.8 可得当 $k = 1$ 时有:

$$\begin{aligned} 1 - \alpha &= P \left\{ (a_i^T \beta - a_i^T \hat{\beta})^T [a_i^T (X^T X)^{-1} a_i]^{-1} (a_i^T \beta - a_i^T \hat{\beta}) \leq \hat{\sigma}^2 F_{1, n-r}(\alpha) \right\} \\ &= P \left\{ (a_i^T \beta - a_i^T \hat{\beta})^2 \leq \hat{\sigma}^2 F_{1, n-r}(\alpha) [a_i^T (X^T X)^{-1} a_i] \right\} \end{aligned}$$

由性质 4.2.2(2) 可知 $F_{1,n-r} = t_{n-r}^2$, 因为服从 t 分布的变量可取负值而服从 F 分布的变量只能为正值, 所以上式把平方变成绝对值时应修改对应的 α 为 $\frac{\alpha}{2}$, 即此时:

$$1 - \alpha = P \left\{ |a_i^T \beta - a_i^T \hat{\beta}| \leq t_{n-r} \left(\frac{\alpha}{2} \right) [\hat{\sigma}^2 a_i^T (X^T X)^{-1} a_i]^{\frac{1}{2}} \right\}$$

由 Bonferroni 校正法可得出结论。 \square

比较

推导 6.4. *Scheffe* 区间与 *Bonferroni* 区间哪个更好? 由二者的公式可以看出只需选择:

$$\min \left\{ [k \hat{\sigma}^2 F_{k,n-r}(\alpha)]^{\frac{1}{2}}, \hat{\sigma} t_{n-r} \left(\frac{\alpha}{2k} \right) \right\}$$

对应的方法即能得到更短的置信区间。

6.2.4 区间预测

Theorem 6.11. 延续 section 6.1.4 处的定义, 假设 $\varepsilon \sim N(0, \sigma^2 I_n)$, $\varepsilon_0 \sim N(0, \sigma^2 I_m)$, 且有

$$\text{Cov}[(y, y_0)^T] = \text{Cov}[(\varepsilon, \varepsilon_0)^T] = \sigma^2 \begin{pmatrix} I_n & V^T \\ V & I_m \end{pmatrix}$$

则 $y_0^{(i)}$ 的 *Scheffe* 置信区间与 *Bonferroni* 置信区间分别为:

$$\hat{y}_0^{(i)} \pm [m \hat{\sigma}^2 F_{m,n-r}(\alpha) T_{ii}]^{\frac{1}{2}}, \quad \hat{y}_0^{(i)} \pm t_{n-r} \left(\frac{\alpha}{2m} \right) \hat{\sigma} (T_{ii})^{\frac{1}{2}}$$

其中 $y_0^{(i)}, \hat{y}_0^{(i)}$ 表示的是 y_0, \hat{y}_0 的第 i 个分量, \hat{y}_0 由定理 6.6 给出, T_{ii} 表示矩阵 T 的 i, i 元, $T = (X_0 - VX)(X^T X)^{-1}(X_0 - VX)^T + I_m - VV^T$ 。

Proof. 由定理 6.6、定理 4.3、性质 3.4.1(3)(5) 和性质 2.5.1(6)(7)(5) 可得:

$$\hat{y}_0 = X_0 \hat{\beta} + V(y - X \hat{\beta}) \sim N_m(X_0 \beta, \sigma^2 VV^T)$$

$$\begin{aligned} \text{Cov}(\hat{y}_0) &= \text{Cov}(Cy) = C \text{Cov}(y) C^T = \sigma^2 C C^T \\ &= \sigma^2 [X_0(X^T X)^{-1} X^T + V - VX(X^T X)^{-1} X^T] \\ &\quad \cdot [X(X^T X)^{-1} X_0^T + V^T - X(X^T X)^{-1} X^T V^T] \\ &= \sigma^2 [X_0(X^T X)^{-1} X^T X(X^T X)^{-1} X_0^T + X_0(X^T X)^{-1} X^T V^T \\ &\quad - X_0(X^T X)^{-1} X^T X(X^T X)^{-1} X^T V^T + VX(X^T X)^{-1} X_0^T \\ &\quad + VV^T - VX(X^T X)^{-1} X^T V^T - VX(X^T X)^{-1} X^T X(X^T X)^{-1} X_0^T \\ &\quad - VX(X^T X)^{-1} X^T V^T + VX(X^T X)^{-1} X^T X(X^T X)^{-1} X^T V^T] \\ &= \sigma^2 [X_0(X^T X)^{-1} X_0^T + X_0(X^T X)^{-1} X^T V^T - X_0(X^T X)^{-1} X^T V^T \\ &\quad + VX(X^T X)^{-1} X_0^T + VV^T - VX(X^T X)^{-1} X^T V^T \\ &\quad - VX(X^T X)^{-1} X_0^T - VX(X^T X)^{-1} X^T V^T + VX(X^T X)^{-1} X^T V^T] \\ &= \sigma^2 [X_0(X^T X)^{-1} X_0^T + VV^T - VX(X^T X)^{-1} X^T V^T] \end{aligned}$$

$$\begin{aligned}
\text{Cov}(\hat{y}_0, y_0) &= \text{Cov}[X_0\hat{\beta} + V(y - X\hat{\beta}), y_0] = \text{Cov}[Vy + (X_0 - VX)\hat{\beta}, y_0] \\
&= V \text{Cov}(y, y_0) + (X_0 - VX) \text{Cov}(\hat{\beta}, y_0) \\
&= V \text{Cov}(y, y_0) + (X_0 - VX) \text{Cov}[(X^T X)^{-1} X^T y, y_0] \\
&= V \text{Cov}(y, y_0) + (X_0 - VX)(X^T X)^{-1} X^T \text{Cov}(y, y_0) \\
&= [V + (X_0 - VX)(X^T X)^{-1} X^T] \text{Cov}(y, y_0) \\
&= \sigma^2 [V + (X_0 - VX)(X^T X)^{-1} X^T] V^T
\end{aligned}$$

令 $A = X_0(X^T X)^{-1} X_0^T + VV^T - VX(X^T X)^{-1} X^T V^T$, $B = [V + (X_0 - VX)(X^T X)^{-1} X^T] V^T$, 所以:

$$\begin{pmatrix} \hat{y}_0 \\ y_0 \end{pmatrix} \sim N_{m+n} \left[\begin{pmatrix} X_0\beta \\ X_0\beta \end{pmatrix}, \sigma^2 \begin{pmatrix} A & B \\ B^T & I_m \end{pmatrix} \right]$$

由定理 4.3 可知:

$$\hat{y}_0 - y_0 = (1, -1) \begin{pmatrix} \hat{y}_0 \\ y_0 \end{pmatrix} \sim N_m \{ \mathbf{0}, \sigma^2 [(X_0 - VX)(X^T X)^{-1} (X_0 - VX) + I_m - VV^T] \}$$

令 $T = (X_0 - VX)(X^T X)^{-1} (X_0 - VX)^T + I_m - VV^T$, 仿照定理 6.8 中的推导能得到此时 $\hat{y}_0 - y_0$ 置信度为 $1 - \alpha$ 的置信椭圆为:

$$[\hat{y}_0 - y_0 : (\hat{y}_0 - y_0)^T (T)^{-1} (\hat{y}_0 - y_0) \leq m\hat{\sigma}^2 F_{m, n-r}(\alpha)]$$

T 的可逆性
证明

其中 T 的可逆性由保证。

由上述置信椭圆, 仿照定理 6.9 的推导, 在其中取 b 为标准基向量即可得到 $y_0^{(i)}$ 的 Scheffe 置信区间为:

$$\hat{y}_0^{(i)} \pm [m\hat{\sigma}^2 F_{m, n-r}(\alpha) T_{ii}]^{\frac{1}{2}}$$

同理可得到 $y_0^{(i)}$ 的 Bonferroni 置信区间为:

$$\hat{y}_0^{(i)} \pm t_{n-r} \left(\frac{\alpha}{2m} \right) \hat{\sigma} (T_{ii})^{\frac{1}{2}}$$

□

6.3 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。

6.3.1 广义最小二乘估计

Definition 6.10. 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 6.5. 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 3.4.1(3) 可知定义 6.10 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, 由于可估函数的定义与协方差矩阵无关, 所以对于线性模型与正态线性模型的那些结论, 广义线性模型也可得到。

6.3.2 最小二乘统一理论

6.4 线性回归模型