

# Chapter 3

## 矩阵的几种重要分解

Matrix Theory

December 15, 2014

黄正华, 数学与统计学院, 武汉大学

3.1

### Contents

1	矩阵的 UR 分解及其推论	1
1.1	满秩方阵的 UR 分解	1
1.2	关于矩阵满秩分解的几个推论和应用	9
2	舒尔引理与正规矩阵的分解	11
2.1	舒尔引理	11
2.2	矩阵的奇异值分解	15
3	幂等矩阵、投影算子及矩阵的谱分解式	18
3.1	投影算子、幂等算子和幂等矩阵	18
3.2	可对角化矩阵的谱分解	26

3.2

## 1 矩阵的 UR 分解及其推论

### 1.1 满秩方阵的 UR 分解

**Definition 1.** 若一个上三角形矩阵的主对角线元素全为正实数, 则称该矩阵为正线上三角阵.

**Theorem 2.** 设  $A \in \mathbb{C}_n^{n \times n}$ , 则存在酉矩阵  $U \in \mathbb{C}^{n \times n}$  及正线上三角阵  $R \in \mathbb{C}_n^{n \times n}$ , 使得  $A = UR$ .

称为满秩方阵的 UR 分解.

**证:** 因为  $A \in \mathbb{C}_n^{n \times n}$ , 故  $\text{rank } A = n$ . 记  $A = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ , 则  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  线性无关. 由施密特正交化过程:

- $\mathbf{u}_1 = \mathbf{x}_1$ ,  
 $\mathbf{y}_1 = \frac{1}{\|\mathbf{u}_1\|} \mathbf{u}_1$ ;

$$2. \mathbf{u}_{m+1} = \mathbf{x}_{m+1} - (\mathbf{x}_{m+1} | \mathbf{y}_1)\mathbf{y}_1 - (\mathbf{x}_{m+1} | \mathbf{y}_2)\mathbf{y}_2 - \cdots - (\mathbf{x}_{m+1} | \mathbf{y}_m)\mathbf{y}_m, \\ \mathbf{y}_{m+1} = \frac{1}{\|\mathbf{u}_{m+1}\|}\mathbf{u}_{m+1}.$$

得到一组标准正交基  $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n$ . 由施密特正交化过程的计算公式得

- $\mathbf{x}_1 = \|\mathbf{u}_1\|\mathbf{y}_1;$
- $\mathbf{x}_i = (\mathbf{x}_i | \mathbf{y}_1)\mathbf{y}_1 + (\mathbf{x}_i | \mathbf{y}_2)\mathbf{y}_2 + \cdots + (\mathbf{x}_i | \mathbf{y}_{i-1})\mathbf{y}_{i-1} + \|\mathbf{u}_i\|\mathbf{y}_i.$

用矩阵形式表达, 则


$$[\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n]\mathbf{R},$$

其中

$$\mathbf{R} = \begin{bmatrix} \|\mathbf{u}_1\| & (\mathbf{x}_2 | \mathbf{y}_1) & (\mathbf{x}_3 | \mathbf{y}_1) & \cdots & (\mathbf{x}_{n-1} | \mathbf{y}_1) & (\mathbf{x}_n | \mathbf{y}_1) \\ 0 & \|\mathbf{u}_2\| & (\mathbf{x}_3 | \mathbf{y}_2) & \cdots & (\mathbf{x}_{n-1} | \mathbf{y}_2) & (\mathbf{x}_n | \mathbf{y}_2) \\ 0 & 0 & \|\mathbf{u}_3\| & \cdots & (\mathbf{x}_{n-1} | \mathbf{y}_3) & (\mathbf{x}_n | \mathbf{y}_3) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \|\mathbf{u}_{n-1}\| & (\mathbf{x}_n | \mathbf{y}_{n-1}) \\ 0 & 0 & 0 & \cdots & 0 & \|\mathbf{u}_n\| \end{bmatrix},$$

注意到  $\mathbf{R}$  为正线上三角矩阵,  $\mathbf{U} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n]$  为酉矩阵, 故得到满足条件的分解  $\mathbf{A} = \mathbf{UR}$ . □

**Corollary 3.** 设  $\mathbf{A} \in \mathbb{R}_n^{n \times n}$ , 则存在正交矩阵  $\mathbf{Q} \in \mathbb{R}^{n \times n}$  及正线上三角阵  $\mathbf{R} \in \mathbb{R}_n^{n \times n}$ , 使得  $\mathbf{A} = \mathbf{QR}$ .

 称为矩阵的 QR 分解. UR 分解和 QR 分解是一致的概念, 更多的教材把它们统称为 QR 分解.

*Example 4.* 试求矩阵  $\mathbf{A} = \begin{bmatrix} 0 & 3 & 1 \\ 0 & 4 & -2 \\ 2 & 1 & 2 \end{bmatrix}$  的 UR 分解.

**解:** 将  $\mathbf{x}_1 = (0, 0, 2)^T$ ,  $\mathbf{x}_2 = (3, 4, 1)^T$ ,  $\mathbf{x}_3 = (1, -2, 2)^T$  标准正交化.

$$\mathbf{u}_1 = \mathbf{x}_1 = (0, 0, 2)^T, \mathbf{y}_1 = \frac{1}{\|\mathbf{u}_1\|}\mathbf{u}_1 = (0, 0, 1)^T.$$

$$\mathbf{u}_2 = \mathbf{x}_2 - (\mathbf{x}_2 | \mathbf{y}_1)\mathbf{y}_1 = (3, 4, 1)^T - (0, 0, 1) \begin{pmatrix} 3 \\ 4 \\ 1 \end{pmatrix} \mathbf{y}_1 = (3, 4, 0)^T.$$

$$\mathbf{y}_2 = \frac{1}{\|\mathbf{u}_2\|}\mathbf{u}_2 = \left(\frac{3}{5}, \frac{4}{5}, 0\right)^T.$$

$$\mathbf{u}_3 = \mathbf{x}_3 - (\mathbf{x}_3 | \mathbf{y}_1)\mathbf{y}_1 - (\mathbf{x}_3 | \mathbf{y}_2)\mathbf{y}_2 = \mathbf{x}_3 - 2\mathbf{y}_1 + \mathbf{y}_2 = \frac{1}{5}(8, -6, 0)^T.$$

$$\mathbf{y}_3 = \frac{1}{\|\mathbf{u}_3\|}\mathbf{u}_3 = \left(\frac{4}{5}, -\frac{3}{5}, 0\right)^T.$$

由正交化过程得

$$\begin{aligned} \mathbf{x}_1 &= 2\mathbf{y}_1, \\ \mathbf{x}_2 &= \mathbf{y}_1 + 5\mathbf{y}_2, \\ \mathbf{x}_3 &= 2\mathbf{y}_1 - \mathbf{y}_2 + 2\mathbf{y}_3, \end{aligned}$$

故

$$[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3] = [\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3] \begin{bmatrix} 2 & 1 & 2 \\ 0 & 5 & -1 \\ 0 & 0 & 2 \end{bmatrix},$$

得满足条件的 UR 分解为

$$\mathbf{A} = \mathbf{U}\mathbf{R} = \begin{bmatrix} 0 & \frac{3}{5} & \frac{4}{5} \\ 0 & \frac{4}{5} & -\frac{3}{5} \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 2 \\ 0 & 5 & -1 \\ 0 & 0 & 2 \end{bmatrix}. \quad \square$$

或者

$$\mathbf{R} = \begin{bmatrix} \|\mathbf{u}_1\| & (\mathbf{x}_2 | \mathbf{y}_1) & (\mathbf{x}_3 | \mathbf{y}_1) \\ 0 & \|\mathbf{u}_2\| & (\mathbf{x}_3 | \mathbf{y}_2) \\ 0 & 0 & \|\mathbf{u}_3\| \end{bmatrix} = \begin{bmatrix} 2 & 1 & 2 \\ 0 & 5 & -1 \\ 0 & 0 & 2 \end{bmatrix}.$$

3.7

3.8

**Corollary 5.** 设  $\mathbf{A} \in \mathbb{C}_n^{m \times n}$ , 则存在酉矩阵  $\mathbf{U} \in \mathbb{C}^{m \times m}$  及正线上三角阵  $\mathbf{R}_1 \in \mathbb{C}_n^{n \times n}$ , 使得

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{O} \end{bmatrix},$$

其中  $\mathbf{O} \in \mathbb{C}^{(m-n) \times n}$ ,  $m > n$ .

**证:** 记  $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$ , 则  $\mathbf{a}_i \in \mathbb{C}^m$ . 因  $\text{rank } \mathbf{A} = n$ , 故  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  线性无关. 注意到  $m > n$ , 可将它们扩充为  $\mathbb{C}^m$  的基底, 记为

$$\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n, \mathbf{a}_{n+1}, \dots, \mathbf{a}_m\}.$$

由施密特正交化过程, 得到一组标准正交基  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$ , 且有正线上三角矩阵  $\mathbf{R}$ , 使得

$$[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n, \mathbf{a}_{n+1}, \dots, \mathbf{a}_m] = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n, \mathbf{y}_{n+1}, \dots, \mathbf{y}_m] \mathbf{R},$$

取  $\mathbf{R}$  的前  $n$  列, 记为矩阵

$$\begin{bmatrix} \mathbf{R}_1 \\ \mathbf{O} \end{bmatrix}, \quad \mathbf{R}_1 \in \mathbb{C}_n^{n \times n}, \quad \mathbf{O} \in \mathbb{C}^{(m-n) \times n}.$$

其中  $\mathbf{R}_1$  取至于  $\mathbf{R}$  的前  $n$  行、前  $n$  列, 仍然是一个正线上三角矩阵, 则有

$$[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n, \mathbf{y}_{n+1}, \dots, \mathbf{y}_m] \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{O} \end{bmatrix},$$

3.9

记  $U = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n, \mathbf{y}_{n+1}, \dots, \mathbf{y}_m]$ , 得

$$A = U \begin{bmatrix} R_1 \\ O \end{bmatrix}. \quad \square$$

3.10

**Corollary 6.** 设  $A \in \mathbb{C}_r^{m \times n}$ , 则存在酉矩阵  $U \in \mathbb{C}^{m \times m}$  和  $V \in \mathbb{C}^{n \times n}$  及正线上三角阵  $R \in \mathbb{C}_r^{r \times r}$ , 使得

$$A = U \begin{bmatrix} R & O_{r \times (n-r)} \\ O_{(m-r) \times r} & O_{(n-r) \times (m-r)} \end{bmatrix} V^H. \quad (1)$$

称为长方阵的正交分解.

3.11

回顾: 正定矩阵

**Definition 7.** 设有二次型  $f = \mathbf{x}^T A \mathbf{x}$ , 如果对于任何  $\mathbf{x} \neq \mathbf{0}$ , 都有  $f(\mathbf{x}) > 0$  (显然  $f(\mathbf{0}) = 0$ ), 则称  $f$  为正定二次型, 并称矩阵  $A$  是正定的.

**Theorem 8.**  $\mathbf{x}^T A \mathbf{x}$  是正定二次型 (或  $A$  是正定矩阵) 的充要条件是下列任何之一:

- (1)  $A$  的正惯性指数为  $n$ .
- (2) 存在可逆矩阵  $P$ , 使得  $A = P^T P$ .
- (3)  $A$  的  $n$  个特征值全为正.
- (4)  $A$  的  $n$  个顺序主子式全为正.

3.12

**Corollary 9.** 设  $A \in \mathbb{R}^{n \times n}$  是一个正定矩阵, 则存在正线上三角形  $R$ , 使得  $A = R^T R$ .

称为 Cholesky 分解或三角 — 三角分解.

**证:** 因为  $A$  为正定矩阵, 故存在满秩矩阵  $P$ , 使得  $A = P^T P$ .

对满秩矩阵  $P$ , 存在正交矩阵  $Q$  及正线上三角矩阵  $R$ , 使得

$$P = QR.$$

因此

$$A = P^T P = R^T Q^T Q R = R^T R. \quad \square$$

3.13

**The Cholesky Factorization** is named in honor of the French military officer Major **André-Louis Cholesky** (1875–1918). Although originally assigned to an artillery branch, Cholesky later became attached to the Geodesic Section of the Geographic Service in France where he became noticed for his extraordinary intelligence and his facility for mathematics. From 1905 to 1909 Cholesky was involved with the problem of adjusting the triangularization grid for France. This was a huge computational task, and there were arguments as to what computational techniques should be employed. It was

during this period that Cholesky invented the ingenious procedure for solving a positive definite system of equations that is the basis for the matrix factorization that now bears his name. Unfortunately, Cholesky's mathematical talents were never allowed to flower. In 1914 war broke out, and Cholesky was again placed in an artillery group—but this time as the commander. On August 31, 1918, Major Cholesky was killed in battle. Cholesky never had time to publish his clever computational methods—they were carried forward by word-of-mouth. Issues surrounding the Cholesky factorization have been independently rediscovered several times by people who were unaware of Cholesky, and, in some circles, the Cholesky factorization is known as the *square root method*.

**Theorem 10** (满秩分解定理). 设  $A \in \mathbb{C}_r^{m \times n}$ , 则存在列满秩矩阵  $B \in \mathbb{C}_r^{m \times r}$  和行满秩矩阵  $C \in \mathbb{C}_r^{r \times n}$ , 使得

$$A = BC.$$

**证:** 将矩阵  $A$  化为标准形, 即存在  $m$  阶可逆矩阵  $P$  和  $n$  阶可逆矩阵  $Q$ , 使得

$$PAQ = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix},$$

则

$$A = P^{-1} \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} Q^{-1} = P^{-1} \begin{bmatrix} I_r \\ O \end{bmatrix} [I_r, O] Q^{-1},$$

取

$$B = P^{-1} \begin{bmatrix} I_r \\ O \end{bmatrix}, \quad C = [I_r, O] Q^{-1},$$

则  $B \in \mathbb{C}_r^{m \times r}$ ,  $C \in \mathbb{C}_r^{r \times n}$ , 得  $A = BC$ . □

**另证:** 因  $\text{rank } A = r$ , 对  $A$  进行初等行变换, 可将  $A$  化为阶梯形矩阵  $\begin{bmatrix} C \\ O \end{bmatrix}$ , 其中  $C \in \mathbb{C}_r^{r \times n}$ .

于是存在  $m$  阶可逆矩阵  $P$ , 使得  $PA = \begin{bmatrix} C \\ O \end{bmatrix}$ , 即


$$A = P^{-1} \begin{bmatrix} C_{r \times n} \\ O_{(m-r) \times n} \end{bmatrix}.$$

取  $P^{-1}$  的前  $r$  列, 记为矩阵  $B$ , 即分块为

$$P^{-1} = [B, F], \quad B \in \mathbb{C}_r^{m \times r}, \quad F \in \mathbb{C}^{m \times (m-r)},$$

则有

$$A = [B_{m \times r}, F_{m \times (m-r)}] \begin{bmatrix} C_{r \times n} \\ O_{(m-r) \times n} \end{bmatrix} = BC. \quad \square$$

 事实上,  $C$  只要行满秩就可以了, 不需要一定是阶梯形.

**Example 11.** 求矩阵  $A = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 2 & -1 & 1 \\ 2 & 2 & -2 & -1 \end{bmatrix}$  的满秩分解.

**解:** 由定理证明过程, 需要求出阶梯形矩阵  $\begin{bmatrix} C \\ O \end{bmatrix}$  及初等矩阵的乘积  $P$ . 为此, 对

矩阵  $[A, I]$  进行初等行变换, 当  $A$  成为阶梯形矩阵  $\begin{bmatrix} C \\ O \end{bmatrix}$  时,  $I$  转化为  $P$ .

$$[A, I] = \left[ \begin{array}{cccc|cccc} -1 & 0 & 1 & 2 & 1 & 0 & 0 & 0 \\ 1 & 2 & -1 & 1 & 0 & 1 & 0 & 0 \\ 2 & 2 & -2 & -1 & 0 & 0 & 1 & 0 \end{array} \right] \xrightarrow{\text{行变换}} \left[ \begin{array}{cccc|cccc} -1 & 0 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 3 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 1 & 0 \end{array} \right].$$

故

$$C = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 3 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix},$$

进一步算得

$$P^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & 1 & 1 \end{bmatrix}.$$

注意到  $\text{rank } A = 2$ , 于是取  $P^{-1}$  的前 2 列得到  $B$ , 即

$$A = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 3 \end{bmatrix}. \quad \square$$

**☞** 需要指出的是, 矩阵的满秩分解不是唯一的. 这是因为任取  $r$  阶可逆矩阵  $D$  ( $D$  不是单位矩阵), 则

$$A = BC = (BD)(D^{-1}C) = \tilde{B}\tilde{C},$$

其中  $\tilde{B} = BD \in \mathbb{C}^{m \times r}$ ,  $\tilde{C} = D^{-1}C \in \mathbb{C}^{r \times n}$ , 从而得到  $A$  的另一个满秩分解.

关于满秩分解, 下面再介绍一种较简便的方法.

*Example 12.* 设矩阵  $A$  经过初等行变换得到其行最简形矩阵为

$$M = \begin{pmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 0 \end{pmatrix}.$$

记  $M = (\beta_1, \beta_2, \beta_3)$ , 则  $\beta_1, \beta_2$  为  $M$  的列向量的极大无关组, 且  $\beta_3 = a\beta_1 + b\beta_2$ .

记  $A = (\alpha_1, \alpha_2, \alpha_3)$ , 注意到初等行变换不改变矩阵列向量的线性关系, 则  $\alpha_1, \alpha_2$  为  $A$  的列向量的极大无关组, 且  $\alpha_3 = a\alpha_1 + b\alpha_2$ . 即

$$\begin{aligned} A &= (\alpha_1, \alpha_2, \alpha_3) = (\alpha_1, \alpha_2, a\alpha_1 + b\alpha_2) \\ &= (\alpha_1, \alpha_2) \begin{pmatrix} 1 & 0 & a \\ 0 & 1 & b \end{pmatrix}. \end{aligned}$$

从而得到矩阵  $A$  的一个满秩分解.



取  $G$  的前  $r$  行构成矩阵  $C$ , 则  $C \in \mathbb{C}_r^{r \times n}$ .

将矩阵  $G$  的第  $k_1, k_2, \dots, k_r$  列记为向量组

$$\mathcal{G} = \{\mathbf{g}_{k_1}, \mathbf{g}_{k_2}, \dots, \mathbf{g}_{k_r}\},$$

则  $\mathcal{G}$  就是  $m$  维基本单位向量  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_r$ .

记  $C$  的第  $i$  个列向量为

$$\mathbf{c}_i = \begin{bmatrix} c_{1i} \\ c_{2i} \\ \vdots \\ c_{ri} \end{bmatrix},$$

则  $G$  的第  $i$  个列向量  $\mathbf{g}_i$  为

$$\mathbf{g}_i = \begin{bmatrix} c_{1i} \\ c_{2i} \\ \vdots \\ c_{ri} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

从而  $\mathbf{g}_i = c_{1i}\mathbf{e}_1 + c_{2i}\mathbf{e}_2 + \dots + c_{ri}\mathbf{e}_r$ , 即

$$\mathbf{g}_i = c_{1i}\mathbf{g}_{k_1} + c_{2i}\mathbf{g}_{k_2} + \dots + c_{ri}\mathbf{g}_{k_r}. \quad (2)$$

设有  $m$  阶可逆矩阵  $P$  使得  $PA = G$ , 则  $A = P^{-1}G$ , 即

$$[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] = [P^{-1}\mathbf{g}_1, P^{-1}\mathbf{g}_2, \dots, P^{-1}\mathbf{g}_n].$$

则  $\mathbf{a}_i = P^{-1}\mathbf{g}_i, i = 1, 2, \dots, n$ .

在 (2) 两边左乘以  $P^{-1}$ , 得

$$P^{-1}\mathbf{g}_i = c_{1i}P^{-1}\mathbf{g}_{k_1} + c_{2i}P^{-1}\mathbf{g}_{k_2} + \dots + c_{ri}P^{-1}\mathbf{g}_{k_r}. \quad (3)$$

即

$$\mathbf{a}_i = c_{1i}\mathbf{a}_{k_1} + c_{2i}\mathbf{a}_{k_2} + \dots + c_{ri}\mathbf{a}_{k_r}.$$

则

$$\mathbf{a}_i = [\mathbf{a}_{k_1}, \mathbf{a}_{k_2}, \dots, \mathbf{a}_{k_r}] \begin{bmatrix} c_{1i} \\ c_{2i} \\ \vdots \\ c_{ri} \end{bmatrix} = B\mathbf{c}_i.$$

故

$$[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] = B[\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n],$$

即

$$A = BC.$$

3.23

3.24



最后还需要说明  $B = [\mathbf{a}_{k_1}, \mathbf{a}_{k_2}, \dots, \mathbf{a}_{k_r}]$  是列满秩的. 设

$$t_1 \mathbf{a}_{k_1} + t_2 \mathbf{a}_{k_2} + \dots + t_r \mathbf{a}_{k_r} = \mathbf{0}, \quad (4)$$

两边左乘以  $P$ , 得  $t_1 P\mathbf{a}_{k_1} + t_2 P\mathbf{a}_{k_2} + \dots + t_r P\mathbf{a}_{k_r} = \mathbf{0}$ , 即

$$t_1 \mathbf{g}_{k_1} + t_2 \mathbf{g}_{k_2} + \dots + t_r \mathbf{g}_{k_r} = \mathbf{0},$$

亦即

$$t_1 \mathbf{e}_1 + t_2 \mathbf{e}_2 + \dots + t_r \mathbf{e}_r = \mathbf{0},$$

故要使 (4) 式成立, 只能有

$$t_1 = t_2 = \dots = t_r = 0,$$

从而  $\mathbf{a}_{k_1}, \mathbf{a}_{k_2}, \dots, \mathbf{a}_{k_r}$  线性无关, 所以  $B$  是列满秩的.  $\square$

*Example 14.* 求矩阵  $A = \begin{bmatrix} 2 & 4 & 1 & 1 \\ 1 & 2 & -1 & 2 \\ -1 & -2 & -2 & 1 \end{bmatrix}$  的满秩分解.

**解:** 将  $A$  进行初等行变换, 化为行最简形, 得

$$A \xrightarrow{\text{初等行变换}} \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

可见  $k_1 = 1, k_2 = 3$ , 故需要取  $A$  的第 1 列、第 3 列构成矩阵  $B$ , 所以

$$A = \begin{bmatrix} 2 & 1 \\ 1 & -1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}. \quad \square$$

## 1.2 关于矩阵满秩分解的几个推论和应用

**回顾:** 半正定矩阵

**Definition 15.** 对任意  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \neq \mathbf{0}$ ,  $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$ , 但至少存在一个  $\mathbf{x}_0 \neq \mathbf{0}$ , 使得  $\mathbf{x}_0^T \mathbf{A} \mathbf{x}_0 = 0$ , 就称  $\mathbf{x}^T \mathbf{A} \mathbf{x}$  是半正定二次型,  $\mathbf{A}$  是半正定矩阵.

**Theorem 16.** 设  $\mathbf{A}$  是  $n$  阶实对称矩阵, 则下列命题等价:

- (i)  $\mathbf{x}^T \mathbf{A} \mathbf{x}$  半正定;
- (ii)  $\mathbf{A}$  的正惯性指数小于  $n$ .
- (iii) 存在降秩矩阵  $\mathbf{P}$  (即  $\text{rank}(\mathbf{P}) < n$ ), 使得  $\mathbf{A} = \mathbf{P}^T \mathbf{P}$ .
- (iv)  $\mathbf{A}$  的  $n$  个特征值全为非负, 但至少有一个等于 0.
- (v)  $\mathbf{A}$  的各阶顺序主子式非负, 且至少有一个顺序主子式等于 0.

**Theorem 17.** 对任意的矩阵  $\mathbf{A} \in \mathbb{C}_r^{m \times n}$ , 都有

1.  $\mathbf{A}^H \mathbf{A}$  与  $\mathbf{A} \mathbf{A}^H$  都是半正定的 Hermite 矩阵;

2.  $\text{rank}(\mathbf{A}^H \mathbf{A}) = \text{rank}(\mathbf{A} \mathbf{A}^H) = \text{rank} \mathbf{A} = r$ .

证: (1)  $\forall \mathbf{x} \in \mathbb{C}^n, \mathbf{y} \in \mathbb{C}^m$ , 有

$$\begin{aligned} \mathbf{x}^H \mathbf{A}^H \mathbf{A} \mathbf{x} &= (\mathbf{A} \mathbf{x} | \mathbf{A} \mathbf{x}) \geq 0, \\ \mathbf{y}^H \mathbf{A} \mathbf{A}^H \mathbf{y} &= (\mathbf{A}^H \mathbf{y} | \mathbf{A}^H \mathbf{y}) \geq 0. \end{aligned}$$

故  $\mathbf{A}^H \mathbf{A}$  与  $\mathbf{A} \mathbf{A}^H$  都是半正定的, 且显然为 Hermite 矩阵.

(2) 只证  $\text{rank}(\mathbf{A}^H \mathbf{A}) = \text{rank} \mathbf{A}$ . 下证方程  $\mathbf{A}^H \mathbf{A} \mathbf{x} = \mathbf{0}$  与  $\mathbf{A} \mathbf{x} = \mathbf{0}$  同解.

显然  $\mathbf{A} \mathbf{x} = \mathbf{0}$  的解都是  $\mathbf{A}^H \mathbf{A} \mathbf{x} = \mathbf{0}$  的解. 反过来, 假设  $\mathbf{x}_1$  是方程  $\mathbf{A}^H \mathbf{A} \mathbf{x} = \mathbf{0}$  的解, 即  $\mathbf{A}^H \mathbf{A} \mathbf{x}_1 = \mathbf{0}$ , 则

$$0 = \mathbf{x}_1^H \mathbf{A}^H \mathbf{A} \mathbf{x}_1 = (\mathbf{A} \mathbf{x}_1 | \mathbf{A} \mathbf{x}_1),$$

故  $\mathbf{A} \mathbf{x}_1 = \mathbf{0}$ . □

3.29

### 左逆、右逆

**Definition 18.** 设  $\mathbf{A} \in \mathbb{C}^{m \times n}, \mathbf{X} \in \mathbb{C}^{n \times m}$ .

1. 若  $\mathbf{X} \mathbf{A} = \mathbf{I}$ , 则称  $\mathbf{X}$  为  $\mathbf{A}$  的左逆, 记为  $\mathbf{A}_L^{-1}$ .

2. 若  $\mathbf{A} \mathbf{X} = \mathbf{I}$ , 则称  $\mathbf{X}$  为  $\mathbf{A}$  的右逆, 记为  $\mathbf{A}_R^{-1}$ .

一般情况下,  $\mathbf{A}_L^{-1} \neq \mathbf{A}_R^{-1}$ . 若  $\mathbf{A}_L^{-1} = \mathbf{A}_R^{-1}$ , 则  $\mathbf{A}^{-1}$  存在, 且

$$\mathbf{A}^{-1} = \mathbf{A}_L^{-1} = \mathbf{A}_R^{-1}.$$

3.30

矩阵的左逆、右逆一般不唯一.

例如对任意复数  $a, b$ , 因

$$\begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

故矩阵

$$\begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \end{bmatrix}$$

都是矩阵

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

的左逆.

3.31

**Corollary 19.** (1) 设  $A \in \mathbb{C}_m^{m \times n}$ , 则必存在  $A$  的右逆, 且

$$A_R^{-1} = A^H(AA^H)^{-1}.$$

(2) 设  $A \in \mathbb{C}_n^{m \times n}$ , 则必存在  $A$  的左逆, 且

$$A_L^{-1} = (A^H A)^{-1} A^H.$$

**证:** (1) 考虑矩阵  $AA^H$ , 由定理 17, 有

$$\text{rank } AA^H = \text{rank } A = m,$$

又  $AA^H$  为  $m$  阶方阵, 故为可逆矩阵.

又


$$AA^H(AA^H)^{-1} = I,$$

故  $A_R^{-1} = A^H(AA^H)^{-1}$ .

同理有 (2) 成立. □

3.32

**Lemma 20.** 设  $F \in \mathbb{C}^{m \times n}$ ,  $G \in \mathbb{C}^{n \times m}$ , 则  $FG$  与  $GF$  有相同的非零特征值.

  $FG$  是  $m$  阶方阵,  $GF$  是  $n$  阶方阵, 特征值个数不同.

**证:** 设  $\lambda \neq 0$  为  $FG$  的特征值,  $x$  为对应的特征向量, 即有

$$FGx = \lambda x. \tag{5}$$

两边同时左乘以  $G$ , 得

$$GFGx = \lambda Gx,$$

记  $y = Gx$ , 则  $y \neq 0$  (否则会导致 (5) 的左侧为 0), 且

$$GFy = \lambda y,$$

故  $\lambda$  也是  $GF$  的特征值.

反过来, 类似可以证明  $GF$  的非零特征值也是  $FG$  的非零特征值. □

用途: 计算  $FG$  或  $GF$  的特征值时, 计算其中阶数较低的一个矩阵的特征值, 以方便求得另一个矩阵的特征值.

3.33

## 2 舒尔引理与正规矩阵的分解

### 2.1 舒尔引理

回顾: 等价、相似、合同、正交相似

- 设  $A, B$  均为  $m \times n$  矩阵,  $A$  与  $B$  等价  $\iff$  存在  $m$  阶可逆阵  $P$  和  $n$  阶可逆阵  $Q$ , 使  $PAQ = B$ .
- 设  $A, B$  均为  $n$  阶方阵,
  - $A$  与  $B$  相似  $\iff$  存在可逆阵  $P$ , 使  $P^{-1}AP = B$ .

- $A$  与  $B$  合同  $\iff$  存在可逆阵  $C$ , 使  $C^T A C = B$ .
- $A$  与  $B$  正交相似  $\iff$  存在正交阵  $Q$ , 使  $Q^T A Q = Q^{-1} A Q = B$ .

等价、相似、合同、正交相似的区别和联系:

- 等价的矩阵不必是方阵, 后面三个都是方阵之间的关系.
- 相似、合同、正交相似都是等价的一种; 正交相似关系最强, 等价关系最弱.
- 相似与合同没有什么关系, 仅当  $Q$  为正交阵时, 有  $Q^T A Q = Q^{-1} A Q$ , 这时相似与合同是一致的.

3.34


**Definition 21.** 设  $A, B$  均为  $n$  阶方阵, 如果存在一个  $n$  阶酉矩阵  $U$ , 使得

$$A = U B U^H,$$

则称  $A$  与  $B$  酉相似.

酉矩阵  $U$  满足  $U^H = U^{-1}$ , 上式也可以记为

$$A = U B U^{-1}, \text{ 或者 } A = U B U^H = U B U^{-1}.$$

 酉相似是正交相似的推广.

3.35

**Lemma 22** (Schur 引理). 设  $A \in \mathbb{C}^{n \times n}$ , 则存在酉矩阵  $U \in \mathbb{C}^{n \times n}$  及上三角矩阵  $T \in \mathbb{C}^{n \times n}$ , 使得

$$U^H A U = T,$$

即  $A$  酉相似于  $T$ , 且  $T$  的主对角元素均为  $A$  的特征值.

**证:** 矩阵  $A$  必相似于一个 Jordan 矩阵  $J$ , 即有可逆矩阵  $P$ , 使得

$$A = P J P^{-1}.$$

对可逆矩阵  $P$ , 存在一个  $n$  阶酉矩阵  $U$  及一个正线上三角矩阵  $R$ , 满足  $P = U R$ . 则有

$$A = U R J (U R)^{-1} = U R J R^{-1} U^{-1} = U R J R^{-1} U^H \triangleq U T U^H$$

其中  $T = R J R^{-1}$  为上三角矩阵, 且主对角线上元素均为  $A$  的特征值.  $\square$

3.36

**Definition 23.** 若  $A \in \mathbb{C}^{n \times n}$  满足条件

$$A^H A = A A^H,$$

则称  $A$  为正规矩阵.

下列 6 类方阵是常见的正规矩阵:

实矩阵的情形	复矩阵的情形
(1) 实对称矩阵 ( $A^T = A$ )	(4) Hermite 矩阵 ( $A^H = A$ )
(2) 实反对称矩阵 ( $A^T = -A$ )	(5) 反 Hermite 矩阵 ( $A^H = -A$ )
(3) 正交矩阵 ( $A^T = A^{-1}$ )	(6) 酉矩阵 ( $U^H = U^{-1}$ )

但是正规矩阵并非只包括上述几类, 例如下面的

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix},$$

是正规矩阵, 因为:

$$\mathbf{A}\mathbf{A}^{\mathrm{H}} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} = \mathbf{A}^{\mathrm{H}}\mathbf{A}.$$

但矩阵  $\mathbf{A}$  既不是酉矩阵, 也不是 Hermite 矩阵或反 Hermite 矩阵.

*Example 24.* 已知  $\mathbf{T}$  是一个上三角矩阵, 若  $\mathbf{T}$  还满足

$$\mathbf{T}^{\mathrm{H}}\mathbf{T} = \mathbf{T}\mathbf{T}^{\mathrm{H}}. \quad (6)$$

试证  $\mathbf{T}$  为对角矩阵.

**证:** 记

$$\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ & t_{22} & \cdots & t_{2n} \\ & & \ddots & \vdots \\ & & & t_{nn} \end{bmatrix}.$$

代入 (6) 得

$$\begin{bmatrix} \overline{t_{11}} & & & \\ \overline{t_{12}} & \overline{t_{22}} & & \\ \vdots & \vdots & \ddots & \\ \overline{t_{1n}} & \overline{t_{2n}} & \cdots & \overline{t_{nn}} \end{bmatrix} \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ & t_{22} & \cdots & t_{2n} \\ & & \ddots & \vdots \\ & & & t_{nn} \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ & t_{22} & \cdots & t_{2n} \\ & & \ddots & \vdots \\ & & & t_{nn} \end{bmatrix} \begin{bmatrix} \overline{t_{11}} & & & \\ \overline{t_{12}} & \overline{t_{22}} & & \\ \vdots & \vdots & \ddots & \\ \overline{t_{1n}} & \overline{t_{2n}} & \cdots & \overline{t_{nn}} \end{bmatrix},$$

两边都用第 1 行乘以第 1 列得

$$\|t_{11}\|^2 = \|t_{11}\|^2 + \|t_{12}\|^2 + \cdots + \|t_{1n}\|^2,$$

故  $t_{12} = \cdots = t_{1n} = 0$ .

两边都用第 2 行乘以第 2 列, 并注意到  $t_{12} = 0$ , 得

$$\|t_{22}\|^2 = \|t_{22}\|^2 + \|t_{23}\|^2 + \cdots + \|t_{2n}\|^2,$$

故  $t_{23} = \cdots = t_{2n} = 0$ . 如此下去, 即可推得  $\mathbf{T}$  是一个对角矩阵.  $\square$

**Theorem 25.** 设  $\mathbf{A} \in \mathbb{C}^{n \times n}$ , 则  $\mathbf{A}$  是正规矩阵的充分必要条件为: 存在一个  $n$  阶酉矩阵  $\mathbf{U}$  及一个  $n$  阶对角矩阵  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ , 使得

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{H}}.$$

这时称  $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{H}}$  为  $\mathbf{A}$  的酉相似对角化分解式, 简称为  $\mathbf{A}$  的酉相似对角分解.

证: 必要性. 若  $A$  是正规矩阵, 则

$$A^H A = A A^H. \quad (7)$$

由 Schur 引理, 存在酉矩阵  $U \in \mathbb{C}^{n \times n}$  及上三角矩阵  $T \in \mathbb{C}^{n \times n}$ , 使得

$$U^H A U = T, \quad \text{即} \quad A = U T U^H.$$

代入 (7) 式得  $(U T U^H)^H (U T U^H) = (U T U^H) (U T U^H)^H$ , 即

$$U T^H U^H U T U^H = U T U^H U T^H U^H.$$

注意到  $U^H U = I$ , 得  $U T^H T U^H = U T T^H U^H$ , 从而  $T^H T = T T^H$ .

又  $T$  为上三角矩阵, 由前例知  $T$  为对角矩阵. 记  $T = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ , 即有

$$A = U T U^H = U \Lambda U^H.$$

充分性. 若  $A$  酉相似于对角矩阵  $\Lambda$ , 即存在酉矩阵  $U$  使得

$$A = U \Lambda U^H,$$

则

$$A^H A = (U \Lambda^H U^H) (U \Lambda U^H) = U \Lambda^H \Lambda U^H,$$

$$A A^H = (U \Lambda U^H) (U \Lambda^H U^H) = U \Lambda \Lambda^H U^H.$$

由  $\Lambda$  为对角矩阵, 有  $\Lambda^H \Lambda = \Lambda \Lambda^H$ , 故

$$A^H A = A A^H,$$

得证  $A$  为正规矩阵. □

**Theorem 26** (Schur 不等式, 1909 年). 若  $n$  阶复矩阵  $A = [a_{ij}]$  的特征值为  $\lambda_1, \lambda_2, \dots, \lambda_n$ , 则有

$$\sum_{i=1}^n |\lambda_i|^2 \leq \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2,$$

等号成立, 当且仅当  $A$  是正规矩阵.

**Corollary 27.** 设  $A \in \mathbb{R}^{n \times n}$  是正规矩阵,  $r_1, r_2, \dots, r_k$  是  $A$  的实特征值,  $c_t \pm id_t$  ( $t = 1, 2, \dots, p$ , 且  $2p + k = n$ ) 是  $A$  的复特征值, 则存在正交矩阵  $Q \in \mathbb{R}^{n \times n}$ , 使得

$$Q^T A Q = \text{diag}(r_1, r_2, \dots, r_k) \oplus \begin{bmatrix} c_1 & -d_1 \\ d_1 & c_1 \end{bmatrix} \oplus \dots \oplus \begin{bmatrix} c_p & -d_p \\ d_p & c_p \end{bmatrix}.$$

**Corollary 28.** 若  $A \in \mathbb{R}^{n \times n}$  为实对称阵, 则  $A$  的特征值  $\lambda_1, \lambda_2, \dots, \lambda_n$  皆为实数, 且存在正交矩阵  $Q \in \mathbb{R}^{n \times n}$ , 使得

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

**Exercise 29** (P.124 习题 (二) 1). 设  $\mathbf{A} = \begin{bmatrix} -1 & i & 0 \\ -i & 0 & -i \\ 0 & i & -1 \end{bmatrix}$ , 问  $\mathbf{A}$  是否为正规矩阵? 若是, 求它的酉相似对角分解.

**解:** 因为  $\mathbf{A}^H = \mathbf{A}$ , 即  $\mathbf{A}$  是 Hermite 矩阵, 所以是正规矩阵.

$$\det(\lambda \mathbf{I} - \mathbf{A}) = \begin{vmatrix} \lambda + 1 & -i & 0 \\ i & \lambda & i \\ 0 & -i & \lambda + 1 \end{vmatrix} = (\lambda - 1)(\lambda + 1)(\lambda + 2),$$

得特征值  $\lambda_1 = 1, \lambda_2 = -1, \lambda_3 = -2$ .


求对应于  $\lambda_1 = 1$  的特征向量.

$$\begin{aligned} \begin{bmatrix} 2 & -i & 0 \\ i & 1 & i \\ 0 & -i & 2 \end{bmatrix} &\xrightarrow{r_2 \times (-i)} \begin{bmatrix} 2 & -i & 0 \\ 1 & -i & 1 \\ 0 & -i & 2 \end{bmatrix} \xrightarrow{r_1 - 2r_2} \begin{bmatrix} 0 & i & -2 \\ 1 & -i & 1 \\ 0 & -i & 2 \end{bmatrix} \\ &\xrightarrow[r_1 \leftrightarrow r_2]{r_3 + r_1} \begin{bmatrix} 1 & -i & 1 \\ 0 & i & -2 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow[r_2 \times (-i)]{r_1 + r_2} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2i \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

得对应于  $\lambda_1 = 1$  的特征向量  $\mathbf{p}_1 = (1, -2i, 1)^T$ . 类似地求得对应于  $\lambda_2 = -1, \lambda_3 = -2$  的特征向量分别为  $\mathbf{p}_2 = (-1, 0, 1)^T, \mathbf{p}_3 = (1, i, 1)^T$ . 将  $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$  标准化, 得标准正交向量组

$$\mathbf{u}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -2i \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ i \\ 1 \end{bmatrix}.$$

于是  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3], \mathbf{\Lambda} = \text{diag}(1, -1, -2)$ , 使得  $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ . □

 Hermite 矩阵的特征值为实数, 并且它的属于不同特征值的特征向量相互正交.

## 2.2 矩阵的奇异值分解

**Definition 30.** 设  $\mathbf{A} \in \mathbb{C}^{m \times n}$ , 则矩阵  $\mathbf{A}^H \mathbf{A}$  的  $n$  个特征值  $\lambda_i$  的算术平方根  $\delta_i = \sqrt{\lambda_i}$  叫做  $\mathbf{A}$  的奇异值.

**Theorem 31.** 设  $\mathbf{A} \in \mathbb{C}_r^{m \times n}$ , 则存在酉矩阵  $\mathbf{U} \in \mathbb{C}^{m \times m}, \mathbf{V} \in \mathbb{C}^{n \times n}$ , 使得

$$\mathbf{U}^H \mathbf{A} \mathbf{V} = \begin{bmatrix} \mathbf{S} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}. \quad (8)$$

其中  $\mathbf{S} = \text{diag}(\delta_1, \delta_2, \dots, \delta_r)$ , 且  $\delta_1 \geq \delta_2 \geq \dots \geq \delta_r > 0$ , 而  $\delta_i$  为  $\mathbf{A}$  的正奇异值.

☞ (8) 式可等价地表达为

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \mathbf{S} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^{\text{H}},$$

称为矩阵  $\mathbf{A}$  的奇异值分解(singular value decomposition, 或 SVD).

3.48

**证:** 因为  $\mathbf{A}^{\text{H}}\mathbf{A}$  是半正定的 Hermite 矩阵, 它的特征值大于等于零, 记为  $\delta_1^2, \delta_2^2, \dots, \delta_r^2, 0, \dots, 0$ , 并将它们的顺序排列成

$$\delta_1 \geq \delta_2 \geq \dots \geq \delta_r > \delta_{r+1} = \delta_{r+2} = \dots = \delta_n = 0.$$

令  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是对应于  $\delta_1^2, \delta_2^2, \dots, \delta_r^2, 0, \dots, 0$  的单位正交特征向量组, 记

$$\begin{aligned} \mathbf{V}_1 &= [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r], \\ \mathbf{V}_2 &= [\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \dots, \mathbf{v}_n], \\ \mathbf{V} &= [\mathbf{V}_1, \mathbf{V}_2]. \end{aligned}$$

则  $\mathbf{V}$  为  $n$  阶酉矩阵. 令  $\mathbf{S} = \text{diag}(\delta_1, \delta_2, \dots, \delta_r)$ , 则有

$$\mathbf{A}^{\text{H}}\mathbf{A}\mathbf{V}_1 = \mathbf{V}_1\mathbf{S}^2.$$

因为  $\mathbf{V}_1^{\text{H}}\mathbf{V}_1 = \mathbf{I}_r$ , 故

$$\mathbf{S}^{-1}\mathbf{V}_1^{\text{H}}\mathbf{A}^{\text{H}}\mathbf{A}\mathbf{V}_1\mathbf{S}^{-1} = \mathbf{I}. \quad (9)$$

令  $\mathbf{U}_1 = \mathbf{A}\mathbf{V}_1\mathbf{S}^{-1}$ , 则公式 (9) 就是  $\mathbf{U}_1^{\text{H}}\mathbf{U}_1 = \mathbf{I}$ .

3.49

注意到  $\mathbf{A}^{\text{H}}\mathbf{A}\mathbf{V}_2 = \mathbf{O}$ , 则  $\mathbf{V}_2^{\text{H}}\mathbf{A}^{\text{H}}\mathbf{A}\mathbf{V}_2 = \mathbf{O}$ , 因此  $\mathbf{A}\mathbf{V}_2 = \mathbf{O}$ .

把  $\mathbf{U}_1$  的列向量可以扩充成一组标准正交基, 记此标准正交基构成的酉矩阵为  $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$ . 则

$$\begin{aligned} \mathbf{U}^{\text{H}}\mathbf{A}\mathbf{V} &= \begin{bmatrix} \mathbf{U}_1^{\text{H}} \\ \mathbf{U}_2^{\text{H}} \end{bmatrix} \mathbf{A}[\mathbf{V}_1, \mathbf{V}_2] = \begin{bmatrix} \mathbf{U}_1^{\text{H}}\mathbf{A}\mathbf{V}_1 & \mathbf{U}_1^{\text{H}}\mathbf{A}\mathbf{V}_2 \\ \mathbf{U}_2^{\text{H}}\mathbf{A}\mathbf{V}_1 & \mathbf{U}_2^{\text{H}}\mathbf{A}\mathbf{V}_2 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{S} & \mathbf{O} \\ \mathbf{U}_2^{\text{H}}\mathbf{U}_1\mathbf{S} & \mathbf{O} \end{bmatrix} = \begin{bmatrix} \mathbf{S} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}. \end{aligned}$$

得证存在酉矩阵  $\mathbf{U}, \mathbf{V}$  使得

$$\mathbf{U}^{\text{H}}\mathbf{A}\mathbf{V} = \begin{bmatrix} \mathbf{S} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}, \quad \text{或} \quad \mathbf{A} = \mathbf{U} \begin{bmatrix} \mathbf{S} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^{\text{H}}. \quad \square$$

3.50

### 奇异值分解的计算

由  $\mathbf{A} = \mathbf{U} \begin{bmatrix} \mathbf{S} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^{\text{H}}$  可得

$$\mathbf{A}\mathbf{A}^{\text{H}} = \mathbf{U} \begin{bmatrix} \mathbf{S} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^{\text{H}}\mathbf{V} \begin{bmatrix} \mathbf{S} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{U}^{\text{H}} = \mathbf{U} \begin{bmatrix} \mathbf{S}^2 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{U}^{\text{H}},$$



即矩阵  $\mathbf{A}\mathbf{A}^H$  的特征值为  $\delta_1^2, \delta_2^2, \dots, \delta_r^2, 0, \dots, 0$ ,  $\mathbf{U}$  的各列是  $\mathbf{A}\mathbf{A}^H$  对应于上述特征值的标准正交特征向量, 称为  $\mathbf{A}$  的左奇异向量.

类似地可知 (或由定理证明过程可知),  $\mathbf{V}$  的各列是  $\mathbf{A}^H\mathbf{A}$  的对应于特征值  $\delta_1^2, \delta_2^2, \dots, \delta_r^2, 0, \dots, 0$  的一组标准正交特征向量, 称为  $\mathbf{A}$  的右奇异向量.

这就给我们提供了计算  $\mathbf{U}, \mathbf{V}$  的方法.

注意到  $\mathbf{A}\mathbf{A}^H$  与  $\mathbf{A}^H\mathbf{A}$  有相同的非零特征值, 实际计算特征值时, 建议选取阶数较小的矩阵.

因  $\mathbf{A} \in \mathbb{C}^{m \times n}$ , 则  $\mathbf{A}\mathbf{A}^H$  为  $m$  阶方阵, 零特征值有  $m-r$  个;  $\mathbf{A}^H\mathbf{A}$  为  $n$  阶方阵, 零特征值有  $n-r$  个.

注意下述两个公式的使用:

$$\mathbf{U}_1 = \mathbf{A}\mathbf{V}_1\mathbf{S}^{-1}, \quad (10)$$

$$\mathbf{V}_1 = \mathbf{A}^H\mathbf{U}_1\mathbf{S}^{-1}. \quad (11)$$

公式 (10) 是在定理证明过程中给出的. 由  $\mathbf{A}^H\mathbf{A}\mathbf{V}_1 = \mathbf{V}_1\mathbf{S}^2$ , 得

$$\mathbf{A}^H\mathbf{U}_1\mathbf{S}^{-1} = \mathbf{A}^H\mathbf{A}\mathbf{V}_1\mathbf{S}^{-1}\mathbf{S}^{-1} = \mathbf{V}_1,$$

故公式 (11) 成立.

有下列两种求  $\mathbf{V}, \mathbf{U}$  的步骤:

1. 求  $\mathbf{A}^H\mathbf{A}$  的特征值及对应的特征向量, 得到  $\mathbf{V}$ . 其中非零特征值对应的特征向量, 构成矩阵  $\mathbf{V}_1$ ; 由公式  $\mathbf{U}_1 = \mathbf{A}\mathbf{V}_1\mathbf{S}^{-1}$  得到  $\mathbf{A}\mathbf{A}^H$  的非零特征值所对应的特征向量, 其余的特征向量可以由 Hermite 矩阵的特征向量的正交性获得 (显然不唯一).
2. 求  $\mathbf{A}\mathbf{A}^H$  的特征值及对应的特征向量, 得到  $\mathbf{U}$ . 其中非零特征值对应的特征向量, 构成矩阵  $\mathbf{U}_1$ ; 由公式  $\mathbf{V}_1 = \mathbf{A}^H\mathbf{U}_1\mathbf{S}^{-1}$  得到  $\mathbf{A}^H\mathbf{A}$  的非零特征值所对应的特征向量, 其余的特征向量可以由 Hermite 矩阵的特征向量的正交性获得 (显然不唯一).

*Example 32.* 求矩阵  $\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}$  的奇异值分解.

**解:** 因为

$$\mathbf{A}\mathbf{A}^H = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix},$$

故  $\mathbf{A}\mathbf{A}^H$  的特征多项式为

$$|\lambda\mathbf{I} - \mathbf{A}\mathbf{A}^H| = \begin{vmatrix} \lambda - 2 & 1 \\ 1 & \lambda - 2 \end{vmatrix} = (\lambda - 3)(\lambda - 1),$$

因此  $\mathbf{A}\mathbf{A}^H$  的特征值为  $\lambda_1 = 3, \lambda_2 = 1$ . 它们对应的单位特征向量分别为

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

则  $\mathbf{A}^H \mathbf{A}$  的 3 个特征值分别为  $\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0$ . 由  $\mathbf{S} = \text{diag}(\delta_1, \delta_2) = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}) = \text{diag}(\sqrt{3}, 1)$ , 得

$$\mathbf{V}_1 = \mathbf{A}^H \mathbf{U}_1 \mathbf{S}^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{3}} & 1 \\ -\frac{1}{\sqrt{3}} & 1 \\ \frac{2}{\sqrt{3}} & 0 \end{bmatrix},$$

故  $\mathbf{A}^H \mathbf{A}$  的 2 个非零特征值  $\lambda_1 = 3, \lambda_2 = 1$  对应的特征向量分别为

$$\mathbf{v}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

因  $\lambda_3 = 0$  对应的特征向量与  $\mathbf{v}_2$  正交, 故可设其为  $(1, -1, z)^T$ . 又需要和  $\mathbf{v}_1$  正交, 故  $z = -1$ . 从而  $\lambda_3 = 0$  对应的单位特征向量为

$$\mathbf{v}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}.$$

令

$$\mathbf{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \end{bmatrix},$$

得矩阵  $\mathbf{A}$  的奇异值分解为

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \mathbf{S} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^H = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{bmatrix}. \quad \square$$

注意答案不唯一. 比如还可以取

$$\mathbf{v}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}.$$

### 3 幂等矩阵、投影算子及矩阵的谱分解式

#### 3.1 投影算子、幂等算子和幂等矩阵

##### 投影概述

空间物体的位置用向量  $(x, y, z)^T$  来表示, 该物体在阳光垂直照射下对应着一个影子, 也就是  $(x, y, 0)^T$ . 这样的变换就是一个投影变换, 它将三维空间中的向量  $(x, y, z)^T$  映射为向量  $(x, y, 0)^T$ .

这个变换可以用矩阵表示为

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

因为对任意一个向量  $(x, y, z)^T$ , 这个矩阵的作用是:

$$\mathbf{P} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}.$$

3.56

记  $z$  轴为子空间  $M$ , 记  $xOy$  面为子空间  $L$ , 则  $\mathbb{R}^3 = L \oplus M$ . 线性变换  $\mathbf{P}$  称为  $\mathbb{R}^3$  上沿  $M$  向  $L$  的投影算子. 它有如下的特点:

1.  $R(\mathbf{P}) = L$ . 因为  $\mathbf{P}$  将  $\mathbb{R}^3$  的任何向量都映射到  $xOy$  面, 所以  $xOy$  面 (即子空间  $L$ ) 是线性变换  $\mathbf{P}$  的值域.
2.  $N(\mathbf{P}) = M$ . 因为  $\mathbf{x}$  被映射为零向量  $\mathbf{0}$ , 当且仅当  $\mathbf{x}$  在  $z$  轴上. 即  $\mathbf{P}\mathbf{x} = \mathbf{0} \Leftrightarrow \mathbf{x} \in M$ .
3.  $\mathbf{P}$  在  $L$  上是恒等变换. 因为  $xOy$  面上的向量, 其投影当然只能是它自己.

正如阳光可以沿着不同的方向将影子投射到地面上, 一般的投影变换也可以是沿着不同的  $M$  到  $L$  上的投影.

我们很容易发现, 即使阳光沿不同的方向, 朝不同的平面做投影, 上述三个特点都是不会变化的.

3.57

另外,

$$\mathbf{P}^2 \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{P} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix},$$

所以  $\mathbf{P} = \mathbf{P}^2$ , 即  $\mathbf{P}$  是一个幂等矩阵, 或幂等算子.

直观上来说: 影子再投影, 只能是它自己.  $\mathbf{P}\mathbf{x}$  已经在空间  $L$  中, 而  $\mathbf{P}$  在  $L$  上是恒等变换, 所以

$$\mathbf{P}^2\mathbf{x} = \mathbf{P}(\mathbf{P}\mathbf{x}) = \mathbf{P}\mathbf{x}.$$

即  $\mathbf{P}^2 = \mathbf{P}$ .

后文我们将会看到:  $T$  是投影算子  $\Leftrightarrow T$  是幂等算子.

3.58

## 投影算子

**Definition 33.** 设  $L, M$  为空间  $\mathbb{C}^n$  的两个非空子空间,  $\mathbb{C}^n = L \oplus M$ ,  $T$  为  $\mathbb{C}^n$  上的线性变换, 如果  $T$  满足条件:

1.  $R(T) = L$ ;
2.  $N(T) = M$ ;
3.  $T$  限制在  $L$  上时为  $L$  上的恒等变换,

则称  $T$  为  $\mathbb{C}^n$  上沿  $M$  向  $L$  的投影算子, 记为  $P_{L,M}$ , 即  $P_{L,M} = T$ , 或者

$$T = P_{R(T), N(T)}.$$

前述的投影是定义在  $\mathbb{C}^n$  上的, 也可以定义一般线性空间  $V$  上的投影.

3.59

见教材 P.69 习题 10.

设  $V$  是数域  $F$  上的线性空间,  $L$  和  $M$  是  $V$  的两个子空间, 且

$$V = L \oplus M.$$

任取  $\mathbf{x} \in V$ , 设  $\mathbf{x} = \mathbf{y} + \mathbf{z}$ ,  $\mathbf{y} \in L$ ,  $\mathbf{z} \in M$ . 令

$$\begin{aligned} T: V &\longrightarrow V \\ \mathbf{x} &\longmapsto \mathbf{y}. \end{aligned}$$

则  $T$  是  $V$  上的一个线性变换. 称  $T$  是平行于  $M$  在  $L$  上的投影, 它满足

$$T\mathbf{x} = \begin{cases} \mathbf{x}, & \text{当 } \mathbf{x} \in L, \quad (\text{即 } T \text{ 在 } L \text{ 上是一个恒等变换}) \\ \mathbf{0}, & \text{当 } \mathbf{x} \in M. \end{cases}$$

对任意  $\mathbf{x} \in V$ , 有  $T\mathbf{x} = T^2\mathbf{x}$ .

3.60

该结论的证明可以参照教材后文的定理 3.3.10 (即本文定理 43).

对上述的  $V = L \oplus M$ , 及  $\forall \mathbf{x} \in V$ ,  $\mathbf{x} = \mathbf{y} + \mathbf{z}$ ,  $\mathbf{y} \in L$ ,  $\mathbf{z} \in M$ , 若定义

$$\begin{aligned} T^\dagger: V &\longrightarrow V \\ \mathbf{x} &\longmapsto \mathbf{z}. \end{aligned}$$

即  $T^\dagger\mathbf{x} = \mathbf{z}$ , 则  $T^\dagger$  是平行于  $L$  向  $M$  的投影. 相仿地,

$$T^\dagger\mathbf{x} = \begin{cases} \mathbf{x}, & \text{当 } \mathbf{x} \in M, \quad (\text{即 } T^\dagger \text{ 在 } M \text{ 上是一个恒等变换}) \\ \mathbf{0}, & \text{当 } \mathbf{x} \in L. \end{cases}$$

显然,  $T$  的值域是  $L$ , 核空间是  $M$ ; 而  $T^\dagger$  的值域是  $M$ , 核空间是  $L$ , 故

$$R(T) = N(T^\dagger), \quad R(T^\dagger) = N(T).$$

3.61

并且有下面两条重要性质:

1.  $TT^\dagger = T^\dagger T = 0^*$ . (即  $TT^\dagger, T^\dagger T$  都是零变换. 或者说  $T$  与  $T^\dagger$  是正交的.)
2.  $T + T^\dagger = E$ . (即  $T + T^\dagger$  是恒等变换.)

事实上,  $\forall \mathbf{x} \in V$ ,  $\mathbf{x} = \mathbf{y} + \mathbf{z}$ ,  $\mathbf{y} \in L$ ,  $\mathbf{z} \in M$ ,

$$TT^\dagger\mathbf{x} = T(T^\dagger\mathbf{x}) = T(\mathbf{z}) = \mathbf{0},$$

故  $TT^\dagger = 0^*$ . 同理  $T^\dagger T = 0^*$ .

又

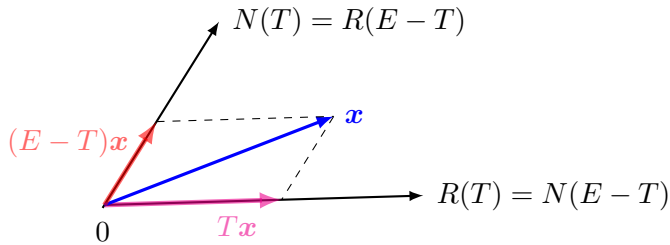
$$(T + T^\dagger)\mathbf{x} = T\mathbf{x} + T^\dagger\mathbf{x} = \mathbf{y} + \mathbf{z} = \mathbf{x},$$

故  $T + T^\dagger = E$ , 即  $T + T^\dagger$  是恒等变换  $E$ .

由  $R(T) = N(T^\dagger)$ ,  $R(T^\dagger) = N(T)$ , 以及  $T + T^\dagger = E$ , 有

$$R(T) = N(E - T), \quad N(T) = R(E - T).$$

即言, 任意一个空间可以是某个投影的值域, 也可能是某个投影的核空间.



教材上本节的各种结论都是围绕以上基本事实展开的.

Example 34. 考虑空间  $\mathbb{C}^4$ , 取其基底为

$$\mathcal{B} = \{(1, 0, 0, 0)^\top, (1, 1, 0, 0)^\top, (1, 1, 1, 0)^\top, (1, 1, 1, 1)^\top\},$$

并令

$$M = \text{span} [(1, 0, 0, 0)^\top, (1, 1, 0, 0)^\top] \triangleq \text{span}[\alpha_M, \beta_M],$$

$$L = \text{span} [(1, 1, 1, 0)^\top, (1, 1, 1, 1)^\top] \triangleq \text{span}[\alpha_L, \beta_L],$$

则有  $\mathbb{C}^4 = L \oplus M$ . 定义线性变换  $T$  为

$$T: \mathbb{C}^4 \rightarrow \mathbb{C}^4$$

$$T(x_1, x_2, x_3, x_4)^\top = (x_3, x_3, x_3, x_4)^\top,$$

则  $T$  就是一个沿  $M$  向  $L$  的投影算子.

证: (1) 验证  $R(T) = L$ . 任取向量  $\mathbf{x} = (x_1, x_2, x_3, x_4)^\top \in \mathbb{C}^4$ ,

$$\begin{aligned} T\mathbf{x} &= (x_3, x_3, x_3, x_4)^\top \\ &= (x_3 - x_4)(1, 1, 1, 0)^\top + x_4(1, 1, 1, 1)^\top \\ &= (x_3 - x_4)\alpha_L + x_4\beta_L. \end{aligned}$$

故  $T\mathbf{x} \in L$ , 因此  $R(T) \subseteq L$ .

反之, 任取向量  $\mathbf{y} \in L$ , 则存在  $k_1, k_2 \in \mathbb{C}$ , 使得

$$\mathbf{y} = k_1\alpha_L + k_2\beta_L = (k_1 + k_2, k_1 + k_2, k_1 + k_2, k_2)^\top.$$

取  $\mathbf{z} = (z_1, z_2, k_1 + k_2, k_2)^\top$ , 则  $T\mathbf{z} = \mathbf{y}$ , 故  $\mathbf{y} \in R(T)$ , 即  $L \subseteq R(T)$ .

综上有  $R(T) = L$ .

(2) 验证  $N(T) = M$ . 任取  $\mathbf{x} = (x_1, x_2, x_3, x_4)^\top \in N(T)$ , 则

$$T\mathbf{x} = (x_3, x_3, x_3, x_4)^\top = \mathbf{0}.$$

故  $x_3 = x_4 = 0$ . 即  $\boldsymbol{x}$  必具形式

$$\boldsymbol{x} = (x_1, x_2, 0, 0)^T = (x_1 - x_2)(1, 0, 0, 0)^T + x_2(1, 1, 0, 0)^T = (x_1 - x_2)\boldsymbol{\alpha}_M + x_2\boldsymbol{\beta}_M.$$

故  $\boldsymbol{x} \in M$ , 从而  $N(T) \subseteq M$ .

反之, 任取  $\boldsymbol{z} \in M$ , 则存在  $c_1, c_2 \in \mathbb{C}$ , 使得

$$\boldsymbol{z} = c_1\boldsymbol{\alpha}_M + c_2\boldsymbol{\beta}_M = (c_1 + c_2, c_2, 0, 0)^T.$$

故  $T\boldsymbol{z} = (0, 0, 0, 0)^T$ , 即  $\boldsymbol{z} \in N(T)$ , 从而  $M \subseteq N(T)$ .

综上有  $N(T) = M$ .

(3) 验证  $T$  限制在  $L$  上为  $L$  上的恒等变换. 任取  $L$  中的向量

$$\boldsymbol{w} = a_1\boldsymbol{\alpha}_L + a_2\boldsymbol{\beta}_L = (a_1 + a_2, a_1 + a_2, a_1 + a_2, a_2)^T,$$

由  $T$  的定义得

$$T\boldsymbol{w} = \boldsymbol{w}.$$

故  $T$  限制在  $L$  上为  $L$  上的恒等变换.

综上所述可知,  $T$  就是一个沿  $M$  向  $L$  的投影算子. □

可以验证, 若改  $T$  的定义为


$$T(x_1, x_2, x_3, x_4)^T = (x_3 + x_4, x_3 + x_4, x_3 + x_4, x_4)^T,$$

则  $T$  不是一个沿  $M$  向  $L$  的投影算子. □

### 幂等算子, 幂等矩阵

**Definition 35.** 设  $T$  为  $\mathbb{C}^n$  上的线性变换, 且满足条件  $T^2 = T$ , 则称  $T$  为幂等算子.

**Definition 36.** 若  $n$  阶方阵  $P$  满足  $P^2 = P$ , 则称  $P$  为幂等矩阵.

 幂等算子的矩阵表示必为幂等矩阵.

**证:** 设  $T$  为  $\mathbb{C}^n$  上的任一幂等算子,  $\mathcal{B}$  为  $\mathbb{C}^n$  的任一基底,  $T$  关于  $\mathcal{B}$  的矩阵表示为  $P$ , 则有  $T\mathcal{B} = \mathcal{B}P$ . 由

$$\mathcal{B}P = T\mathcal{B} = T^2\mathcal{B} = TT\mathcal{B} = T\mathcal{B}P = \mathcal{B}P^2,$$

而  $\mathcal{B}$  为  $\mathbb{C}^n$  的基底, 从而为可逆矩阵, 故  $P = P^2$ ,  $P$  为幂等矩阵. □

**Lemma 37.**  $\mathbb{C}^n$  上的投影算子必为幂等算子.

**证:** 设  $\mathbb{C}^n = L \oplus M$ ,  $T$  是一个沿  $M$  向  $L$  的投影算子, 则  $\mathbb{C}^n = R(T) \oplus N(T)$ , 且  $T$  限制在  $R(T)$  上为  $R(T)$  上的恒等变换.

任取  $\boldsymbol{x} \in \mathbb{C}^n$ , 则由直和的定义,  $\exists \boldsymbol{y} \in R(T), \exists \boldsymbol{z} \in N(T)$ , 使得  $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}$ . 因此

$$T\boldsymbol{x} = T(\boldsymbol{y} + \boldsymbol{z}) = T\boldsymbol{y} + T\boldsymbol{z} = \boldsymbol{y} + \mathbf{0} = \boldsymbol{y},$$

$$T^2\boldsymbol{x} = T\boldsymbol{y} = \boldsymbol{y} = T\boldsymbol{x}.$$

故  $T^2 = T$ , 即  $T$  为  $\mathbb{C}^n$  上的幂等算子. □

**Lemma 38.** 投影算子的任意矩阵表示必为幂等矩阵.

**证:** 由引理 37, 投影算子是幂等算子, 而幂等算子的矩阵表示必为幂等矩阵, 因此结论成立.  $\square$

3.70

一个幂等矩阵  $P \in \mathbb{C}^{n \times n}$  也可以视为  $\mathbb{C}^n$  上的幂等算子.

**Lemma 39.** 若  $P \in \mathbb{C}^{n \times n}$  为幂等矩阵, 则下面的命题成立:

1.  $P^H$  和  $I - P$  也为幂等矩阵, 这里  $I$  为  $n$  阶单位矩阵;
2.  $P$  的特征值非 0 即 1, 且  $P$  可以对角化;
3.  $\text{rank } P = \text{tr } P$ ;
4.  $P(I - P) = (I - P)P = O$ ;
5.  $Px = x \Leftrightarrow x \in R(P)$ ;
6.  $N(P) = R(I - P)$ .

**证:** (1) 因  $P^H P^H = (PP)^H = P^H$ , 故  $P^H$  为幂等矩阵.

因  $(I - P)(I - P) = I - 2P + P^2 = I - 2P + P = I - P$ , 故  $I - P$  为幂等矩阵.

(4) 因  $P(I - P) = P - P^2 = O$ ,  $(I - P)P = P - P^2 = O$ , 故  $P(I - P) = (I - P)P$ .

3.71

下面先证 (6), 再证 (2), (5), 最后证 (3).

(6) 要证  $N(P) = R(I - P)$ . 任取  $x \in N(P)$ , 则有  $Px = O$ , 从而

$$(I - P)x = Ix - Px = x - O = x,$$

所以  $x = (I - P)x \in R(I - P)$ , 因而  $N(P) \subseteq R(I - P)$ .

反之, 任取  $y \in R(I - P)$ , 则存在  $z \in \mathbb{C}^n$ , 使得

$$y = (I - P)z = (I - P)^2 z = (I - P)y,$$

所以

$$Py = O,$$

故  $y \in N(P)$ , 因而  $R(I - P) \subseteq N(P)$ . 得证  $N(P) = R(I - P)$ .

3.72

(2) 证明  $P$  的特征值非 0 即 1, 且  $P$  可以对角化.

由  $P^2 - P = O$ , 知  $\lambda^2 - \lambda$  是  $P$  的一个零化多项式. 而  $P$  的最小多项式  $m_P(\lambda)$  满足

$$m_P(\lambda) \mid \lambda^2 - \lambda,$$

又  $\lambda^2 - \lambda$  的根为 0 和 1, 故  $m_P(\lambda)$  的根非 0 即 1. 从而  $P$  的特征值也非 0 即 1.

又因为  $\lambda^2 - \lambda = \lambda(\lambda - 1)$ , 即  $\lambda^2 - \lambda$  可以分解为不同一次因式的乘积, 故  $m_P(\lambda)$  也可分解为不同一次因式的乘积, 所以  $P$  可对角化.

3.73


(5) 证明  $Px = x \Leftrightarrow x \in R(P)$ .

任取  $x \in R(P)$ , 则存在  $z \in \mathbb{C}^n$  使得  $Pz = x$ . 于是

$$x = Pz = P^2 z = PPz = Px.$$

反之, 若  $P\mathbf{x} = \mathbf{x}$ , 则  $\mathbf{x} \in R(P)$ .

故  $P\mathbf{x} = \mathbf{x} \Leftrightarrow \mathbf{x} \in R(P)$ .

  $P$  是  $R(P)$  上的恒等变换.

(3) 证明  $\text{rank } P = \text{tr } P$ .

设  $\text{rank } P = r$ , 因  $P$  的特征值非 0 即 1, 则  $P$  必以 1 为其  $r$  重特征值. 于是

$$\text{tr } P = \sum_{i=1}^n \lambda_i = r = \text{rank } P.$$

证毕.

□

3.74

**Lemma 40.**  $\mathbb{C}^n$  上的幂等算子必为投影算子.

**证:** 设  $T$  为  $\mathbb{C}^n$  上的幂等算子, 由“幂等算子的矩阵表示一定是幂等矩阵”, 知  $T$  的任一矩阵表示  $P$  为幂等矩阵. 则  $P$  的特征值非 0 即 1, 且  $P$  可对角化. 故  $T$  的特征值也是非 0 即 1, 且  $T$  可对角化.

记  $T$  的关于特征值 0 和 1 的特征子空间分别为  $E_T(0)$  和  $E_T(1)$ , 则由可对角化线性变换的性质, 有

$$\mathbb{C}^n = E_T(0) \oplus E_T(1). \quad (12)$$

另一方面, 由  $\mathbf{x} \in E_T(1) \Leftrightarrow T\mathbf{x} = \mathbf{x} \Leftrightarrow \mathbf{x} \in R(T)$ , 故  $E_T(1) = R(T)$ . 同理  $E_T(0) = N(T)$ , 因此 (12) 式可改写为

$$\mathbb{C}^n = R(T) \oplus N(T).$$

下证  $T$  限制在  $R(T)$  上为  $R(T)$  上的恒等变换. 任取  $\mathbf{y} \in R(T)$ , 则存在  $\mathbf{x} \in \mathbb{C}^n$ , 使得

$$T\mathbf{x} = \mathbf{y}.$$

从而

$$T\mathbf{y} = T^2\mathbf{x} = T\mathbf{x} = \mathbf{y},$$

故  $T$  为空间  $R(T)$  上的恒等变换.

综上,  $T$  是  $\mathbb{C}^n$  上沿  $N(T)$  向  $R(T)$  的投影算子.

□

上述引理可以换一个更为明确的表述:

如果  $T$  是  $\mathbb{C}^n$  上的幂等算子, 那么

$$\mathbb{C}^n = R(T) \oplus N(T),$$

且  $T$  是  $\mathbb{C}^n$  上沿  $N(T)$  向  $R(T)$  的投影算子.

下面给出  $\mathbb{C}^n = R(T) \oplus N(T)$  的另一个证明.

任取  $\mathbf{x} \in \mathbb{C}^n$ , 则  $T\mathbf{x} \in R(T)$ . 由于

$$T(\mathbf{x} - T\mathbf{x}) = T\mathbf{x} - T^2\mathbf{x} = T\mathbf{x} - T\mathbf{x} = \mathbf{0},$$

3.75

3.76

3.77



因此  $\mathbf{x} - T\mathbf{x} \in N(T)$ . 故任意  $\mathbf{x} \in \mathbb{C}^n$ , 有

$$\mathbf{x} = T\mathbf{x} + (\mathbf{x} - T\mathbf{x}), \quad T\mathbf{x} \in R(T), \mathbf{x} - T\mathbf{x} \in N(T),$$

所以

$$\mathbb{C}^n = R(T) + N(T). \quad (13)$$

任取  $\boldsymbol{\xi} \in R(T) \cap N(T)$ . 由于  $\boldsymbol{\xi} \in R(T)$ , 故存在  $\boldsymbol{\eta} \in \mathbb{C}^n$ , 使得  $T\boldsymbol{\eta} = \boldsymbol{\xi}$ . 由于  $\boldsymbol{\xi} \in N(T)$ , 故  $T\boldsymbol{\xi} = \mathbf{0}$ . 从而

$$\mathbf{0} = T\boldsymbol{\xi} = T(T\boldsymbol{\eta}) = T^2\boldsymbol{\eta} = T\boldsymbol{\eta} = \boldsymbol{\xi}.$$

于是

$$R(T) \cap N(T) = \{\mathbf{0}\}. \quad (14)$$

综合 (13), (14) 知  $\mathbb{C}^n = R(T) \oplus N(T)$ .

**Theorem 41.**  $T$  是  $\mathbb{C}^n$  上的投影算子  $\Leftrightarrow T$  是  $\mathbb{C}^n$  上的幂等算子.

**Corollary 42.** 对任一幂等矩阵  $\mathbf{P} \in \mathbb{C}^{n \times n}$ , 都有

$$\mathbb{C}^n = R(\mathbf{P}) \oplus N(\mathbf{P}),$$

且  $\mathbf{P}$  是  $\mathbb{C}^n$  上沿  $N(\mathbf{P})$  向  $R(\mathbf{P})$  的投影算子. 反之亦然.

**Theorem 43.** 设  $\mathbb{C}^n = L \oplus M$ , 则一定存在一个  $\mathbb{C}^n$  上的沿  $M$  向  $L$  的投影算子  $T$ .

**证:** 任取  $\mathbf{x} \in \mathbb{C}^n$ , 因  $\mathbb{C}^n = L \oplus M$ , 故  $\exists! \mathbf{y} \in L, \exists! \mathbf{z} \in M$ , 使得  $\mathbf{x} = \mathbf{y} + \mathbf{z}$ .

定义  $\mathbb{C}^n$  上的映射  $T$  为

$$T\mathbf{x} = \mathbf{y}. \quad (15)$$

则  $T$  为  $\mathbb{C}^n$  上的线性变换. 以下验证  $T$  为幂等算子, 从而为投影算子.

对上述的  $\mathbf{y}$ , 作为  $\mathbb{C}^n$  中的向量有唯一的分解式

$$\mathbf{y} = \mathbf{y} + \mathbf{0}, \quad \mathbf{y} \in L, \mathbf{0} \in M.$$

由  $T$  的定义知  $T\mathbf{y} = \mathbf{y}$ . 故

$$T^2\mathbf{x} = TT\mathbf{x} = T\mathbf{y} = \mathbf{y} = T\mathbf{x},$$

由  $\mathbf{x}$  的任意性, 得  $T$  为幂等算子.

下证  $R(T) = L, M = N(T)$ , 从而  $T$  是  $\mathbb{C}^n$  上的沿  $M$  向  $L$  的投影算子.

(1) 对任意  $\mathbf{x} \in \mathbb{C}^n$ , 有  $T\mathbf{x} \in R(T)$ , 由  $T\mathbf{x} = \mathbf{y} \in L$ , 知  $R(T) \subseteq L$ .

任意  $\mathbf{y} \in L, \mathbf{y} = T\mathbf{y} \in R(T)$ , 故  $L \subseteq R(T)$ .

(2) 任意  $\mathbf{z} \in M$ , 作为  $\mathbb{C}^n$  中的向量有唯一的分解式  $\mathbf{z} = \mathbf{0} + \mathbf{z}$ , 故  $T\mathbf{z} = \mathbf{0}$ , 从而  $M \subseteq N(T)$ .

任意  $\mathbf{x} \in N(T)$ , 有  $T\mathbf{x} = \mathbf{0}$ . 设  $\mathbf{x} = \mathbf{y} + \mathbf{z}, \mathbf{y} \in L, \mathbf{z} \in M$ , 则  $T\mathbf{x} = \mathbf{y} = \mathbf{0}$ , 从而  $\mathbf{x} = \mathbf{0} + \mathbf{z} = \mathbf{z} \in M$ . 故  $N(T) \subseteq M$ .  $\square$

### 3.2 可对角化矩阵的谱分解

#### 谱分解概览

矩阵  $\mathbf{A}$  的特征值的集合, 有时也称为谱(Spectrum).

谱分解 (Spectral decomposition), 也称为特征分解(Eigendecomposition),

注意: 只有可对角化矩阵才可以进行谱分解.

设  $\mathbf{A}$  可对角化, 则

$$\begin{aligned}
\mathbf{A} &= \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \\
&= [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{y}_1^T \\ \mathbf{y}_2^T \\ \vdots \\ \mathbf{y}_n^T \end{bmatrix} \\
&= \lambda_1 \mathbf{x}_1 \mathbf{y}_1^T + \lambda_2 \mathbf{x}_2 \mathbf{y}_2^T + \dots + \lambda_n \mathbf{x}_n \mathbf{y}_n^T \\
&= \sum_{i=1}^n \lambda_i \mathbf{P}_i,
\end{aligned}$$

其中  $\mathbf{P}_i = \mathbf{x}_i \mathbf{y}_i^T$ ,  $\mathbf{y}_i^T$  是矩阵  $\mathbf{P}^{-1}$  的行向量, 从而  $\mathbf{A}$  表达为  $n$  个矩阵  $\mathbf{P}_i$  之和的形式, 其组合系数是  $\mathbf{A}$  的谱.

此所谓谱分解.

因

$$\begin{aligned}
\mathbf{I} &= \mathbf{P}^{-1}\mathbf{P} \\
&= \begin{bmatrix} \mathbf{y}_1^T \\ \mathbf{y}_2^T \\ \vdots \\ \mathbf{y}_n^T \end{bmatrix} [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \\
&= \begin{bmatrix} \mathbf{y}_1^T \mathbf{x}_1 & \mathbf{y}_1^T \mathbf{x}_2 & \dots & \mathbf{y}_1^T \mathbf{x}_n \\ \mathbf{y}_2^T \mathbf{x}_1 & \mathbf{y}_2^T \mathbf{x}_2 & \dots & \mathbf{y}_2^T \mathbf{x}_n \\ \vdots & \vdots & & \vdots \\ \mathbf{y}_n^T \mathbf{x}_1 & \mathbf{y}_n^T \mathbf{x}_2 & \dots & \mathbf{y}_n^T \mathbf{x}_n \end{bmatrix},
\end{aligned}$$

故

$$\mathbf{y}_i^T \mathbf{x}_j = \begin{cases} 1, & \text{若 } i = j, \\ 0, & \text{若 } i \neq j. \end{cases}$$

或  $\mathbf{y}_i^T \mathbf{x}_j = \delta_{ij}$ .

对于  $\mathbf{P}_i$  有下面的性质.

(1)  $\mathbf{P}_i \mathbf{P}_j = \mathbf{O}$ ,  $i \neq j$ .  $\mathbf{P}_i^2 = \mathbf{P}_i$ .

事实上,

$$\begin{aligned}
\mathbf{P}_i \mathbf{P}_j &= (\mathbf{x}_i \mathbf{y}_i^T) (\mathbf{x}_j \mathbf{y}_j^T) \\
&= \mathbf{x}_i (\mathbf{y}_i^T \mathbf{x}_j) \mathbf{y}_j^T
\end{aligned}$$

$$\begin{aligned}
&= \mathbf{x}_i \delta_{ij} \mathbf{y}_j^T \\
&= \begin{cases} \mathbf{x}_i \mathbf{y}_i^T, & \text{若 } i = j, \\ \mathbf{O}, & \text{若 } i \neq j. \end{cases} \\
&= \begin{cases} \mathbf{P}_i, & \text{若 } i = j, \\ \mathbf{O}, & \text{若 } i \neq j. \end{cases}
\end{aligned}$$

故  $\mathbf{P}_i^2 = \mathbf{P}_i$ , 即  $\mathbf{P}_i$  为投影算子, 且两两正交.

$$(2) \sum_{i=1}^k \mathbf{P}_i = \mathbf{I}.$$

事实上,

$$\begin{aligned}
\mathbf{I} &= \mathbf{P}\mathbf{P}^{-1} \\
&= [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \begin{bmatrix} \mathbf{y}_1^T \\ \mathbf{y}_2^T \\ \vdots \\ \mathbf{y}_n^T \end{bmatrix} \\
&= \mathbf{x}_1 \mathbf{y}_1^T + \mathbf{x}_2 \mathbf{y}_2^T + \dots + \mathbf{x}_n \mathbf{y}_n^T \\
&= \sum_{i=1}^n \mathbf{P}_i.
\end{aligned}$$

3.86

3.87

Example 44. 求矩阵  $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix}$  的谱分解.

解: 由

$$|\lambda \mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda - 1 & -1 \\ -4 & \lambda - 1 \end{vmatrix} = (\lambda - 3)(\lambda + 1),$$

得特征值  $\lambda_1 = 3, \lambda_2 = -1$ . 对应的特征向量为

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

记  $\mathbf{P} = [\mathbf{x}_1, \mathbf{x}_2]$ , 则

$$\mathbf{P}^{-1} = \begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & -\frac{1}{4} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1^T \\ \mathbf{y}_2^T \end{bmatrix}.$$

从而

$$\mathbf{P}_1 = \mathbf{x}_1 \mathbf{y}_1^T = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ 1 & \frac{1}{2} \end{bmatrix}, \quad \mathbf{P}_2 = \mathbf{x}_2 \mathbf{y}_2^T = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} \\ -1 & \frac{1}{2} \end{bmatrix}.$$

故  $\mathbf{A}$  的谱分解为

$$\mathbf{A} = \lambda_1 \mathbf{P}_1 + \lambda_2 \mathbf{P}_2 = 3 \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ 1 & \frac{1}{2} \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} \\ -1 & \frac{1}{2} \end{bmatrix}. \quad \square$$

3.88

3.89

谱分解, 谱算子

**Theorem 45.** 设  $A \in \mathbb{C}^{n \times n}$  共有  $k$  个互异的特征值  $\lambda_1, \lambda_2, \dots, \lambda_k$ , 则  $A$  可对角化的充分必要条件是: 存在  $k$  个投影算子  $P_1, P_2, \dots, P_k$ , 满足

1.  $P_i P_j = O, i \neq j, P_i \in \mathbb{C}^{n \times n}, i = 1, 2, \dots, k;$
2.  $\sum_{i=1}^k P_i = I;$
3.  $A = \sum_{i=1}^k \lambda_i P_i.$

其中  $A = \sum_{i=1}^k \lambda_i P_i$  称为矩阵  $A$  的谱分解式, 而  $P_i$  称为  $A$  的谱算子.

3.90

**证:** 充分性. 设  $r_i = \text{rank } P_i$ , 令  $X_i \in \mathbb{C}^{n \times r_i}$  是  $R(P_i)$  的某一基底构成的列满秩矩阵,  $i = 1, 2, \dots, k$ , 并记

$$X \triangleq [X_1, X_2, \dots, X_k].$$

则

$$\begin{aligned} X \text{ 的列数} &= \sum_{i=1}^k r_i = \sum_{i=1}^k \text{rank } P_i && (r_i = \text{rank } P_i) \\ &= \sum_{i=1}^k \text{tr } P_i && (P_i \text{ 为幂等矩阵, } \text{rank } P_i = \text{tr } P_i) \\ &= \text{tr } \sum_{i=1}^k P_i && (\text{对角线元素之和不变}) \\ &= \text{tr } I && (\sum_{i=1}^k P_i = I) \\ &= n, \end{aligned}$$

3.91

这表明  $X \in \mathbb{C}^{n \times n}$ . 因为  $X_i$  是  $R(P_i)$  的基底, 故  $P_i$  的列向量可以由  $X_i$  的列向量线性表示. 即存在矩阵  $Y_i \in \mathbb{C}^{r_i \times n}$ , 使得

$$P_i = X_i Y_i, \quad i = 1, 2, \dots, k.$$

记

$$Y \triangleq \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_k \end{bmatrix},$$

则有

$$XY = [X_1, X_2, \dots, X_k] \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_k \end{bmatrix} = \sum_{i=1}^k X_i Y_i = \sum_{i=1}^k P_i = I.$$

可见  $\mathbf{X}$  可逆, 且  $\mathbf{X}^{-1} = \mathbf{Y}$ .

于是得

$$\begin{aligned} \mathbf{A} &= \sum_{i=1}^k \lambda_i \mathbf{P}_i && \text{(已知条件)} \\ &= \sum_{i=1}^k \lambda_i \mathbf{X}_i \mathbf{Y}_i && (\mathbf{P}_i = \mathbf{X}_i \mathbf{Y}_i) \\ &= [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k] \begin{bmatrix} \lambda_1 \mathbf{I}_{r_1} & & & \\ & \lambda_2 \mathbf{I}_{r_2} & & \\ & & \ddots & \\ & & & \lambda_k \mathbf{I}_{r_k} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_k \end{bmatrix} \\ &\triangleq \mathbf{X} \mathbf{\Lambda} \mathbf{Y} && \text{(记 } \mathbf{\Lambda} = \text{diag}(\lambda_1 \mathbf{I}_{r_1}, \lambda_2 \mathbf{I}_{r_2}, \dots, \lambda_k \mathbf{I}_{r_k})) \\ &= \mathbf{X} \mathbf{\Lambda} \mathbf{X}^{-1}, && (\mathbf{X}^{-1} = \mathbf{Y}) \end{aligned}$$

故  $\mathbf{A}$  可对角化.

必要性. 如果  $\mathbf{A} \in \mathbb{C}^{n \times n}$  可对角化, 设  $\mathbf{A}$  有  $k$  个互异的特征值  $\lambda_1, \lambda_2, \dots, \lambda_k$ , 其几何重数分别为  $r_1, r_2, \dots, r_k$ , 且  $\sum_{i=1}^k r_i = n$ .

将  $\lambda_i$  的  $r_i$  个线性无关特征向量, 构成矩阵  $\mathbf{X}_i$ , 则  $\mathbf{X}_i$  是列满秩的, 且  $\mathbf{X}_i \in \mathbb{C}_{r_i}^{n \times r_i}$ .

记  $\mathbf{X} \triangleq [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k]$ ,  $\mathbf{\Lambda} = \text{diag}(\lambda_1 \mathbf{I}_{r_1}, \lambda_2 \mathbf{I}_{r_2}, \dots, \lambda_k \mathbf{I}_{r_k})$ , 则

$$\mathbf{A} = \mathbf{X} \mathbf{\Lambda} \mathbf{X}^{-1}.$$

对  $\mathbf{X}^{-1}$  进行行分块,

$$\mathbf{X}^{-1} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_k \end{bmatrix}, \quad \mathbf{Y}_i \in \mathbb{C}_{r_i}^{r_i \times n}.$$

令  $\mathbf{P}_i = \mathbf{X}_i \mathbf{Y}_i \in \mathbb{C}^{n \times n}$ , 则有

$$\sum_{i=1}^k \mathbf{P}_i = \sum_{i=1}^k \mathbf{X}_i \mathbf{Y}_i = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k] \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_k \end{bmatrix} = \mathbf{X} \mathbf{X}^{-1} = \mathbf{I}.$$

即满足条件 (2). 又

$$\mathbf{I} = \mathbf{X}^{-1} \mathbf{X} = \begin{bmatrix} \mathbf{Y}_1 \mathbf{X}_1 & \mathbf{Y}_1 \mathbf{X}_2 & \cdots & \mathbf{Y}_1 \mathbf{X}_k \\ \mathbf{Y}_2 \mathbf{X}_1 & \mathbf{Y}_2 \mathbf{X}_2 & \cdots & \mathbf{Y}_2 \mathbf{X}_k \\ \vdots & \vdots & & \vdots \\ \mathbf{Y}_k \mathbf{X}_1 & \mathbf{Y}_k \mathbf{X}_2 & \cdots & \mathbf{Y}_k \mathbf{X}_k \end{bmatrix},$$

故

$$Y_i X_j = \begin{cases} I_{r_i}, & \text{若 } i = j, \\ O, & \text{若 } i \neq j. \end{cases}$$

从而

$$P_i^2 = P_i P_i = X_i Y_i X_i Y_i = X_i I_{r_i} Y_i = X_i Y_i = P_i.$$

故  $P_i$  为投影算子.

且满足: 若  $i \neq j$ , 则

$$P_i P_j = X_i Y_i X_j Y_j = X_i O Y_j = O.$$

以及

$$\begin{aligned} A &= X \Lambda X^{-1} = [X_1, X_2, \dots, X_k] \begin{bmatrix} \lambda_1 I_{r_1} & & & \\ & \lambda_2 I_{r_2} & & \\ & & \ddots & \\ & & & \lambda_k I_{r_k} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{bmatrix} \\ &= \sum_{i=1}^k \lambda_i X_i Y_i = \sum_{i=1}^k \lambda_i P_i. \end{aligned}$$

证毕. □

**Corollary 46.** 设可对角化矩阵  $A \in \mathbb{C}^{n \times n}$  共有  $k$  个互异的特征值  $\lambda_1, \lambda_2, \dots, \lambda_k$ , 则

1.  $A$  的谱分解式  $A = \sum_{i=1}^k \lambda_i P_i$  中的幂等矩阵  $P_i$  是由  $A$  唯一确定的;
2. 若  $f(\lambda) = \sum_{l=0}^m a_l \lambda^l$  为任一多项式, 则有

$$f(A) = \sum_{i=1}^k f(\lambda_i) P_i.$$

3. 设  $\varphi_i(\lambda) = \prod_{\substack{j=1 \\ j \neq i}}^k (\lambda - \lambda_j)$ , 则有

$$P_i = \frac{\varphi_i(A)}{\varphi_i(\lambda_i)}, \quad i = 1, 2, \dots, k.$$

4. 矩阵  $B$  与  $A$  可交换  $\Leftrightarrow B$  与  $P_i$  可交换.

**证:** (1) 设  $Q_1, Q_2, \dots, Q_k$  也满足  $Q_i^2 = Q_i$ ;  $Q_i Q_j = O$ ,  $i \neq j$ ;  $\sum_{i=1}^k Q_i = I$ ;

$A = \sum_{i=1}^k \lambda_i Q_i$ . 下证  $Q_i = P_i$ .

$$P_i A = P_i \sum_{l=1}^k \lambda_l P_l = \sum_{l=1}^k \lambda_l P_i P_l = \lambda_i P_i. \quad (16)$$

同理可得  $AP_i = \lambda_i P_i$ , 以及  $AQ_i = Q_i A = \lambda_i Q_i$ , 从而

$$P_i(AQ_j) = P_i(\lambda_j Q_j) = \lambda_j P_i Q_j, \quad (P_i A)Q_j = \lambda_i P_i Q_j,$$

因此

$$\lambda_j P_i Q_j = \lambda_i P_i Q_j,$$

即

$$(\lambda_j - \lambda_i) P_i Q_j = O,$$

因此当  $i \neq j$  时, 有  $P_i Q_j = O$ .

从而有

$$P_i = P_i \left( \sum_{l=1}^k Q_l \right) = P_i Q_i = \left( \sum_{l=1}^k P_l \right) Q_i = Q_i.$$

从而  $P_i$  是由  $A$  唯一确定的.

(2) 由式 (16) 可知, 对任意的正整数  $n$ , 有

$$P_i A^n = (P_i A) A^{n-1} = \lambda_i P_i A^{n-1} = \cdots = \lambda_i^n P_i,$$

从而

$$A^n = A A^{n-1} = \left( \sum_{i=1}^k \lambda_i P_i \right) A^{n-1} = \sum_{i=1}^k \lambda_i (P_i A^{n-1}) = \sum_{i=1}^k \lambda_i^n P_i,$$

特别地  $A^0 = I = \sum_{i=1}^k P_i$ .

因此

$$\begin{aligned} f(A) &= \sum_{l=0}^m a_l A^l && (f(\lambda) = \sum_{l=0}^m a_l \lambda^l) \\ &= \sum_{l=0}^m a_l \sum_{i=1}^k \lambda_i^l P_i && (A^n = \sum_{i=1}^k \lambda_i^n P_i) \\ &= \sum_{i=1}^k \sum_{l=0}^m a_l \lambda_i^l P_i \\ &= \sum_{i=1}^k \left( \sum_{l=0}^m a_l \lambda_i^l \right) P_i \\ &= \sum_{i=1}^k f(\lambda_i) P_i. && (f(\lambda) = \sum_{l=0}^m a_l \lambda^l) \end{aligned}$$

(3) 已知  $\varphi_i(\lambda) = \prod_{\substack{j=1 \\ j \neq i}}^k (\lambda - \lambda_j)$ , 即

$$\varphi_1(\lambda) = (\lambda - \lambda_2)(\lambda - \lambda_3) \cdots (\lambda - \lambda_k),$$

$$\varphi_2(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_3) \cdots (\lambda - \lambda_k),$$

$\vdots$

$$\varphi_k(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_{k-1}),$$

则  $i \neq j$  时,

$$\varphi_i(\lambda_j) = 0.$$

(例如,  $\varphi_1(\lambda_2) = 0, \varphi_1(\lambda_3) = 0, \dots, \varphi_1(\lambda_k) = 0$ .)

由  $f(\mathbf{A}) = \sum_{i=1}^k f(\lambda_i) \mathbf{P}_i$ , 知

$$\varphi_i(\mathbf{A}) = \sum_{l=1}^k \varphi_i(\lambda_l) \mathbf{P}_l = \varphi_i(\lambda_i) \mathbf{P}_i,$$

所以

$$\mathbf{P}_i = \frac{\varphi_i(\mathbf{A})}{\varphi_i(\lambda_i)}, \quad i = 1, 2, \dots, k.$$

例如

$$\begin{aligned} \mathbf{P}_1 &= \frac{\varphi_1(\mathbf{A})}{\varphi_1(\lambda_1)} = \frac{(\mathbf{A} - \lambda_2 \mathbf{I})(\mathbf{A} - \lambda_3 \mathbf{I}) \cdots (\mathbf{A} - \lambda_k \mathbf{I})}{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3) \cdots (\lambda_1 - \lambda_k)}, \\ \mathbf{P}_2 &= \frac{\varphi_2(\mathbf{A})}{\varphi_2(\lambda_2)} = \frac{(\mathbf{A} - \lambda_1 \mathbf{I})(\mathbf{A} - \lambda_3 \mathbf{I}) \cdots (\mathbf{A} - \lambda_k \mathbf{I})}{(\lambda_2 - \lambda_1)(\lambda_2 - \lambda_3) \cdots (\lambda_2 - \lambda_k)}. \end{aligned}$$

这为谱分解的计算提供了一个新的方法.

(4) 因  $\mathbf{P}_i = \frac{\varphi_i(\mathbf{A})}{\varphi_i(\lambda_i)}$ , 故:

矩阵  $\mathbf{B}$  与  $\mathbf{A}$  可交换  $\Leftrightarrow$  矩阵  $\mathbf{B}$  与  $\varphi_i(\mathbf{A})$  可交换  $\Leftrightarrow \mathbf{B}$  与  $\mathbf{P}_i$  可交换.  $\square$

Example 47. 求矩阵  $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix}$  的谱分解.

解: 由

$$|\lambda \mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda - 1 & -1 \\ -4 & \lambda - 1 \end{vmatrix} = (\lambda - 3)(\lambda + 1),$$

得特征值  $\lambda_1 = 3, \lambda_2 = -1$ . 故  $\varphi_1(\lambda) = (\lambda - \lambda_2), \varphi_2(\lambda) = (\lambda - \lambda_1)$ .

$$\begin{aligned} \mathbf{P}_1 &= \frac{\varphi_1(\mathbf{A})}{\varphi_1(\lambda_1)} = \frac{\mathbf{A} - \lambda_2 \mathbf{I}}{\lambda_1 - \lambda_2} = \frac{1}{4} \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ 1 & \frac{1}{2} \end{bmatrix}, \\ \mathbf{P}_2 &= \frac{\varphi_2(\mathbf{A})}{\varphi_2(\lambda_1)} = \frac{\mathbf{A} - \lambda_1 \mathbf{I}}{\lambda_2 - \lambda_1} = \frac{1}{-4} \begin{bmatrix} -2 & 1 \\ 4 & -2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} \\ -1 & \frac{1}{2} \end{bmatrix}. \end{aligned}$$

故  $\mathbf{A}$  的谱分解为

$$\mathbf{A} = \lambda_1 \mathbf{P}_1 + \lambda_2 \mathbf{P}_2 = 3 \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ 1 & \frac{1}{2} \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} \\ -1 & \frac{1}{2} \end{bmatrix}. \quad \square$$

## 正交投影

**Definition 48.** 设  $T$  是  $\mathbb{C}^n = L \oplus M$  上的沿  $M$  向  $L$  的投影算子, 并且  $L \perp M$  (即  $L^\perp = M$ ), 则称  $T$  为  $\mathbb{C}^n$  上的沿  $M$  向  $L$  的正交投影算子.



**Theorem 49.**  $P \in \mathbb{C}^{n \times n}$  为  $\mathbb{C}^n = L \oplus M$  上的沿  $M$  向  $L$  的正交投影算子的充分必要条件是:  $P$  为幂等的 Hermite 矩阵, 且  $L = R(P)$ ,  $M = N(P)$ .

**证:** 必要性.  $P$  为  $\mathbb{C}^n = L \oplus M$  上的沿  $M$  向  $L$  的正交投影算子, 则  $L = R(P)$ ,  $M = N(P)$ ,  $P^2 = P$ . 由  $L \perp M$ , 即有  $R(P) \perp N(P)$ . 下证  $P^H = P$ .

由幂等矩阵性质 (6), 有  $N(P) = R(I_n - P)$ , 故

$$R(P) \perp R(I_n - P).$$

对任意  $y \in \mathbb{C}^n$ , 都有

$$Py \in R(P), \quad (I_n - P)y \in R(I_n - P),$$

所以

$$Py \perp (I_n - P)y,$$

即二者内积为零:

$$(Py | (I_n - P)y) = 0,$$

或

$$((I_n - P)y)^H Py = 0, \quad y^H (I_n - P^H) Py = 0.$$

当  $y$  取  $\mathbb{C}^n$  的自然基底  $e_i$  时, 上式也成立. 故

$$e_i^H (I_n - P^H) P e_i = 0, \quad i = 1, 2, \dots, n.$$

即  $e_i^T (I_n - P^H) P e_i = 0$ . 等价于:

$$\begin{bmatrix} e_1^T \\ \vdots \\ e_n^T \end{bmatrix} (I_n - P^H) P \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} = O.$$

即

$$I_n (I_n - P^H) P I_n = O.$$

故  $(I_n - P^H)P = O$ , 得  $P^H P = P$ , 故  $P^H = P$ .

充分性.  $P$  是幂等矩阵, 故  $P$  是沿  $M = N(P)$  向  $L = R(P)$  的投影算子.

下证  $R(P) \perp N(P)$ .

任取  $x \in N(P)$ ,  $y \in R(P)$ , 则  $Px = 0$ , 且存在  $u \in \mathbb{C}^n$ , 使得  $Pu = y$ . 从而

$$\begin{aligned} (x | y) &= (x | Pu) = u^H P^H x \\ &= u^H Px && (P^H = P) \\ &= u^H 0 = 0. && (Px = 0) \end{aligned}$$

由  $x, y$  的任意性, 得证  $R(P) \perp N(P)$ .  $\square$

**Theorem 50.** 设  $A \in \mathbb{C}^{n \times n}$  共有  $k$  个互异的特征值  $\lambda_1, \lambda_2, \dots, \lambda_k$ , 则  $A$  是正规矩阵的充分必要条件为: 存在  $k$  个正交投影算子  $P_1, P_2, \dots, P_k$ , 满足:

1.  $P_i P_j = O, i \neq j, P_i \in \mathbb{C}^{n \times n}, i = 1, 2, \dots, k;$
2.  $\sum_{i=1}^k P_i = I;$
3.  $A = \sum_{i=1}^k \lambda_i P_i.$

**证:** 充分性.  $P_i$  为正交投影算子, 则  $P_i$  为幂等的 Hermite 矩阵. 由  $A = \sum_{i=1}^k \lambda_i P_i,$  有

$$A^H = \sum_{i=1}^k \bar{\lambda}_i P_i^H = \sum_{i=1}^k \bar{\lambda}_i P_i.$$

再由  $AP_i = P_i A = \lambda_i P_i$  (见 (16) 式), 则有

$$AA^H = A \sum_{i=1}^k \bar{\lambda}_i P_i = \sum_{i=1}^k \bar{\lambda}_i AP_i = \sum_{i=1}^k \bar{\lambda}_i P_i A = A^H A,$$

即证  $A$  是正规矩阵.

*Example 51.* 考虑空间  $\mathbb{C}^4$ , 取其自然基底  $e_1, e_2, e_3, e_4$ , 并令  $L = \text{span}[e_1, e_2],$   $M = \text{span}[e_3, e_4],$  则显然有  $\mathbb{C}^4 = L \oplus M,$  且  $L \perp M.$  试求  $\mathbb{C}^4$  上沿  $M$  向  $L$  的正交投影算子  $P \in \mathbb{C}^{4 \times 4}.$

**解:** 由投影算子的性质知  $P$  应满足条件:

- (1)  $M = N(P),$  故  $Pe_3 = 0, Pe_4 = 0.$
  - (2)  $L = R(P),$  且  $P$  是  $R(P)$  上的恒等变换, 故  $Pe_1 = e_1, Pe_2 = e_2.$
- 因此

$$P[e_1, e_2, e_3, e_4] = [Pe_1, Pe_2, Pe_3, Pe_4] = [e_1, e_2, 0, 0],$$

而  $[e_1, e_2, e_3, e_4] = I_4,$  故上式即为

$$P = [e_1, e_2, 0, 0]. \quad \square$$

**Exercise 52** (P.135 习题 (一) 6). 证明正交投影算子  $P \in \mathbb{C}^{n \times n}$  为半正定矩阵.

**证:** 因  $P \in \mathbb{C}^{n \times n}$  是正交投影算子, 故  $P$  为幂等的 Hermite 矩阵, 故

$$P^H P = P P = P.$$

任意  $x \in \mathbb{C}^n,$  有

$$x^H P x = x^H P^H P x = (P x | P x) \geq 0,$$

所以  $P$  为半正定矩阵. □

或者, 对任意矩阵  $A,$  有  $A^H A$  为半正定矩阵 (见教材 P.118, 或本文的定理 17), 故  $P = P^H P$  为半正定矩阵.

**Exercise 53** (P.94 习题 (二) 5). 设  $T^2 = T, S^2 = S, T, S \in L(X, X),$  证明

- (1)  $T$  与  $S$  有相同的值域  $\Leftrightarrow TS = S, ST = T.$
- (2)  $T$  与  $S$  有相同的核  $\Leftrightarrow TS = T, ST = S.$

(3)  $T$  与  $S$  的特征值非 0 即 1.

**证:** 因  $T$  和  $S$  为幂等算子, 故  $T$  是沿  $N(T)$  向  $R(T)$  的投影算子;  $S$  是沿  $N(S)$  向  $R(S)$  的投影算子.

(1) 必要性. 设  $R(T) = R(S)$ . 任取  $\alpha \in X$ , 有  $S\alpha \in R(S)$ , 从而  $S\alpha \in R(T)$ . 因  $T$  在  $R(T)$  上为恒等算子, 故

$$(TS)\alpha = T(S\alpha) = S\alpha,$$

由  $\alpha$  的任意性, 知  $TS = S$ . 对称地,  $ST = T$ .

充分性. 设  $TS = S, ST = T$ . 任取  $\gamma \in R(T)$ , 因  $T$  在  $R(T)$  上为恒等算子, 故  $T\gamma = \gamma$ . 又

$$S\gamma = S(T\gamma) = (ST)\gamma = T\gamma = \gamma,$$

于是  $\gamma = S\gamma \in R(S)$ , 故  $R(T) \subseteq R(S)$ . 对称地,  $R(S) \subseteq R(T)$ . 得证  $R(T) = R(S)$ .

(2) 必要性. 设  $N(T) = N(S)$ . 由投影算子的定义知  $X = R(T) \oplus N(T)$ . 任取  $\alpha \in X$ , 有

$$\alpha = x + y, \quad x \in R(T), y \in N(T).$$

于是  $\alpha - x \in N(T)$ . 又  $N(T) = N(S)$ , 故  $\alpha - x \in N(S)$ , 于是

$$S(\alpha - x) = 0, \quad \Rightarrow \quad S\alpha = Sx.$$

因  $T$  是沿  $N(T)$  向  $R(T)$  的投影算子, 故  $T\alpha = x$ . 从而

$$(ST)\alpha = S(T\alpha) = Sx = S\alpha.$$

由  $\alpha$  的任意性, 知  $ST = S$ . 对称地,  $ST = S$ .

充分性. 设  $TS = T, ST = S$ . 任取  $\delta \in N(T)$ , 则  $T\delta = 0$ , 从而

$$(ST)\delta = S(T\delta) = 0.$$

由  $ST = S$ , 故

$$S\delta = (ST)\delta = 0.$$

故  $\delta \in N(S)$ , 从而  $N(T) \subseteq N(S)$ . 对称地有  $N(S) \subseteq N(T)$ , 得证  $N(T) = N(S)$ .

(3)  $T$  和  $S$  为幂等算子, 则其矩阵表示均为幂等矩阵. 而幂等矩阵的特征值非 0 即 1, 又线性变换与其矩阵表示有相同的特征值, 得证  $T$  与  $S$  的特征值非 0 即 1. □