

# 线性代数讲义

## 目录

1. 特征值与特征向量	2
1.1. 特征值与特征向量	2
1.1.1. 引入	2
1.1.2. 特征向量与特征值的定义	3
1.1.3. 行列式与迹	6
1.1.4. 矩阵与其特征值的关系	6
1.2. 矩阵的对角化, 矩阵相似	8
1.2.1. 矩阵可对角化的定义	8
1.2.2. 马氏链	9
1.2.3. 阅读	11
1.2.4. 矩阵可对角化的条件(一)	12
1.2.5. 矩阵可对角化的条件(二)	14
1.2.6. Cayley-Hamilton定理	17
1.2.7. $AB$ 的特征值	17
1.2.8. 矩阵的相似	19
1.3. 对称矩阵	21
1.3.1. 对称矩阵的谱定理	21
1.3.2. 实对称矩阵的线性代数基本定理	23
1.3.3. 如何对角化对称矩阵	23
1.3.4. 对称矩阵的主元与特征值(阅读)	25
1.4. 正定对称矩阵	25
1.4.1. 正定对称矩阵的定义	26
1.4.2. 正定对称矩阵的判别	27
1.4.3. 半正定对称矩阵的判别	30
1.4.4. Rayleigh商	31

1.4.5. 二次曲面分类(阅读)	32
2. 奇异值分解 (Singular Value Decomposition)	35
2.1. 奇异值分解定理	35
2.1.1. 定理的陈述	35
2.1.2. SVD定理的证明	35
2.1.3. 奇异值分解的几何意义	37
2.1.4. SVD与线性代数基本定理 (v.3.0)	38
2.2. 矩阵的低秩逼近	39
2.2.1. 矩阵的范数与矩阵空间上的距离	41
2.3. 奇异值分解的应用	43
2.3.1. 数据压缩	43
2.3.2. 广义逆	44
2.3.3. 广义逆与正交投影	47
2.3.4. 最小二乘问题	48

## 1. 特征值与特征向量

### 1.1. 特征值与特征向量.

1.1.1. 引入. 考虑某地的城镇化。假定每年有97%的人口选则留在城市，3%的人口选择由城镇迁移至乡村；每年有95%的人口选则留在农村，5%的人口选择由乡村迁移至城镇。如果初始有 $u_{country,0} = 60$ 万人在农村， $u_{city,0} = 40$ 万人在城市，总人口共100万。一年后，

$$\begin{bmatrix} u_{city,1} \\ u_{country,1} \end{bmatrix} = \begin{bmatrix} 0.97u_{city,0} + 0.05u_{country,0} \\ 0.03u_{city,0} + 0.95u_{country,0} \end{bmatrix} = \begin{bmatrix} 0.97 & 0.05 \\ 0.03 & 0.95 \end{bmatrix} \begin{bmatrix} u_{city,0} \\ u_{country,0} \end{bmatrix} = \begin{bmatrix} 418,000 \\ 582,000 \end{bmatrix}$$

于是下面的矩阵记住了人口迁移的情况

$$A = \begin{bmatrix} 0.97 & 0.05 \\ 0.03 & 0.95 \end{bmatrix}$$

我们希望预测多年以后的人口分布情况。做两个假定：总人口不变，每年的迁移比例也不变。第 $k$ 年人口分布为：

$$\mathbf{u}_k = A\mathbf{u}_{k-1}, \dots, \mathbf{u}_2 = A\mathbf{u}_1, \mathbf{u}_1 = A\mathbf{u}_0.$$

$$\mathbf{u}_k = \begin{bmatrix} u_{city} \\ u_{country} \end{bmatrix}_k = A^k \begin{bmatrix} u_{city} \\ u_{country} \end{bmatrix}_0.$$

计算矩阵的方幂作用在一个向量上, 找矩阵的特征向量和对矩阵进行对角化是一个常用方法。

1.1.2. 特征向量与特征值的定义. 在本章中我们考虑复矩阵, 复特征值和复特征向量。关于复矩阵, 高斯消元解方程的过程, 行列式的计算和定义等和实矩阵一样。不同之处在于 $\mathbb{C}^n$ 的内积应该定义为 $\bar{\mathbf{v}}^T \mathbf{w}$ , 其中 $\bar{\mathbf{v}} = [\bar{v}_1, \dots, \bar{v}_n]^T$ ,  $v_i \in \mathbb{C}$ . 正交性改为 $\mathbf{v} \perp \mathbf{w}$ , 如果 $\bar{\mathbf{v}}^T \mathbf{w} = 0$ . 于是 $\bar{\mathbf{v}}^T \mathbf{v} = 0$ 当且仅当 $|v_1|^2 + \dots + |v_n|^2 = 0$ , 当且仅当 $v_1 = \dots = v_n = 0$ . 另外, 实向量空间的概念也需要推广到复向量空间。

**定义1.1** (特征向量、特征值). 设 $A \in M_{n \times n}(\mathbb{C})$ 为 $n$ 阶方阵。若有非零向量 $\mathbf{x} \in \mathbb{C}^n$ ,  $\lambda \in \mathbb{C}$ , 满足

$$A\mathbf{x} = \lambda\mathbf{x}.$$

则称 $\lambda$ 为 $A$ 的特征值(eigenvalue),  $\mathbf{x}$ 为 $A$ 的特征值为 $\lambda$ 的特征向量(eigenvector)。

**注记1.2.** (1) 零向量不是特征向量。

(2) 如果 $\mathbf{x}$ 是特征值为 $\lambda$ 的特征向量, 则 $c\mathbf{x}$ ,  $c \neq 0$ 也是 $A$ 的特征值为 $\lambda$ 的特征向量。

(3) 如果 $\mathbf{x} \neq 0$ 满足 $A\mathbf{x} = \lambda\mathbf{x}$ . 则

$$A^k \mathbf{x} = \lambda^k \mathbf{x}, (k \geq 1)$$

即 $\mathbf{x}$ 为 $A^k$ 的特征值为 $\lambda^k$ 的特征向量. ( $A^k(\mathbf{x}) = A^{k-1}(\lambda\mathbf{x}) = \dots = \lambda^k \mathbf{x}$ )。

根据定义, 容易看出以下几条等价。

**命题1.3.** 以下几条等价:

- (1)  $\lambda \in \mathbb{C}$ 为 $A$ 的特征值.
- (2) 方程 $(A - \lambda I_n)\mathbf{x} = \mathbf{0}$ 有非零解.
- (3) 零空间 $N(A - \lambda I_n)$ 包含非零向量.
- (4) 方阵 $A - \lambda I_n$ 不可逆.

(5) 行列式  $|A - \lambda I_n| = 0$ .

**定义1.4.** 设  $\lambda$  为  $A$  的特征值.  $\mathbb{C}^n$  的子空间  $N(A - \lambda I_n)$  称为  $A$  的特征值为  $\lambda$  的特征子空间 (*eigenspace*).

**例1.5.** 根据定义  $N(A) = \{0\} \cup \{\text{特征值为0的特征向量}\}$ ;  $N(A - \lambda I_n) = \{0\} \cup \{\text{特征值为}\lambda\text{的特征向量}\}$ .

**例1.6.** 如果  $A$  可逆,  $\mathbf{x}$  是特征值为  $\lambda$  的特征向量, 那么  $\lambda \neq 0$  且  $A^{-1}(\mathbf{x}) = \lambda^{-1}\mathbf{x}$ .

证明. 首先  $\lambda \neq 0$ , 否则  $A\mathbf{x} = 0\mathbf{x} = \mathbf{0}$  有非零解, 与  $A$  可逆矛盾.

设  $\mathbf{x}$  为特征向量,  $\mathbf{x} = I\mathbf{x} = A^{-1}A\mathbf{x} = A^{-1}\lambda\mathbf{x}$ . 得到  $A^{-1}\mathbf{x} = \lambda^{-1}\mathbf{x}$ .

□

从第(5)条出发, 找  $A$  的特征值.

**定义1.7** (特征多项式). 将  $\lambda$  视为不定变元. 令  $f_A(\lambda) = |A - \lambda I_n|$ .  $f_A(\lambda)$  称为  $A$  的特征多项式 (**characteristic polynomial**).  $f(\lambda)$  为  $\lambda$  的  $n$  次多项式, 最高次数项为  $(-1)^n \lambda^n$ .

根据(1)(5)等价, 我们有:

**命题1.8.**  $\lambda$  为  $A$  的特征值当且仅当  $f_A(\lambda) = 0$ .

**例1.9** (二阶矩阵).  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ .

$$f_A(\lambda) = \det(A - \lambda I_2) = \begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = \lambda^2 - (a + d)\lambda + (ad - bc).$$

$a + d$  为  $A$  的迹 (*trace*),  $ad - bc$  为  $A$  的行列式.

**例1.10** (3阶矩阵).  $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ .  $|A - \lambda I_3| = \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{vmatrix} = (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda) + \deg \leq 1$ .

**例1.11.** 上三角或下三角矩阵的特征值为对角线上的元素。

以下三角矩阵为例, 设  $L = \begin{bmatrix} a_{11} & & \\ \vdots & \ddots & \\ a_{1n} & \cdots & a_{nn} \end{bmatrix}$ .

$$f_L(\lambda) = |L - \lambda I| = (a_{11} - \lambda) \cdots (a_{nn} - \lambda),$$

$f_L(\lambda) = 0$  当且仅当  $\lambda \in \{a_{11}, \dots, a_{nn}\}$ .

**命题1.12.**  $f_A(\lambda) = f_{A^T}(\lambda)$ . 因此,  $A$  与  $A^T$  有相同的特征值。

证明.  $f_A(\lambda) = \det(A - \lambda I) = \det(A - \lambda I)^T = \det(A^T - \lambda I) = f_{A^T}(\lambda)$ .

□

**例1.13** (旋转矩阵).  $Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ , 特征多项式为  $\lambda^2 + 1$ , 在  $\mathbb{R}$  中无解。所以,  $Q$  作为  $M_{2 \times 2}(\mathbb{R})$  中的矩阵没有特征值与特征向量。这与直观符合:  $\mathbb{R}^2$  逆时针旋转  $90^\circ$ ,  $Q\mathbf{x}$  不可能与  $\mathbf{x}$  共线。

求  $Q$  的特征值与特征向量需要将  $Q$  视为复矩阵,  $\lambda^2 + 1 = 0$  有两个解  $i, -i$ , 且有

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = -i \begin{bmatrix} 1 \\ i \end{bmatrix}, \quad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} i \\ 1 \end{bmatrix} = i \begin{bmatrix} i \\ 1 \end{bmatrix}.$$

$Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ .  $f_Q(\lambda) = \lambda^2 + 1 = (\lambda - i)(\lambda + i)$ .  $\det Q = 1 = (-i)(i)$ ,  $\text{trace}(Q) = -i + i = 0 = a_{11} + a_{22}$ .

**注记1.14.** 求特征向量的一般方法:

- (1) 计算行列式  $f_A(\lambda) = \det(A - \lambda I)$ ;
- (2) 找到多项式  $f_A(\lambda)$  的所有根, 即  $f_A(\lambda) = 0$  的所有解。它们是  $A$  的所有特征值;
- (3) 对每个特征值  $\lambda$ , 求解方程  $(A - \lambda I)\mathbf{x} = \mathbf{0}$ 。

**定理1.15** (代数基本定理). 一个  $n$  次的一元多项式  $f(x) \in \mathbb{C}[x]$  在  $\mathbb{C}$  中有  $n$  个根 (计算重数)。

由代数基本定理,  $A \in M_{n \times n}(\mathbb{C})$ ,

$$f_A(\lambda) = (\lambda_1 - \lambda)^{n_1} (\lambda_2 - \lambda)^{n_2} \cdots (\lambda_k - \lambda)^{n_k},$$

其中 $\lambda_i \in \mathbb{C}$ ,互不相同.  $n_1, \dots, n_k \geq 1$ ,  $n_1 + \dots + n_k = n$ .

$AM(\lambda_i) = n_i$ 称为特征值 $\lambda_i$ 对应的**代数重数(algebraic multiplicity)**。 $GM(\lambda_i) = d_i := \dim N(A - \lambda_i I_n)$ 称为 $\lambda_i$ 的**几何重数 (geometric multiplicity)**。

1.1.3. 行列式与迹. 在 $2 \times 2$ 的例子中, 我们看到 $A$ 的行列式与迹在 $f_A(\lambda)$ 中出现, 这是一般现象。

**定义1.16.**  $A = (a_{ij})$ 的**迹(trace)**为 $\text{trace}(A) := a_{11} + a_{22} + \dots + a_{nn}$ 。

**定理1.17.** (1) 行列式(=  $\pm$  主元乘积)=特征值之积, 即 $|A| = \lambda_1 \lambda_2 \dots \lambda_n$ .

(2) 迹=特征值之和, 即 $\text{trace}(A) = \lambda_1 + \lambda_2 + \dots + \lambda_n$ .

(3)  $f_A(\lambda) = |A| + \dots + (-1)^{n-1}(\text{trace}(A))\lambda^{n-1} + (-1)^n \lambda^n$ .

证明. 设 $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ 为 $A$ 的特征值。设

$$f_A(\lambda) = a_0 + a_1 \lambda + \dots + a_{n-1} \lambda^{n-1} + (-1)^n \lambda^n = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda).$$

- $|A| = |(a_{ij})| = f_A(0) = \lambda_1 \lambda_2 \dots \lambda_n$ .
- 将 $f_A(\lambda) = |A - \lambda I|$ 沿第一行展开, 得到

$$\begin{aligned} f_A(\lambda) &= (a_{11} - \lambda)(a_{22} - \lambda) \dots (a_{nn} - \lambda) + (\deg \leq n-2) \\ &= (-1)^n \lambda^n + (a_{11} + \dots + a_{nn})(-1)^{n-1} \lambda^{n-1} + (\deg \leq n-2) \end{aligned}$$

另一方面,

$$f_A(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda) = (-1)^n \lambda^n + (-1)^{n-1}(\lambda_1 + \dots + \lambda_n) \lambda^{n-1} + (\deg \leq n-2)$$

对比得知,  $a_{11} + \dots + a_{nn} = \lambda_1 + \dots + \lambda_n$ . □

**注记1.18.** 不少教材也将 $|\lambda I_n - A|$ 定义为 $A$ 的特征多项式. 这不影响特征值. 区别是特征多项式以 $\lambda^n$ 为最高次项,  $(-1)^n |A|$ 为常数项.

1.1.4. 矩阵与其特征值的关系. 我们再看几个矩阵及其特征值的关系。

实矩阵的复特征值:

$A \in M_{n \times n}(\mathbb{R})$ ,  $f_A(\lambda)$  在  $\mathbb{C}$  中总有  $n$  个根 (计重数)。如果  $\lambda_0 \in \mathbb{C} \setminus \mathbb{R}$  为  $A$  的复特征值, 即  $f_A(\lambda_0) = 0$ , 解方程  $(A - \lambda_0 I)\mathbf{x} = 0$  得到解  $\mathbf{x} \in \mathbb{C}$  满足

$$A\mathbf{x} = \lambda_0 \mathbf{x},$$

取共轭, 我们有

$$\overline{A\mathbf{x}} = \overline{\lambda_0 \mathbf{x}}.$$

由于  $A$  为实矩阵,  $\overline{A} = A$ , 我们有

$$A\overline{\mathbf{x}} = \overline{\lambda_0} \overline{\mathbf{x}}.$$

所以,  $\overline{\lambda_0}$  也同为  $A$  的特征值, 对应特征向量为  $\overline{\mathbf{x}}$ 。

由于  $A$  为实矩阵,  $f_A(\lambda) \in \mathbb{R}[\lambda]$  为实多项式. 设  $n_0 = AM(\lambda_0)$ , 设  $n'_0 = AM(\overline{\lambda_0})$ , 则  $f_A(\lambda) = (\lambda_0 - \lambda)^{n_0}(\overline{\lambda_0} - \lambda)^{n'_0} \cdots f_A(\lambda)$  为实多项式得到  $n_0 = n'_0$ . 由于  $N(A - \lambda_0 I)$  中线性无关的向量  $\mathbf{x}_1, \dots, \mathbf{x}_d$  取共轭后得到线性无关的向量  $\overline{\mathbf{x}}_1, \dots, \overline{\mathbf{x}}_d$ , 我们也有  $GM(\lambda_0) = GM(\overline{\lambda_0})$ .

总结: 对于实矩阵, 复特征值 ( $\lambda \in \mathbb{C} \setminus \mathbb{R}$ ) 成对出现。“成对”意思是: (1) 如果  $\lambda \in \mathbb{C} \setminus \mathbb{R}$  是特征值,  $\overline{\lambda}$  也是特征值; (2)  $\lambda$  与  $\overline{\lambda}$  作为特征值的重数相等.

*Gershgorin* 圆盘定理 (对角占优矩阵的应用):

矩阵的特征值的分布依赖于矩阵的系数. 例如,  $A = (a_{ij}) \in M_{3 \times 3}$ . 如果  $\lambda$  是  $A$  的特征值, 则矩阵  $A - \lambda I = \begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{bmatrix}$  不可逆.

另一方面, 利用对角占优矩阵, 如果  $|a_{11} - \lambda| > |a_{12}| + |a_{13}|$ ,  $|a_{22} - \lambda| > |a_{21}| + |a_{23}|$ ,  $|a_{33} - \lambda| > |a_{31}| + |a_{32}|$ , 则  $(A - \lambda I)$  可逆。

于是, 如果  $\lambda$  是  $A$  的特征值, 则  $\lambda$  一定满足  $|\lambda - a_{11}| \leq |a_{12}| + |a_{13}|$  或者  $|\lambda - a_{22}| \leq |a_{21}| + |a_{23}|$  或者  $|\lambda - a_{33}| \leq |a_{31}| + |a_{32}|$ . 即  $\lambda$  一定落在上述三个圆盘的中一个。

同样的理由我们有 *Gershgorin* 圆盘定理:

**命题 1.19** (*Gershgorin* 圆盘定理). 对矩阵  $A = (a_{ij})$  定义如下  $n$  个圆盘

$$G_i(A) = \{z \mid |z - a_{ii}| \leq \sum_{j \neq i} |a_{ij}|\}.$$

$A$ 的全部特征值一定落在这 $n$ 个圆盘中.

## 1.2. 矩阵的对角化, 矩阵相似.

### 1.2.1. 矩阵可对角化的定义.

**例1.20** (投影矩阵).  $\mathbf{0} \neq \mathbf{u} \in \mathbb{R}^2$ ,  $\|\mathbf{u}\| = 1$ ,  $P = \mathbf{u}\mathbf{u}^T$ . 则  $P: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  为向  $\mathbb{R}\mathbf{u}$  投影的投影矩阵.

$P(\mathbf{u}) = \mathbf{u}$ , 所以  $\mathbf{u}$  为特征值为1的特征向量;

取非零向量  $\mathbf{u}^\perp \perp \mathbf{u}$ , 则  $\mathbf{u}^\perp$  为特征值为0的特征向量.

两个相互垂直的特征向量  $\mathbf{u}, \mathbf{u}^\perp$  构成  $\mathbb{R}^2$  的一组基.

$$P[\mathbf{u}, \mathbf{u}^\perp] = [\mathbf{u}, \mathbf{u}^\perp] \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

令  $X = [\mathbf{u}, \mathbf{u}^\perp] \in M_{2 \times 2}(\mathbb{R})$ , 得到  $PX = X \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ , 即  $X^{-1}PX = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ .

一般地, 设  $P$  为  $\mathbb{R}^m = C(A) \oplus N(A^T)$  向子空间  $C(A)$  投影的投影矩阵. 令  $\mathbf{x}_1, \dots, \mathbf{x}_r$  为  $C(A)$  的一组基,  $\mathbf{x}_{r+1}, \dots, \mathbf{x}_m$  为  $N(A^T)$  的一组基. 则  $P(\mathbf{x}_i) = \mathbf{x}_i$ ,  $1 \leq i \leq r$ ,  $P(\mathbf{x}_i) = \mathbf{0}$ ,  $r+1 \leq i \leq m$ . 令  $X = [\mathbf{x}_1, \dots, \mathbf{x}_m]$ ,

$$AX = X \operatorname{diag}(1, \dots, 1, 0, \dots, 0), \quad X^{-1}AX = \operatorname{diag}(1, \dots, 1, 0, \dots, 0).$$

**例1.21** (镜面反射矩阵).  $R = I - 2\mathbf{u}\mathbf{u}^T$ ,  $\mathbf{u} \in \mathbb{R}^2$ ,  $\|\mathbf{u}\| = 1$ .  $R^T = R$ . 例如, 取  $\mathbf{u} = [1/\sqrt{2}, -1/\sqrt{2}]^T$ ,  $R = \begin{bmatrix} & 1 \\ 1 & \end{bmatrix}$ .

取  $\mathbf{u}^\perp$  非零且正交于  $\mathbf{u}$

- $R\mathbf{u} = -\mathbf{u}$ ;
- $R\mathbf{u}^\perp = \mathbf{u}^\perp$ .

$\mathbf{u}, \mathbf{u}^\perp$  构成  $\mathbb{R}^2$  的一组基, 它们都是  $R$  的特征向量, 特征值为1与-1. 令  $X = [\mathbf{u}, \mathbf{u}^\perp]$ .

$$RX = X \begin{bmatrix} -1 & \\ & 1 \end{bmatrix}, \quad X^{-1}RX = \begin{bmatrix} -1 & \\ & 1 \end{bmatrix}.$$



$$R = \begin{bmatrix} & 1 \\ 1 & \end{bmatrix}, f_R(\lambda) = \lambda^2 - 1 = (\lambda - 1)(\lambda + 1).$$

$$(R - I)\mathbf{x} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \mathbf{x} = 0, \text{ 得到 } \mathbf{u}^\perp = [1, 1]^T.$$

$$(R - (-1)I)\mathbf{x} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \mathbf{x} = 0, \text{ 得到 } \mathbf{u} = [1, -1]^T.$$

**注记1.22.**  $P, R$  为对称矩阵, 存在一组相互正交的特征向量构成它们的一组基。这是一个一般现象, 我们会在对称矩阵一节进一步研究。

**定义1.23.**  $A$  称为可对角化的 (diagonalizable), 如果存在可逆矩阵  $X$  使得

$$X^{-1}AX = \Lambda$$

其中,  $\Lambda$  为对角矩阵。

**命题1.24.** 如果  $X^{-1}AX = \Lambda$ ,  $\Lambda$  为对角矩阵, 那么  $\Lambda$  对角线元素为  $A$  的特征值。

证明.  $f_A(\lambda) = |A - \lambda I| = |X\Lambda X^{-1} - \lambda I| = |X||\Lambda - \lambda I||X^{-1}| = |\Lambda - \lambda I| = (\lambda_1 - \lambda) \cdots (\lambda_n - \lambda)$ , 其中  $\lambda_1, \dots, \lambda_n$  为  $\Lambda$  对角线上元素。□

**例1.25.** 对于可对角化的矩阵容易计算对应线性映射的迭代, 即矩阵的幂: 如果  $A$  可对角化, 即有可逆矩阵  $X$  使得

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

那么  $A^k = X\Lambda^k X^{-1}$ .

证明.  $A^2 = X\Lambda X^{-1}X\Lambda X^{-1} = X\Lambda^2 X^{-1}; A^k = X\Lambda^k X^{-1}$ . □

1.2.2. 马氏链. 回到本章开始的例子

$$A = \begin{bmatrix} 0.97 & 0.05 \\ 0.03 & 0.95 \end{bmatrix}$$

即每年有97%的人口选则留在城市, 3%的人口选择由城镇迁移至乡村; 每年有95%的人口选则留在农村, 5%的人口选择由城镇迁移至乡村。 $A$  是 (马尔可夫矩阵) Markov matrix 或者 transition matrix.

如果当年有 $u_{country} = 60$ 万人在农村,  $u_{city} = 40$ 万人在城市, 总人口共100万。我们将总人数做单位化 $u_{city,0} + u_{country,0} = 1$ , 那么

$$\begin{bmatrix} u_{city} \\ u_{country} \end{bmatrix}_0 = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}.$$

$\lambda_1 = 1$ 是 $A$ 的一个特征值, 这是因为 $A - I_2 = \begin{bmatrix} -0.03 & 0.05 \\ 0.03 & -0.05 \end{bmatrix}$ ,  $(A - I_2)\mathbf{x} = \mathbf{0}$ 有解 $[5, 3]^T$ 。由于我们假设总人口数不变, 为1, 我们将得到的特征向量单位化,  $\mathbf{x}_1 = [5/8, 3/8]^T$ 。

由trace,  $A$ 的另一个特征值 $\lambda_2 = 0.97 + 0.95 - 1 = 0.92 < 1$ .  $\mathbf{x}_2 = [1, -1]$ 为其特征向量。

事实上, 在二阶马氏矩阵情形, 由于 $|\lambda_2| < 1$ ,  $\lambda_1 \neq \lambda_2$ 。矩阵 $A$ 可对角化, 此时的另一个特征向量一定为 $[1, -1]^T$ 。这是因为 $A = \begin{bmatrix} a & b \\ 1-a & 1-b \end{bmatrix}$ ,  $\lambda_1 = 1$ ,  $\lambda_2 = (a + 1 - b) - 1 = a - b$ .  $A - \lambda_2 I = \begin{bmatrix} b & b \\ 1-a & 1-a \end{bmatrix}$ 。所以,  $[1, -1]^T$ 为特征向量。

因此, 我们有

$$A[\mathbf{x}_1, \mathbf{x}_2] = [\mathbf{x}_1, \mathbf{x}_2] \begin{bmatrix} 1 & \\ & \lambda_2 \end{bmatrix}.$$

我们容易计算 $A$ 的方幂:

- $A^2 = X \Lambda X^{-1} X \Lambda X^{-1} = X \Lambda^2 X^{-1}$ ,
- $A^k = X \Lambda^k X^{-1} = X \begin{bmatrix} 1 & 0 \\ 0 & \lambda_2^k \end{bmatrix} X^{-1}$

下面我们计算 $A^\infty$ : 设 $\mathbf{u}_0 = [u_1, u_2]^T$ 为任一初始状态满足

$$(*) \quad u_1 + u_2 = 1.$$

$\mathbf{x}_1, \mathbf{x}_2$ 构成 $\mathbb{R}^2$ 的一组基, 有 $c_1, c_2$ 满足 $\mathbf{u}_0 = c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2$ 。我们有 $c_1 = 1$ 。这是因为 $1 = u_0 + u_1 = c_1((\mathbf{x}_1)_1 + (\mathbf{x}_1)_2) + c_2(1 + (-1)) = c_1$ 。因此,  $\mathbf{u}_0 = \mathbf{x}_1 + c_2 \mathbf{x}_2$ 。

于是, 我们有

$$\mathbf{u}_k = A^k \mathbf{u}_0 = \mathbf{x}_1 + c_2 \lambda_2^k \mathbf{x}_2$$

由于  $0 < \lambda_2 < 1$ , 我们得到  $k \rightarrow \infty$ ,  $\mathbf{u}_\infty = \mathbf{x}_1$ . 即无论初始状态如何, 最后都收敛到  $\mathbf{x}_1$ ,  $\mathbf{x}_1$  也称为**稳定状态向量(steady-state vector)**。

由于  $\mathbf{e}_1, \mathbf{e}_2$  满足条件(\*),  $A^k = A^k[\mathbf{e}_1, \mathbf{e}_2] = [\mathbf{x}_1 + c\lambda_2^k \mathbf{x}_2, \mathbf{x}_1 + d\lambda_2^k \mathbf{x}_2]$ ,  $c, d \in \mathbb{R}$ ; 取  $k \rightarrow \infty$ ,

$$A^\infty = [\mathbf{x}_1, \mathbf{x}_1].$$

结论: 不论从什么初始状态出发, 人口模型最后都会达到稳定态  $\mathbf{x}_1 = [5/8, 3/8]^T$ .

### 1.2.3. 阅读.

**定义1.26.** 一个方阵  $A = (a_{ij})_{1 \leq i, j \leq n}$  称为 (马尔可夫矩阵) **Markov matrix** 或者 **transition matrix**, 如果

- (1)  $a_{ij} \geq 0$ ;
- (2)  $a_{1j} + a_{2j} + \cdots + a_{nj} = 1, \forall j$ . 即每列求和为1。

**注记1.27.** 对于第(2)条, 我们也可用行代替列。

马尔可夫矩阵  $A$  有下列好的性质:

- (1) 1是  $A$  的一个特征值。

证明. 由于  $[1, \dots, 1]A = [1, \dots, 1]$ ,  $[1, \dots, 1](A - I_n) = 0$ . 所以,  $[1, \dots, 1]^T \in N((A - I_n)^T)$ . 因此,  $A - I_n$  奇异。□

- (2)  $A^k$  也是马氏矩阵。

证明. 首先  $A^k$  的每项大于等于0.

如果  $A$  是马氏矩阵,  $[1, \dots, 1]A = [1, \dots, 1]$ . 所以,

$$[1, \dots, 1]A^k = ([1, \dots, 1]A)A^{k-1} = [1, \dots, 1]A^{k-1} = \cdots = [1, \dots, 1].$$

因此,  $A^k$  也是马氏矩阵。□

我们做如下假定:

• 假定  $a_{ij} > 0$ . 事实上, 由于我们考虑长期的过程我们只需要有某个  $N$  使得  $A^N$  满足每项  $> 0$ .

在这个假定下, 我们还有

- (3)  $A$  的其它特征值  $\lambda$  满足  $|\lambda| < 1$ . 特别地,  $\lambda^k \rightarrow 0, k \rightarrow \infty$ .

因此,  $A$  有唯一的一个特征向量  $\mathbf{x} = [x_1, \dots, x_n]^T \in \mathbb{R}^n$ , 满足  $A\mathbf{x} = \mathbf{x}$  且  $x_1 + \dots + x_n = 1$ .  $\mathbf{x}$  称为 steady-state vector.

**定理1.28.** 如果  $A$  为马氏矩阵且  $a_{ij} > 0$ , 令  $\mathbf{q}$  为 steady-state vector. 那么, 对任意  $\mathbf{u}_0 = [u_1, \dots, u_n]^T$  满足  $u_i \geq 0$ ,  $u_1 + \dots + u_n = 1$ , 我们有

$$\lim_k A^k \mathbf{u}_0 = \mathbf{q}.$$

因此,  $A^\infty = [\mathbf{q}, \dots, \mathbf{q}]$ .

**注记1.29.** 马氏链也是 Google 网页搜索的理论基础, 关于此方面内容可参见

<http://faculty.winthrop.edu/polaskit/spring11/math550/chapter.pdf>

#### 1.2.4. 矩阵可对角化的条件(一).

**定理1.30.**  $A$  可对角化当且仅当  $A$  有  $n$  个线性无关特征向量.

证明. 如果  $A$  可对角化, 即有可逆矩阵  $X$  使得

$$X^{-1}AX = \Lambda.$$

于是  $AX = \Lambda X$ . 记  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ ,  $X = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ . 那么  $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$ . 即  $X$  的列向量为  $A$  的特征向量. 因为  $X$  可逆,  $\mathbf{x}_i$  线性无关.

反之, 设  $\mathbf{x}_1, \dots, \mathbf{x}_n$  为  $A$  的特征向量, 特征值分别为  $\lambda_1, \dots, \lambda_n$ .  $\mathbf{x}_1, \dots, \mathbf{x}_n$  线性无关, 它们构成  $\mathbb{R}^n$  的一组基. 令

$$X := [\mathbf{x}_1, \dots, \mathbf{x}_n].$$

于是,  $A[\mathbf{x}_1, \dots, \mathbf{x}_n] = [\mathbf{x}_1, \dots, \mathbf{x}_n] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$ . 因此,  $X^{-1}AX = \Lambda$ . □

**注记1.31.** 由  $Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$  的例子可以看出, 矩阵是否可对角化与将其视为实矩阵或复矩阵有密切关系.  $Q$  在  $M_{2 \times 2}(\mathbb{R})$  中不可对角化, 因为它的特征值不在  $\mathbb{R}$  中; 但  $Q$  在  $M_{2 \times 2}(\mathbb{C})$  中是可对角化的, 两个线性无关的特征向量分别为  $[1, i]^T$  与  $[i, 1]^T$ .

如不特殊指明, 本章考虑的可对角化都是作为复矩阵考虑。

**注记1.32.** 如果 $A$ 可对角化, 则在 $\mathbb{R}^n$ 中有一组基 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 由 $A$ 的特征向量构成. $A$ 对应的线性映射 $\mathcal{A}$ 即是对每个方向 $\mathbf{x}_i$ 伸缩 $\lambda_i$ 倍。

此时, 对于 $\mathbf{u}_0 = c_1\mathbf{x}_1 + \dots + c_n\mathbf{x}_n$ ,

$$\mathbf{u}_k = A\mathbf{u}_{k-1} = A^k\mathbf{u}_0 = c_1\lambda_1^k\mathbf{x}_1 + \dots + c_n\lambda_n^k\mathbf{x}_n.$$

证明.  $A^k(\mathbf{u}_0) = A^k(c_1\mathbf{x}_1 + \dots + c_n\mathbf{x}_n) = c_1A^k\mathbf{x}_1 + \dots + c_nA^k\mathbf{x}_n = c_1\lambda_1^k\mathbf{x}_1 + \dots + c_n\lambda_n^k\mathbf{x}_n.$   $\square$

**命题1.33.** 设 $\lambda_1, \dots, \lambda_k$ 为矩阵 $A$ 互不相同的特征值,  $\mathbf{x}_i$ 分别为特征值为 $\lambda_i$ 的特征向量, 那么 $\mathbf{x}_1, \dots, \mathbf{x}_k$  线性无关。

证明. 以 $n = 3$ 为例。设有 $c_1, c_2, c_3$ 使得

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 = \mathbf{0} \quad (1)$$

作用 $A, A^2$ , 得到

$$c_1\lambda_1\mathbf{x}_1 + c_2\lambda_2\mathbf{x}_2 + c_3\lambda_3\mathbf{x}_3 = \mathbf{0} \quad (2)$$

和

$$c_1\lambda_1^2\mathbf{x}_1 + c_2\lambda_2^2\mathbf{x}_2 + c_3\lambda_3^2\mathbf{x}_3 = \mathbf{0} \quad (3)$$

将(1)(2)(3)写为

$$[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3] \begin{bmatrix} c_1 & c_1\lambda_1 & c_1\lambda_1^2 \\ c_2 & c_2\lambda_2 & c_2\lambda_2^2 \\ c_3 & c_3\lambda_3 & c_3\lambda_3^2 \end{bmatrix} = [\mathbf{0}, \mathbf{0}, \mathbf{0}]$$

即

$$[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3] \begin{bmatrix} c_1 & & \\ & c_2 & \\ & & c_3 \end{bmatrix} \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 \\ 1 & \lambda_2 & \lambda_2^2 \\ 1 & \lambda_3 & \lambda_3^2 \end{bmatrix} = [\mathbf{0}, \mathbf{0}, \mathbf{0}]$$

由于 $\lambda_i$ 互异,  $\begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 \\ 1 & \lambda_2 & \lambda_2^2 \\ 1 & \lambda_3 & \lambda_3^2 \end{bmatrix}$ 可逆。于是,

$$[c_1 \mathbf{x}_1, c_2 \mathbf{x}_2, c_3 \mathbf{x}_3] = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3] \begin{bmatrix} c_1 & & \\ & c_2 & \\ & & c_3 \end{bmatrix} = [\mathbf{0}, \mathbf{0}, \mathbf{0}]$$

由于 $\mathbf{x}_i$ 非零,  $c_i = 0$ . □

**定理1.34.** 如果 $A$ 有 $n$ 个互不相同特征值, 则 $A$ 可对角化。

证明. 设 $A$ 的特征值为 $\lambda_1, \dots, \lambda_n$ , 它们对应的特征向量分别为 $\mathbf{x}_1, \dots, \mathbf{x}_n$ , 由上述命题 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 线性无关。因此,  $A$ 可对角化。 □

**注记1.35.** 判定 $A$ 是否有 $n$ 个互不相同特征值是看 $f_A(\lambda)$ 没有重根, 即是看 $f_A(\lambda)$ 与 $f'_A(\lambda)$ 没有次数 $\geq 1$ 的公因子。

**例1.36.**  $\begin{bmatrix} 5 & 1 & 4 \\ 0 & 0 & 3 \\ 0 & 0 & 9 \end{bmatrix}$ 可对角化, 因为它的特征值为5, 0, 9.

**注记1.37.** (1) 实际中碰到的矩阵从概率上看100%都是有 $n$ 个互异的特征值.

(2)  $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ 是个不可对角化的例子。因为 $A$ 的特征值为2, 重数为2, 如果 $A$ 可对角化, 那么 $A = X 2 I_2 X^{-1} = 2 I_2$ . 我们在下一小节进一步研究矩阵的可对角化的。

特征值有重数的矩阵许多也是可对角化的, 例如 $I_2$ .

1.2.5. 矩阵可对角化的条件 (二). 设 $\lambda_1, \dots, \lambda_k$ 分别为 $A$ 互不相同的特征值, 重数分别为 $n_1, \dots, n_k$ . 我们有

- $f_A(\lambda) = (\lambda_1 - \lambda)^{n_1} \cdots (\lambda_k - \lambda)^{n_k}$ ;
- $n_1 + \cdots + n_k = n$ .

回顾我们称 $n_i$ 为特征值 $\lambda_i$ 对应的代数重数(algebraic multiplicity)。  $d_i := \dim N(A - \lambda_i I_n)$ 称为 $\lambda_i$ 的几何重数 (geometric multiplicity)。  $d_i$ 等于特征值为 $\lambda_i$ 的线性无关的特征向量的个数。

**命题1.38.**  $\forall \lambda_i, d_i \leq n_i$ , 即几何重数  $\leq$  代数重数。

证明. 设  $\mathbf{x}_1, \dots, \mathbf{x}_{d_i}$  为  $N(A - \lambda_i)$  的一组基, 我们将它扩充为  $\mathbb{R}^n$  的一组基  $\mathbf{x}_1, \dots, \mathbf{x}_{d_i}, \mathbf{u}, \dots, \mathbf{v}$ .

$$A[\mathbf{x}_1, \dots, \mathbf{x}_{d_i}, \mathbf{u}, \dots, \mathbf{v}] = [\mathbf{x}_1, \dots, \mathbf{x}_{d_i}, \mathbf{u}, \dots, \mathbf{v}] \begin{bmatrix} \lambda_i I_{d_i} & B \\ 0 & A_1 \end{bmatrix}$$

$$\begin{aligned} f_A(\lambda) &= \det \left( \begin{bmatrix} \lambda_i I_{d_i} & B \\ 0 & A_1 \end{bmatrix} - \lambda I_n \right) \\ &= \begin{vmatrix} (\lambda_i - \lambda) I_{d_i} & B \\ 0 & A_1 - \lambda I_{n-d_i} \end{vmatrix} = (\lambda_i - \lambda)^{d_i} |A_1 - \lambda I_{n-d_i}|. \end{aligned}$$

所以,  $(\lambda_i - \lambda)^{d_i} |f_A(\lambda)| = (\lambda_1 - \lambda)^{n_1} \cdots (\lambda_k - \lambda)^{n_k}$ . 由整除性, 我们得到  $d_i \leq n_i$ .

□

**定理1.39.**  $A$  可对角化当且仅当  $d_i = n_i, \forall i$ , 即的每个特征值都有足够多的线性无关的特征向量 (几何重数 = 代数重数)。

证明. 由命题1.33, 分属于  $A$  的不同特征值对应的特征子空间的向量是线性无关的。所以,  $A$  拥有线性无关的特征向量最多有  $d_1 + d_2 + \cdots + d_k \leq n_1 + \cdots + n_k = n$ .  $A$  可对角化当且仅当  $d_1 + \cdots + d_k = n$ , 当且仅当  $d_i = n_i, \forall i$ .

□

**注记1.40.** (1) 推论说明如果矩阵不可对角化, 那么一定存在一个特征值, 它的几何重数  $<$  代数重数。

(2) 求  $(A - \lambda_i I)\mathbf{x} = 0$  的特殊零解得到  $N(A - \lambda_i I)$  的一组基。

**例1.41.**  $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ . 秩1矩阵  $A = \mathbf{u}\mathbf{v}^T = (u_i v_j)_{1 \leq i, j \leq n}$ . 由于  $A$  的秩为1,  $N(A)$  的维数为  $n - 1$ . 于是0作为特征值代数重数至少为  $n - 1$ .  $\text{trace}(A) = u_1 v_1 + \cdots + u_n v_n = \mathbf{u}^T \mathbf{v}$ . 得到  $\mathbf{u}^T \mathbf{v}$  也是  $A$  的特征值. 我们有  $f_A(\lambda) = (-1)^n \lambda^{n-1} (\lambda - \mathbf{u}^T \mathbf{v})$ .

如果  $\mathbf{u}^T \mathbf{v} = 0$ , 则  $A$  一定不可对角化 (此时  $GM = n - 1 < AM = n$ ); 如果  $\mathbf{u}^T \mathbf{v} \neq 0$ ,  $A$  可对角化 ( $GM = n - 1 = AM$ ).

定义1.42 (若当块(Jordan block)). 形如  $J_n(\lambda_0) = \begin{bmatrix} \lambda_0 & 1 & & \\ & \lambda_0 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_0 \end{bmatrix}$  的  $n$  阶

方阵称为若当块。

例1.43.  $f_{J_n(\lambda_0)}(\lambda) = (\lambda_0 - \lambda)^n$ . 因此,  $J_n(\lambda_0)$  的特征值为  $\lambda_0$ , 其代数

重数为  $n$ .  $\dim N(J_n(\lambda_0) - \lambda_0 I_n) = \dim N\left(\begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}\right) = n -$

$r\left(\begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}\right) = 1$ . 因此, 如果  $n \geq 2$ ,  $J_n(\lambda_0)$  不可对角化.

总结:  $A \in M_{n \times n}(\mathbb{C})$ .  $\lambda_1, \dots, \lambda_k$  为  $A$  互不相同的特征值.

特征值: $\lambda_1$	$\lambda_2 \dots$	$\lambda_k$
特征多项式: $f_A(\lambda) = (\lambda_1 - \lambda)^{n_1}$	$(\lambda_2 - \lambda)^{n_2} \dots$	$(\lambda_k - \lambda)^{n_k}$
$n = n_1$	$+n_2 + \dots +$	$n_k$
代数重数(AM): $n_1$	$n_2 \dots$	$n_k$
几何重数(GM): $d_1$	$d_2 \dots$	$d_k$
$d_i = \dim N(A - \lambda_i I), \quad d_i \leq n_i$		
$\mathbb{C}^n \supseteq N(A - \lambda_1 I) \oplus N(A - \lambda_2 I) \oplus \dots \oplus N(A - \lambda_k I)$		
线性无关向量总数 $d = d_1$	$+d_2 + \dots +$	$d_k$

注意: 上面线性无关向量是指的作为复数上的线性空间的线性无关向量.

$A$  可对角化当且仅当  $A$  有  $n$  个线性无关特征向量, 当且仅当  $d = n$ , 当且仅当  $d_i = n_i$  (即对每个特征值  $\lambda_i$ ,  $GM(\lambda_i) = AM(\lambda_i)$ ).



## 1.2.6. Cayley-Hamilton定理.

**定义1.44.** 设  $f(\lambda) = a_0 + a_1\lambda + \cdots + a_n\lambda^n \in \mathbb{C}[\lambda]$  为  $\mathbb{C}$  上多项式.  $A \in M_{n \times n}(\mathbb{C})$ . 定义

$$f(A) = a_0I_n + a_1A + \cdots + a_nA^n \in M_{n \times n}(\mathbb{C}).$$

**定理1.45.** 设  $f_A(\lambda)$  为  $A$  的特征多项式, 则  $f_A(A) = 0_{n \times n}$ . 即若  $f_A(\lambda) = a_0 + a_1\lambda + \cdots + (-1)^n\lambda^n$ , 则

$$a_0 + a_1A + \cdots + (-1)^nA^n = 0_{n \times n}.$$

证明思路. (1)  $A$  为对角矩阵时, 可直接验证。

(2)  $A$  为可对角化时, 利用相似矩阵有相同的特征多项式。(见习题)

(3) 对一般的  $A$ , 想法是转化成可对角化情形。由命题1.57,  $A$  可上三角化, 即  $A = BUB^{-1}$ .  $U$  对角线上可能有相同的元素。我们找一系列矩阵  $A_k = BU_kB^{-1}$ , 满足  $U_k$  上三角, 除对角线以外其元素与  $U$  元素相同, 对角线元素互不相同, 且  $U_k \rightarrow U, k \rightarrow \infty$ 。由于  $U_k \rightarrow U, k \rightarrow \infty$ , 我们有  $A_k \rightarrow A$  (矩阵的entry的极限),  $f_{A_k}(\lambda) \rightarrow f_A(\lambda)$  (多项式系数的极限)。因此,

$$f_{A_k}(A_k) \rightarrow f_A(A)$$

得到  $f_A(A) = 0_{n \times n}$ .

□

**注记1.46.** 由  $f_A(\lambda) = a_0 + a_1\lambda + \cdots + (-1)^n\lambda^n, a_0 = |A|$ , 我们得到

$$a_0I_n + A(a_1 + a_2A + \cdots + (-1)^nA^{n-1}) = 0_{n \times n}.$$

因此,  $|A|I_n = A(-a_1 - a_2A - \cdots - (-1)^nA^{n-1})$ . 因此,  $A$  的伴随矩阵  $C^T = (-a_1 - a_2A - \cdots - (-1)^nA^{n-1})$ . 如果  $A$  可逆,  $|A| \neq 0$ ,  $A^{-1} = \frac{C^T}{|A|}$ .

1.2.7.  $AB$  的特征值. 假定  $A, B$  都是可对角化矩阵。一般来说,  $AB$  的特征值不等于  $A$  与  $B$  的特征值相乘 (见习题)。如果假设  $\mathbf{x}$  同时为  $A, B$  特征值分别为  $\lambda_A, \lambda_B$  的特征向量, 则

$$AB\mathbf{x} = A(\lambda_B\mathbf{x}) = \lambda_B A\mathbf{x} = \lambda_B \lambda_A \mathbf{x}.$$

此时 $\mathbf{x}$ 也为 $AB$ 的特征向量, 特征值为 $\lambda_A\lambda_B$ .

**定义1.47.** 设 $A, B$ 为可对角化矩阵. 称 $A, B$ 可同时对角化, 如果有可逆矩阵 $X$ 使得

$$AX = X\Lambda(A), \quad BX = X\Lambda(B)$$

其中 $\Lambda(A), \Lambda(B)$ 为对角矩阵.

将 $X$ 写为列向量形式 $X = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ .  $A, B$ 可由 $X$ 同时对角化即是说 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 为 $A, B$ 公共的线性无关的特征向量.

**命题1.48.** 设 $A, B$ 为可对角化矩阵. 如果 $A, B$ 可同时对角化, 则 $AB$ 也为可对角化矩阵, 且 $AB = BA$ .

证明. 设 $X = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ 满足

$$AX = X\Lambda(A), \quad BX = X\Lambda(B)$$

那么,  $ABX = AX\Lambda(B) = X\Lambda(A)\Lambda(B)$ . 于是 $AB$ 也可由 $X$ 对角化.

$$AB = X\Lambda(A)\Lambda(B)X^{-1} = X\Lambda(B)\Lambda(A)X^{-1} = (X\Lambda(B)X^{-1})(X\Lambda(A)X^{-1}) = BA.$$

□

**定理1.49.** 设 $A, B$ 为可对角化矩阵.  $A, B$ 可同时对角化当且仅当 $AB = BA$ .

证明. 上述命题已经证明: 如果 $A, B$ 可同时对角化, 则 $AB$ 可对角化且 $AB = BA$ .

反过来, 我们只证明命题一个简单情况:

假定 $B$ 有 $n$ 个互不相同的特征值.

一般情形较为复杂, 见习题课材料.

根据假设 $BX = X\Lambda(B)$ . 令 $A' := X^{-1}AX = (a'_{ij})_{1 \leq i, j \leq n}$ . 断言 $A'$ 是对角矩阵. 因此,  $A$ 可由 $X$ 对角化.

满足

$$A'\Lambda(B) = X^{-1}AX\Lambda(B) = X^{-1}ABX = X^{-1}BAX = (X^{-1}BX)(X^{-1}AX) = \Lambda(B)A'.$$

所以,

$$\lambda_j(B)a'_{ij} = \lambda_i(B)a'_{ij}$$

由于对  $i \neq j$ ,  $\lambda_j(B) \neq \lambda_i(B)$ , 我们得到  $a'_{ij} = 0$  如果  $i \neq j$ . 因此,  $A'$  为对角矩阵, 断言成立.  $\square$

1.2.8. 矩阵的相似. 将可对角化的概念推广到矩阵的相似 (similar)。

**定义1.50.** 称  $n$  阶矩阵  $A, C$  相似(similar), 如果有可逆矩阵  $B$  ( $B$  代表 basis) 使得

$$A = BCB^{-1}.$$

**命题1.51.** 矩阵的相似是一个等价关系, 即:

- (1) 自反性:  $A$  与自身相似; ( $A = I_n^{-1}AI_n$ )
- (2) 对称性: 若  $A$  与  $B$  相似, 则  $B$  与  $A$  相似; ( $A = X^{-1}BX$ ,  $B = XAX^{-1}$ .)
- (3) 传递性: 若  $A$  与  $B$  相似,  $B$  与  $C$  相似, 则  $A$  与  $C$  相似. ( $A = X_1^{-1}BX_1$ ,  $B = X_2^{-1}CX_2$ , 则  $A = (X_1X_2)^{-1}C(X_1X_2)$ .)

与  $C$  相似的矩阵  $\{BCB^{-1} \mid B \text{ 可逆}\}$  构成一类, 称为  $C$  的等价类.  $I_n$  的等价类只有  $I_n$  自己。

**命题1.52.** 相似矩阵有相同的特征多项式。

证明. 设  $A = BCB^{-1}$ .  $|A - \lambda I| = |BCB^{-1} - \lambda I| = |B(C - \lambda I)B^{-1}| = |C - \lambda I|$ .  $\square$

**注记1.53.** 反之, 如果  $A, C$  有相同的特征多项式, 但  $A, C$  可能并不相似, 例如,  $\begin{bmatrix} 1 & \\ & 1 \end{bmatrix}$  与  $\begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}$ .

**推论1.54.** 相似矩阵有相同的特征值。

**注记1.55.**  $A = BCB^{-1}$ . 如果  $\mathbf{x}$  为  $C$  的特征值为  $\lambda$  的特征向量, 那么  $B\mathbf{x}$  为  $A$  的特征值为  $\lambda$  的特征向量。

**定理1.56** (Jordan canonical form). (1)  $A \in M_{n \times n}(\mathbb{R})$ 。则有可逆矩阵  $B \in M_{n \times n}(\mathbb{C})$  使得

$$B^{-1}AB = \begin{bmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{n_s}(\lambda_s) \end{bmatrix}$$

这里  $\lambda_i$  可能相同.  $\begin{bmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{n_s}(\lambda_s) \end{bmatrix}$  称为  $A$  的若当标准型 (Jordan canonical form) .

(2) 两个矩阵相似当且仅当它们具有相同的若当标准型。

我们应理解为: 若当标准型 (Jordan form) 是在  $A$  的相似等价类中选取一个最简单的矩阵代表形式。在线性映射一章我们将回到这个问题。

若当标准型的证明有兴趣的同学可参见中文教材阅读5.4.7. 在这里, 我们证明一个课堂上省略没讲的简单版本: 每个等价类里有一个上三角矩阵。

**命题1.57.** 设  $A \in M_{n \times n}(\mathbb{R})$ , 则存在  $B \in M_{n \times n}(\mathbb{C})$  可逆使得  $B^{-1}AB$  为上三角复方阵, 对角线元素为  $A$  的复特征值。

证明.  $A$  总存在一个复特征值, 设为  $\lambda_1$ , 设  $\mathbf{x}_1$  为对应的一个特征向量。将  $\mathbf{x}_1$  扩充为  $\mathbb{C}^n$  一组基  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ . 记矩阵

$$X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n].$$

那么

$$\begin{aligned} AX &= [A\mathbf{x}_1, A\mathbf{x}_2, \dots, A\mathbf{x}_n] = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \begin{bmatrix} \lambda_1 & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & \cdots & \vdots \\ 0 & * & \cdots & * \end{bmatrix} \\ &= [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \begin{bmatrix} \lambda_1 & \mathbf{a}^T \\ & A_1 \end{bmatrix} = X \begin{bmatrix} \lambda_1 & \mathbf{a}^T \\ & A_1 \end{bmatrix} \end{aligned}$$

因此,

$$X^{-1}AX = \begin{bmatrix} \lambda_1 & \mathbf{a}^T \\ & A_1 \end{bmatrix}.$$

由归纳假设,  $A_1$  为  $n-1$  阶方阵, 它相似于上三角, 即有  $B_1$  使得  $B_1^{-1}A_1B_1 = U_1$ . 令

$$B = X \begin{bmatrix} 1 & \\ & B_1 \end{bmatrix},$$

$$\begin{aligned} B^{-1}AB &= \begin{bmatrix} 1 & \\ & B_1 \end{bmatrix}^{-1} X^{-1}AX \begin{bmatrix} 1 & \\ & B_1 \end{bmatrix} = \begin{bmatrix} 1 & \\ & B_1 \end{bmatrix}^{-1} \begin{bmatrix} \lambda_1 & \mathbf{a}^T \\ & A_1 \end{bmatrix} \begin{bmatrix} 1 & \\ & B_1 \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \mathbf{a}^T B_1 \\ & B_1^{-1}A_1B_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{a}^T B_1 \\ & U_1 \end{bmatrix} \end{aligned}$$

为上三角矩阵。  $\square$

### 1.3. 对称矩阵.

G. Strang: “It is no exaggeration to say that symmetric matrices  $S$  are the most important matrices the world will ever see – in the theory of linear algebra and also in the applications.”

1.3.1. 对称矩阵的谱定理. 令  $S \in M_{n \times n}(R)$  为对称矩阵, 即  $S^T = S$ . 我们考虑  $S$  的特征值, 特征向量, 对角化问题. 我们会看到实对称矩阵可由实方阵对角化。

**定理1.58** (对称矩阵谱定理-特征值(Spectrum Theorem - Part I)).

如果  $\lambda$  为实对称矩阵  $S$  的特征值, 那么  $\lambda \in \mathbb{R}$ . 有实向量  $\mathbf{x} \in \mathbb{R}^n$  为  $S$  的特征值为  $\lambda \in \mathbb{R}$  的特征向量.

证明. 设

$$S\mathbf{x} = \lambda\mathbf{x}, \quad \lambda \in \mathbb{C}, \quad \mathbf{x} \in \mathbb{C}^n \quad (1)$$

取复共轭, 由于  $\overline{S} = S$ ,

$$S\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$$

取转置得到

$$\overline{\mathbf{x}}^T S = \overline{\lambda} \overline{\mathbf{x}}^T \quad (2)$$

对(1)左乘 $\bar{\mathbf{x}}^T$ ,对(2)右乘 $\mathbf{x}$ , 得到

$$\lambda \bar{\mathbf{x}}^T \mathbf{x} = \bar{\mathbf{x}}^T S \mathbf{x} = \bar{\lambda} \bar{\mathbf{x}}^T \mathbf{x}.$$

由于 $\bar{\mathbf{x}}^T \mathbf{x} = |x_1|^2 + \cdots + |x_n|^2 > 0$ , 我们有 $\bar{\lambda} = \lambda$ , 即 $\lambda \in \mathbb{R}$ .

解实系数方程 $(S - \lambda I)\mathbf{x} = 0$ , 可知一定有实向量解 $\mathbf{x} \in \mathbb{R}^n$ . □

**注记1.59.** 事实上, 我们证明对满足 $\bar{S}^T = S$ 的矩阵 $S$ , 其特征值一定是实数。

**命题1.60** (反对称矩阵). 如果 $\lambda$ 为反对称矩阵的特征值, 那么 $\lambda \in i\mathbb{R}$ .

证明. 留为习题. □

**命题1.61** (正交矩阵). 如果 $\lambda$ 为正交矩阵的特征值, 那么 $|\lambda| = 1$ .

证明. 留为习题. □

**注记1.62.** 对称矩阵类比实数, 反对称矩阵类比纯虚数, 正交矩阵类比模长为1的复数。

**定理1.63** (对称矩阵谱定理-对角化(Spectrum Theorem - Part II)).

存在正交矩阵 $Q$ (即 $Q \in M_{n \times n}(\mathbb{R})$ ,  $Q^T Q = I$ )使得 $S = Q \Lambda Q^{-1} = Q \Lambda Q^T$ . 即 $S$ 有 $n$ 个相互正交的单位特征向量。

记 $Q = [\mathbf{q}_1, \dots, \mathbf{q}_n]$ ,  $S = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T$ , 称为 $S$ 的谱分解。

证明.  $S \in M_{n \times n}(\mathbb{R})$ , 对 $n$ 用数学归纳法。

假设命题对 $n - 1$ 阶实对称矩阵成立。令 $S$ 为 $n$ 阶实对称, 设 $S$ 的一个特征值 $\lambda_1$ . 令 $\mathbf{q}_1 \in \mathbb{R}^n$ 为特征值为 $\lambda_1$ 的一个特征向量(由谱定理的第一部分 $\mathbf{q}_1$ 可取为实向量)且 $\|\mathbf{q}_1\| = 1$ . 令 $\mathbf{q}_2, \dots, \mathbf{q}_n$ 为 $N(\mathbf{q}_1^T) = \mathbb{R}\mathbf{q}_1^\perp$ 的一组标准正交基.  $\mathbf{q}_1, \dots, \mathbf{q}_n$ 构成 $\mathbb{R}^n$ 的一组标准正交基. 令 $Q = [\mathbf{q}_1, \dots, \mathbf{q}_n]$ ,

$$SQ = S[\mathbf{q}_1, \dots, \mathbf{q}_n] = [\mathbf{q}_1, \dots, \mathbf{q}_n] \begin{bmatrix} \lambda_1 & \mathbf{b}^T \\ 0 & S_1 \end{bmatrix}$$

所以,

$$Q^T S Q = \begin{bmatrix} \lambda_1 & \mathbf{b}^T \\ \mathbf{0} & S_1 \end{bmatrix}.$$

由于 $Q^T S Q$ 为对称矩阵,  $\mathbf{b} = \mathbf{0}$ 且 $S_1$ 为对称矩阵。于是,

$$Q^T S Q = \begin{bmatrix} \lambda_1 & 0^T \\ \mathbf{0} & S_1 \end{bmatrix}.$$

根据归纳假设, 有正交矩阵 $Q_1$ 使得

$$Q_1^T S_1 Q_1 = \begin{bmatrix} \lambda_2 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

令 $Q^{new} = Q \begin{bmatrix} 1 & \\ & Q_1 \end{bmatrix}$ 。可以直接验证 $Q^{new}$ 为正交矩阵。我们有 $(Q^{new})^T S Q^{new} = \Lambda$ . □

**注记1.64.** 由上面的谱定理, 实对称矩阵的对角化可在 $M_{n \times n}(\mathbb{R})$ 中用正交矩阵进行对角化。我们也称正交对角化。

更广泛的一类常用的可对角化矩阵是正规矩阵 $N$ , 请见6.4 (31)题。

### 1.3.2. 实对称矩阵的线性代数基本定理.

**命题1.65.**  $S$ 实对称矩阵, 根据谱定理有 $S = Q \Lambda Q^T$ ,  $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$ ,  $\lambda_i \neq 0$ . 则

(1)  $r = r(S)$ .

(2)  $\mathbf{q}_1, \dots, \mathbf{q}_r$ 构成 $C(S)(= C(S^T))$ 行空间和列空间的一组标准正交基;  $\mathbf{q}_{r+1}, \dots, \mathbf{q}_n$ 构成零空间和左零空间的一组基。

证明. 由于 $Q, Q^T$ 可逆,  $r(\Lambda) = r(Q \Lambda Q^T) = r(S)$ . 所以,  $r = r(S)$ 。

由于 $1 \leq i \leq r$ ,  $S \mathbf{q}_i = \lambda_i \mathbf{q}_i$ ,  $\lambda_i \neq 0$ , 我们有 $\mathbf{q}_i \in C(S)$ . 因此,  $\text{Span}(\mathbf{q}_1, \dots, \mathbf{q}_r) \subseteq C(S)$ . 由于两者有相同的维数, 它们相等。 □

### 1.3.3. 如何对角化对称矩阵. 对于实对称矩阵我们考虑正交对角化。

**命题1.66.** 不同特征值对应的特征子空间相互正交。即如果 $\lambda_1, \lambda_2$ 为对称矩阵 $S$ 两个不同的特征值,  $\mathbf{x}_1, \mathbf{x}_2$ 分别为特征值为 $\lambda_1, \lambda_2$ 的特征向量。则 $\mathbf{x}_1 \perp \mathbf{x}_2$ .

证明. 由 $S\mathbf{x}_1 = \lambda_1\mathbf{x}_1$ , 我们有 $\mathbf{x}_2^T S\mathbf{x}_1 = \lambda_1\mathbf{x}_2^T \mathbf{x}_1$ ; 由 $S\mathbf{x}_2 = \lambda_2\mathbf{x}_2$ , 我们有 $\mathbf{x}_1^T S\mathbf{x}_2 = \lambda_2\mathbf{x}_1^T \mathbf{x}_2$ . 由于 $S$ 对称且 $\mathbf{x}_1^T S\mathbf{x}_2$ 为一个数,

$$\mathbf{x}_1^T S\mathbf{x}_2 = (\mathbf{x}_1^T S\mathbf{x}_2)^T = \mathbf{x}_2^T S^T \mathbf{x}_1 = \mathbf{x}_2^T S\mathbf{x}_1.$$

我们得到

$$\lambda_2\mathbf{x}_1^T \mathbf{x}_2 = \lambda_1\mathbf{x}_2^T \mathbf{x}_1.$$

由于 $\mathbf{x}_1^T \mathbf{x}_2 = \mathbf{x}_2^T \mathbf{x}_1$ 且 $\lambda_1 \neq \lambda_2$ , 我们有 $\mathbf{x}_2^T \mathbf{x}_1 = 0$ . □

根据上述命题, 对称矩阵 $S$ 的对角化可按如下步骤进行:

(1) 求 $S$ 的特征多项式与所以特征值 $\lambda_1, \dots, \lambda_k$ (互不相同), 代数重数分别为 $n_1, \dots, n_k$ ;

(2) 对每个 $\lambda_i$ , 解方程 $(S - \lambda_i I)\mathbf{x} = 0$ . 由于我们已知 $S$ 在 $M_{n \times n}(\mathbb{R})$ 中可对角化, 对每个 $\lambda_i$ , 几何重数等于代数重数。所以, 方程 $(S - \lambda_i I)\mathbf{x} = 0$ 的解空间维数为 $n_i$ , 即我们一定得到 $n_i$ 个线性无关的解, 我们记为 $\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}$ ;

(3) 由上述命题, 我们已经知道如果 $i \neq j$ ,  $N(S - \lambda_i I_n) \perp N(S - \lambda_j I_n)$ . 我们利用Gram-Schmidt正交化将 $\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}$ 化为标准正交的向量 $\mathbf{q}_{i1}, \dots, \mathbf{q}_{in_i}$ ;

(4) 记 $Q = [\mathbf{q}_{11}, \dots, \mathbf{q}_{1n_1}, \dots, \mathbf{q}_{k1}, \dots, \mathbf{q}_{kn_k}]$ , 我们有

$$SQ = Q\Lambda,$$

其中 $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_1, \dots, \lambda_k, \dots, \lambda_k)$ .

**例1.67.**  $S = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{bmatrix}.$

特征多项式 $f_S(\lambda) = |S - \lambda I| = (1 - \lambda)^2(2 + \lambda)$ . 所以, 特征值为1(重数2),  $-2$ (重数1).

解方程 $(S + 2I)\mathbf{x} = 0$ , 得到 $\mathbf{x}_1 = [1, -1, -1]^T$ ;  $\mathbf{q}_1 = [1/\sqrt{3}, -1/\sqrt{3}, -1/\sqrt{3}]^T$ ;

解方程 $(S - I)\mathbf{x} = 0$ , 得到 $\mathbf{x}_2 = [1, 1, 0]^T$ ,  $\mathbf{x}_3 = [1, 0, 1]^T$ ;



对  $\mathbf{x}_2, \mathbf{x}_3$  进行 *Gram-Schmidt* 正交化得到  $\mathbf{q}_2 = [1/\sqrt{2}, 1/\sqrt{2}, 0]^T$ .

$$\mathbf{x}'_3 = \mathbf{x}_3 - \mathbf{q}_2 \mathbf{q}_2^T \mathbf{x}_3 = [1/2, -1/2, 1]^T$$

$$\mathbf{q}_3 = \frac{\mathbf{x}'_3}{\|\mathbf{x}'_3\|} = [1/\sqrt{6}, -1/\sqrt{6}, \sqrt{\frac{2}{3}}]^T.$$

$$\text{因此, } Q = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3] = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ -1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ -1/\sqrt{3} & 0 & \sqrt{\frac{2}{3}} \end{bmatrix}.$$

$$S = Q \begin{bmatrix} -2 & & \\ & 1 & \\ & & 1 \end{bmatrix} Q^T.$$

1.3.4. 对称矩阵的主元与特征值(阅读). 设  $S$  为可逆对称矩阵, 利用LU分解  $S = LDL^T$  与谱定理  $S = Q\Lambda Q^T$ , 我们建立主元与特征值的关系。假定  $S$  在高斯消元中无需换行操作, 由引理1.75 这等价于  $S$  的所有顺序主子式非零。

**命题1.68.**  $S$  正主元的数目等于正特征值的数目; 负主元的数目等于负特征值的数目。

证明. 从LU分解出发  $S = LDL^T$ ,  $L$  为对角线元素均为1的下三角矩阵,  $D$  为对角矩阵, 特别地,  $S$  的主元即为  $D$  对角线上元素。令  $L(t) = (1-t)L + tI_n$ ,  $t \in [0, 1]$ . 令  $S(t) = L(t)DL(t)^T$ . 我们有  $S(t)$  均为对称矩阵且都是可逆矩阵。

我们关心在  $t$  从0到1整个变换过程中, 矩阵  $S(t)$  特征值的变化。矩阵的特征值也随时间连续变化, 但是在整个变化过程中特征值符号不会改变, 否则会有某个时刻有0作为特征值, 与  $S(t)$  一直保持可逆矛盾。由于  $t = 0$  时刻, 矩阵的特征值为  $S$  的特征值, 而  $t = 1$  时刻  $S(1) = D$ , 其特征值就是  $S$  的主元。所以,  $S$  正主元的数目等于正特征值的数目; 负主元的数目等于负特征值的数目。□

**注记1.69.** 可由这个命题证明二次型中的惯性定理。

1.4. 正定对称矩阵. 对称矩阵类比于实数, 而正定对称矩阵类比于正实数, 它在应用数学中有广泛应用。

**例1.70.** 考虑 $\mathbb{R}^3$ 中二次曲面 $F(x, y) = ax^2 + 2bxy + cy^2$ 。  $F(0, 0) = 0$ ,  $\frac{\partial F}{\partial x}|_{(0,0)} = \frac{\partial F}{\partial y}|_{(0,0)} = 0$ .  $(0, 0)$ 称为驻点。问题：什么条件保证 $F(x, y)$ 在 $(0, 0)$ 处取得最小值？

将 $F(x, y)$ 写成矩阵形式

$$F(x, y) = [x, y] \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{x}^T S \mathbf{x}$$

其中,  $S = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ 为对称矩阵。 $F(x, y)$ 在 $(0, 0)$ 处取最小即是要求

$$\mathbf{x}^T S \mathbf{x} \geq 0, \quad \forall \mathbf{x} \in \mathbb{R}^2.$$

满足这个条件的 $S$ 称为半正定矩阵。 $(0, 0)$ 是唯一的最小值点当且仅当

$$\mathbf{x}^T S \mathbf{x} > 0, \quad \forall \mathbf{x} \neq \mathbf{0} \in \mathbb{R}^2.$$

满足条件的 $S$ 称为正定。

事实上,  $S$ 为 $F$ 在 $(0, 0)$ 点处的二阶导数给出的矩阵

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \frac{\partial^2 F}{\partial x \partial x} & \frac{\partial^2 F}{\partial x \partial y} \\ \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial y \partial y} \end{bmatrix}$$

$S$ 的正定性推广了一元函数时的极值条件 $f''(x_0) > 0$ 。

#### 1.4.1. 正定对称矩阵的定义.

**定义1.71.** 对称矩阵 $S \in M_{n \times n}(\mathbb{R})$ 称为正定的(positive definite) $\mathbf{x}^T S \mathbf{x} > 0$ 对所有非零向量 $\mathbf{x} \in \mathbb{R}^n$ .  $S$ 称为半正定 (positive semi-definite),  $\mathbf{x}^T S \mathbf{x} \geq 0$ 对所有 $\mathbf{x} \in \mathbb{R}^n$ .

**注记1.72.** 这个定义也称为正定矩阵的能量判定法 (energy test)。

**例1.73.** (1) 如果 $S = D = \text{Diag}(d_1, \dots, d_n)$ 为对角矩阵.  $\mathbf{x}^T D \mathbf{x} = d_1 x_1^2 + \dots + d_n x_n^2$ .  $D$ 为正定当且仅当 $d_1, d_2, \dots, d_n > 0$ .

(2) 如果 $S$ 为正定矩阵,  $A$ 可逆, 那么 $A^T S A$ 为正定矩阵。这是因为设 $\mathbf{x} \neq \mathbf{0}$ .  $\mathbf{x}^T (A^T S A) \mathbf{x} = (A\mathbf{x})^T S (A\mathbf{x})$ ,  $A$ 可逆,  $A\mathbf{x} \neq \mathbf{0}$ . 由 $S$ 的正定性得到 $(A\mathbf{x})^T S (A\mathbf{x}) > 0$ .

(3) 设  $A \in M_{m \times n}(\mathbb{R})$ . 从  $A$  出发可以得到对称矩阵  $S = A^T A$ . 这个对称矩阵对于  $A$  的研究有重要作用(参见下一章奇异值分解内容). 由于  $\mathbf{x}^T S \mathbf{x} = \mathbf{x}^T A^T A \mathbf{x} = \|A \mathbf{x}\|^2 \geq 0$ . 于是,

$S = A^T A$  是半正定的实对称矩阵.  $S$  正定当且仅当  $A$  列满秩.

1.4.2. 正定对称矩阵的判别. 我们讨论实对称正定性的几个判别方法. 回顾主子式, 顺序主子式的定义:

**定义1.74.**  $A = (a_{ij})_{1 \leq i, j \leq n} \in M_{n \times n}(\mathbb{R})$ .

$A_k = (a_{ij})_{1 \leq i, j \leq k}$  为  $A$  的左上角  $k$  阶子方阵, 称为  $A$  的顺序主子阵.  $|A_k|$  称为  $A$  的  $k$  阶顺序主子式.

**引理1.75.**  $A$  在高斯消元中无需进行行对换操作 (这等价于  $A$  有  $LU$  分解) 当且仅当  $A$  的所有顺序主子式  $|A_1|, \dots, |A_n|$  非零.

证明. 假设  $A$  的所有顺序主子式  $|A_1|, \dots, |A_n|$  非零, 特别地,  $a_{11} \neq 0$ , 为主元. 利用高斯消元得到

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22}^{(1)} & \cdots & a_{2n} \\ \vdots & * & * & * \\ 0 & * & * & * \end{bmatrix}$$

由于这样的行变换不改变行列式,  $|A_2| = a_{11}a_{22}^{(1)}$ , 所以

$$a_{22}^{(1)} = \frac{|A_2|}{|A_1|} \neq 0.$$

因此,  $a_{22}^{(1)}$  为主元. 用  $a_{22}^{(1)}$  进行行消去,

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22}^{(1)} & \cdots & a_{2n} \\ 0 & 0 & a_{33}^{(1)} & * \\ \vdots & 0 & * & * \\ 0 & 0 & * & * \end{bmatrix}$$

所以,  $|A_3| = a_{11}a_{22}^{(1)}a_{33}^{(1)}$ , 得到

$$a_{33}^{(1)} = \frac{|A_3|}{|A_2|} \neq 0.$$

以此类推, 我们得到高斯消元时无需行交换。

反过来,  $a_{11} = |A_1| \neq 0$ , 可用 $a_{11}$ 进行高斯消元。 $|A_2| \neq 0$ 保证 $a_{22}^{(1)} \neq 0$ , 于是 $a_{22}^{(1)}$ 是第二行的主元, 可由它进行高斯消元。以此类推,  $|A_1|, \dots, |A_n| \neq 0$ 可得到高斯消元无需行对换。  $\square$

**定理1.76.** 对于实对称矩阵 $S$ , 以下陈述等价:

(1)  $S$ 正定, 即 $\mathbf{x}^T S \mathbf{x} > 0$ 对所有非零向量 $\mathbf{x} \in \mathbb{R}^n$ . (能量判定)

(2)  $S$ 的所有特征值 $> 0$ . (特征值判定)

(3) 存在可逆矩阵 $A$ 使得 $S = A^T A$ .

(4)  $S$ 有 $LDL^T$ 分解, 且 $D$ 的对角元均为正数。回顾:  $L$ 为对角线元素均为1的下三角矩阵,  $D$ 为对角矩阵, 其对角元素为 $S$ 的主元。这条也可称为主元判定。

(5)  $S$ 的 $n$ 个顺序主子式均 $> 0$ . (顺序主子式判定)

证明. (1) $\Rightarrow$ (2): 设 $\lambda$ 是特征值,  $\mathbf{x}$ 为 $\lambda$ 对应的一个特征向量, 则 $\mathbf{x}^T S \mathbf{x} = \lambda \mathbf{x}^T \mathbf{x}$ . 于是

$$\lambda = \frac{\mathbf{x}^T S \mathbf{x}}{\mathbf{x}^T \mathbf{x}} > 0.$$

$$(2) \Rightarrow (3): S = Q \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} Q^T, \lambda_i > 0. \text{ 令 } A = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix} Q^T,$$

我们有

$$S = A^T A.$$

(3) $\Rightarrow$ (4): 令 $S = A^T A$ . 考虑 $A$ 的QR分解:  $A = QR$ .  $S = R^T Q^T Q R = R^T R$ . 将 $R^T$ 写为 $LD_1$ ,  $D_1$ 为可逆对角矩阵. 则 $A = LD_1 D_1^T L^T = LDL^T$ ,  $D = D_1^2$ . 因此,  $D$ 的对角元 $> 0$ .

$$(4) \Rightarrow (5): \text{假设 } S = LDL^T, D \text{ 的对角元 } > 0. \text{ 记 } L = \begin{bmatrix} L_k & \\ C & L' \end{bmatrix},$$

$$D = \begin{bmatrix} D_k & \\ & D' \end{bmatrix}.$$

$$S = \begin{bmatrix} L_k & \\ C & L' \end{bmatrix} \begin{bmatrix} D_k & \\ & D' \end{bmatrix} \begin{bmatrix} L_k^T & C^T \\ & L'^T \end{bmatrix} = \begin{bmatrix} L_k D_k L_k^T & L_k D_k C^T \\ C D_k L_k^T & C D_k C^T + L' D' L'^T \end{bmatrix}.$$

所以,  $S_k = L_k D_k L_k^T$ . 因此,  $|S_k| = |D_k| > 0$ . 实际上, 由于  $D_k$  正定, 根据例1.73,  $S_k = L_k D_k L_k^T$  也是正定矩阵.

(5)  $\Rightarrow$  (1): 根据引理,  $S$  有  $LU$  分解, 由于  $S$  对称,

$$S = LDL^T = \begin{bmatrix} L_k & \\ C & L' \end{bmatrix} \begin{bmatrix} D_k & \\ & D' \end{bmatrix} \begin{bmatrix} L_k^T & C^T \\ & L'^T \end{bmatrix} = \begin{bmatrix} L_k D_k L_k^T & L_k D_k C^T \\ C D_k L_k^T & C D_k C^T + L' D' L'^T \end{bmatrix}$$

$S$  的  $k$  阶顺序主子式  $|S_k| = |D_k| > 0, 1 \leq k \leq n$ . 由此得到  $D$  的对角元均  $> 0$ , 于是  $D$  是正定矩阵.  $S = LDL^T$  也是正定矩阵.

□

**注记1.77.** 对  $\mathbf{x} \in \mathbb{R}^n$ ,  $\frac{\mathbf{x}^T S \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$  称为  $\mathbf{x}$  对于  $S$  的 Rayleigh 商 (Rayleigh quotient).

**注记1.78.** 如果  $S$  正定矩阵, 那么  $S$  有两个分解

$$S = LDL^T = Q\Lambda Q^T,$$

$D, \Lambda$  均为对角矩阵且对角线元素均大于零。这个等式连接了这门课程两个重要概念: 主元和特征值!

另外,  $S$  有多种分解成  $A^T A$  的形式:

- $S = (\sqrt{\Lambda} Q^T)^T \sqrt{\Lambda} Q^T$ ;
- $S = (Q \sqrt{\Lambda} Q^T)^T (Q \sqrt{\Lambda} Q^T) = (Q \sqrt{\Lambda} Q^T)(Q \sqrt{\Lambda} Q^T)$ ; 注意  $S_1 = Q \sqrt{\Lambda} Q^T$  也是正定矩阵.  $S = S_1^2$ , 于是, 正定矩阵可写成正定矩阵的平方.
- $S = (\sqrt{D} L^T)^T \sqrt{D} L^T$ . (Cholesky factorization).

**例1.79.**  $S = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ . 何时正定?

(1) 利用特征值判别:  $S$  正定说明其两个特征值  $\lambda_1, \lambda_2$  均大于 0. 这等价于  $\text{trace}(S) = a + c = \lambda_1 + \lambda_2 > 0$ ,  $\det S = ac - b^2 = \lambda_1 \lambda_2 > 0$ .

(2) 利用主元判别:  $a > 0$ ,  $\frac{ac-b^2}{a} > 0$ .

(3) 利用主子式判别,  $|S_1| = a > 0$ ,  $|S_2| = ac - b^2 > 0$ .

一般的,  $S$  正定说明其对角线元素  $s_{ii} = \mathbf{e}_i^T S \mathbf{e}_i > 0$ . 其次, 对角线以外的元素要被对角线元素“控制”。

例1.80.  $S = \begin{bmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{bmatrix}$  何时正定?

利用主子式判别:  $\begin{vmatrix} 1 & a \\ a & 1 \end{vmatrix} > 0$ , 得到  $1 - a^2 > 0$ ;  $|S| > 0$ , 得到  $1 + 2abc > a^2 + b^2 + c^2$ . 所以, 两个条件  $1 - a^2 > 0$ ,  $1 + 2abc > a^2 + b^2 + c^2$  保证  $S$  正定。

命题1.81. 如果  $S, T$  为正定对称矩阵, 那么  $S + T$  也为正定对称矩阵。

证明. 对非零向量  $\mathbf{x}$ ,  $\mathbf{x}^T S \mathbf{x} > 0$ ,  $\mathbf{x}^T T \mathbf{x} > 0$ , 得到  $\mathbf{x}^T (S + T) \mathbf{x} > 0$ .  $\square$

#### 1.4.3. 半正定对称矩阵的判别.

定义1.82.  $A = (a_{ij})_{1 \leq i, j \leq n} \in M_{n \times n}(\mathbb{R})$ .

$A$  一个  $k$  阶主子式是指去掉  $i_1, \dots, i_{n-k}$  行与  $i_1, \dots, i_{n-k}$  列剩下方阵的行列式。

对于半正定对称矩阵, 我们有下列等价判别:

- (1)  $S$  的特征值均  $\geq 0$ ;
- (2)  $S$  所有主元非负;
- (3)  $S$  的所有主子式均  $\geq 0$ ; 这里, 一个主子式是指去掉  $i_1, \dots, i_k$  行与  $i_1, \dots, i_k$  列剩下方阵的行列式。
- (4)  $S = A^T A$ ;
- (5)  $\mathbf{x}^T S \mathbf{x} \geq 0$  对所有非零向量  $\mathbf{x} \in \mathbb{R}^n$ .

注记1.83. 条件(3)对于正定和半正定情况的区别: 在半正定下需要考虑所有主子式, 而不仅仅是顺序主子式。

例如,  $S = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$  的顺序主子式均大于等于零, 但是  $[0, 1]S[0, 1]^T = -1 < 0$ .

#### 1.4.4. Rayleigh商.

**定义1.84.** 给定实(对称)矩阵  $S$  与非零向量  $\mathbf{x} \in \mathbb{R}^n$ , 实数  $\frac{\mathbf{x}^T S \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$  称为  $\mathbf{x}$  对于  $S$  的 Rayleigh 商(Rayleigh quotient).

通过对  $Q^T S Q = \Lambda$  左乘适当的置换矩阵  $P^T$ , 右乘  $P$ , 我们可以将对角矩阵调整为  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

**命题1.85.** 设对称矩阵  $S$  的特征值为  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ .  $S = Q \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} Q^T$ ,

$Q = [\mathbf{q}_1, \dots, \mathbf{q}_n]$ . 则,

$$\lambda_1 = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T S \mathbf{x}}{\mathbf{x}^T \mathbf{x}}, \quad \lambda_i = \max_{\mathbf{x} \neq \mathbf{0}, \mathbf{q}_1^T \mathbf{x} = \cdots = \mathbf{q}_{i-1}^T \mathbf{x} = 0} \frac{\mathbf{x}^T S \mathbf{x}}{\mathbf{x}^T \mathbf{x}}, \quad 2 \leq i \leq n.$$

证明. 记  $S = Q \Lambda Q^T$  为  $S$  的谱分解. 则

$$\begin{aligned} \max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T S \mathbf{x}}{\mathbf{x}^T \mathbf{x}} &= \max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T Q \Lambda Q^T \mathbf{x}}{\mathbf{x}^T Q Q^T \mathbf{x}} = \max_{\mathbf{y} = Q^T \mathbf{x} \neq \mathbf{0}} \frac{\mathbf{y}^T \Lambda \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \\ &= \max_{\mathbf{y} \neq \mathbf{0}} \frac{\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2}{y_1^2 + \cdots + y_n^2} \leq \frac{\lambda_1 (y_1^2 + \cdots + y_n^2)}{y_1^2 + \cdots + y_n^2} = \lambda_1. \end{aligned}$$

令  $\mathbf{y} = [1, 0, \dots, 0]^T$  可取得等号。

$$\begin{aligned} \max_{\mathbf{x} \neq \mathbf{0}, \mathbf{q}_1^T \mathbf{x} = \cdots = \mathbf{q}_{i-1}^T \mathbf{x} = 0} \frac{\mathbf{x}^T S \mathbf{x}}{\mathbf{x}^T \mathbf{x}} &= \max_{\mathbf{x} \neq \mathbf{0}, \mathbf{q}_1^T \mathbf{x} = \cdots = \mathbf{q}_{i-1}^T \mathbf{x} = 0} \frac{\mathbf{x}^T Q \Lambda Q^T \mathbf{x}}{\mathbf{x}^T Q Q^T \mathbf{x}} = \max_{\mathbf{y} = Q^T \mathbf{x} \neq \mathbf{0}, y_1 = \cdots = y_{i-1} = 0} \frac{\mathbf{y}^T \Lambda \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \\ &= \max_{\mathbf{y} \neq \mathbf{0}, y_1 = \cdots = y_{i-1} = 0} \frac{\lambda_i y_i^2 + \cdots + \lambda_n y_n^2}{y_i^2 + \cdots + y_n^2} \leq \frac{\lambda_i (y_i^2 + \cdots + y_n^2)}{y_i^2 + \cdots + y_n^2} = \lambda_i. \end{aligned}$$

取  $\mathbf{y} = \mathbf{e}_i$  可取得等号。 □

**注记1.86.** 对  $-S$  运用上面命题, 得到

$$\lambda_n = \min_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T S \mathbf{x}}{\mathbf{x}^T \mathbf{x}}, \quad \lambda_i = \min_{\mathbf{x} \neq \mathbf{0}, \mathbf{q}_{i+1}^T \mathbf{x} = \cdots = \mathbf{q}_n^T \mathbf{x} = 0} \frac{\mathbf{x}^T S \mathbf{x}}{\mathbf{x}^T \mathbf{x}}, \quad 1 \leq i \leq n-1.$$

## 1.4.5. 二次曲面分类(阅读).

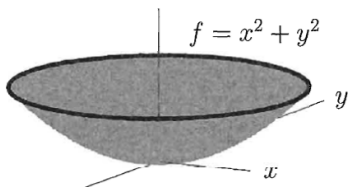
考虑 $\mathbb{R}^3$ 中二次曲面 $F(x, y) = ax^2 + 2bxy + cy^2$ 。  $F(0, 0) = 0$ ,  $\frac{\partial F}{\partial x}|_{(0,0)} = \frac{\partial F}{\partial y}|_{(0,0)} = 0$ . 将 $F(x, y)$ 写成矩阵形式

$$F(x, y) = [x, y] \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{x}^T S \mathbf{x}$$

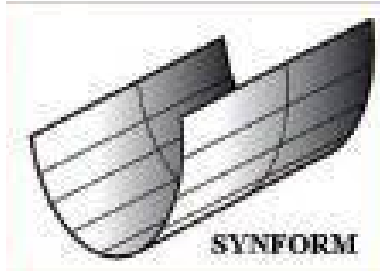
其中,  $S = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$  为对称矩阵。  $S$  有两个特征值 $\lambda_1, \lambda_2$ , 我们分如下情况讨论:

(1)  $\lambda_1 > 0, \lambda_2 > 0$ ; (2)  $\lambda_1 > 0, \lambda_2 = 0$ ; (3)  $\lambda_1 > 0, \lambda_2 < 0$ ; (4)  $\lambda_1 < 0, \lambda_2 < 0$ ; (5)  $\lambda_1 < 0, \lambda_2 = 0$ . 如果 $S$ 满足(4)或(5)(分别称为negative definit与negative semi-definit), 那么 $-S$ 满足(1)或(2), 因此, 我们只需考虑(1),(2),(3).

• 如果 $S$ 正定, 例如,  $S = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ,  $F(x, y) = x^2 + y^2$ , 为一个“碗”形。

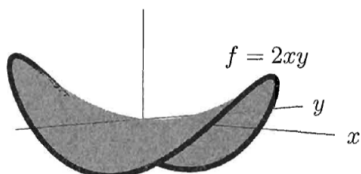


• 如果 $S$ 半正定, 例如,  $S = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ ,  $F(x, y) = x^2$ , 为一个“山谷”形状。





• 如果 $S$ 不定, 且特征值一正一负, 例如 $S = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ .  $F(x, y) = x^2 - y^2 = XY$ ,  $X = x + y, Y = x - y$ . 这时的二次曲面为一个“马鞍面”:



进一步研究正定情形:

$$S = Q\Lambda Q^T, \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \lambda_1, \lambda_2 > 0.$$

$$F(x, y) = \mathbf{x}^T S \mathbf{x} = [x, y] Q \Lambda Q^T \begin{bmatrix} x \\ y \end{bmatrix}.$$

做变量替换

$$\begin{bmatrix} X \\ Y \end{bmatrix} := Q^T \begin{bmatrix} x \\ y \end{bmatrix}$$

$F(X, Y) = [X, Y] \Lambda \begin{bmatrix} X \\ Y \end{bmatrix} = \lambda_1 X^2 + \lambda_2 Y^2$ . 考虑两条曲线  $1 = F(x, y)$  与  $1 = F(X, Y)$ .  $F(X, Y) = 1$  为椭圆  $\lambda_1 X^2 + \lambda_2 Y^2 = 1$ .  $F(x, y) = 1$  为椭圆  $ax^2 + 2bxy + cy^2 = 1$ , 两个椭圆正交变换  $Q$  (例如旋转) 相联系。

例如,  $F(x, y) = 5x^2 + 8xy + 5y^2 = [x, y] \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$ .

$$S = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} \cos \pi/4 & -\sin \pi/4 \\ \sin \pi/4 & \cos \pi/4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

取  $P_1 = (1/3, 0)$ ,  $P_2 = (0, 1)$  为  $XY$  平面上椭圆上的点。由于

$$\begin{bmatrix} x \\ y \end{bmatrix} = Q \begin{bmatrix} X \\ Y \end{bmatrix}$$

，我们将 $XY$ 上椭圆先沿 $X$ 轴反射，再逆时针旋转 $\pi/4$ 得到 $xy$ 上的椭圆。因此， $P_1, P_2$ 分别对应 $xy$ 上的点

$$p_1 = QP_1 = 1/3(1/\sqrt{2}, 1/\sqrt{2}), \quad p_2 = QP_2 = (1/\sqrt{2}, -1/\sqrt{2}).$$

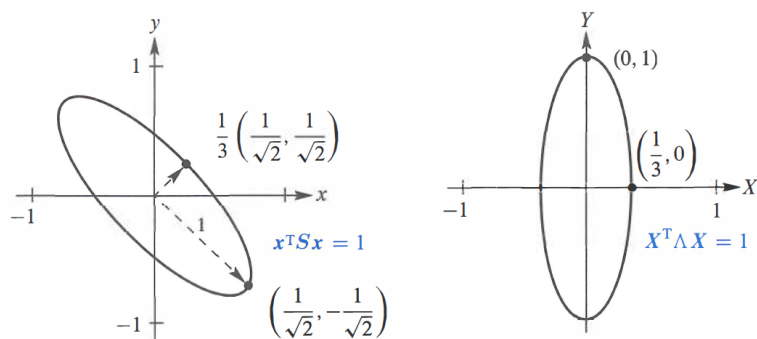


Figure 6.7: The tilted ellipse  $5x^2 + 8xy + 5y^2 = 1$ . Lined up it is  $9X^2 + Y^2 = 1$ .

## 2. 奇异值分解 (SINGULAR VALUE DECOMPOSITION)

未加特殊说明, 本章考虑的矩阵都是  $m \times n$  阶实矩阵。对  $A \in M_{m \times n}(\mathbb{R})$ , 一个重要研究手段是利用  $A^T A$  为半正定实对称矩阵。

### 2.1. 奇异值分解定理.

#### 2.1.1. 定理的陈述.

**定理2.1.** 设  $A \in M_{m \times n}(\mathbb{R})$  秩为  $r$ . 则有正交矩阵  $U \in M_{m \times m}(\mathbb{R})$ ,  $V \in M_{n \times n}(\mathbb{R})$  使得

$$A = U \Sigma V^T,$$

其中,  $\Sigma \in M_{m \times n}(\mathbb{R})$  为 (广义) 对角矩阵

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix}$$

且  $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ . (按重要性顺序对奇异值排序, 详见矩阵的低秩逼近部分).

**注记2.2.**  $\sigma_1, \dots, \sigma_r$  称为  $A$  的奇异值 (*singular value*), 分解中  $U, V$  不一定唯一, 但是奇异值由  $A$  唯一决定。

2.1.2. SVD定理的证明. 证明思路: 假设如果有定理中的分解  $A = U \Sigma V^T$ . 那么  $A^T A = V \Sigma^T \Sigma V^T$ , 即

$$A^T A V = V \Sigma^T \Sigma, \quad \Sigma^T \Sigma = \begin{bmatrix} \sigma_1^2 & & & \\ & \ddots & & \\ & & \sigma_r^2 & \\ & & & \end{bmatrix}$$

因此,  $V = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ , 满足

$$A^T A \mathbf{v}_i = \sigma_i^2 \mathbf{v}_i, \quad \forall i.$$

由于  $A^T A$  为对称半正定矩阵, 它可以由正交矩阵对角化, 且特征值非负。因此,  $\sigma_i^2$  为  $A^T A$  的非零特征值,  $\mathbf{v}_i$  为  $\sigma_i^2$  对应的特征向量。

$AV = U\Sigma$ 表明: 如果 $\sigma_i > 0$ ,

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i$$

即

$$\mathbf{u}_i = \frac{A\mathbf{v}_i}{\sigma_i}.$$

因此, 我们证明的思路即是从半正定矩阵 $A^T A$ 的对角化开始。

定理的证明. 记 $r := r(A^T A) = r(A)$ .  $A^T A$ 对称半正定, 令 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ 为 $A^T A$ 的所有非零特征值。 $V = [\mathbf{v}_1, \dots, \mathbf{v}_r, \dots, \mathbf{v}_n]$ 为特征向量构成的正交矩阵, 因此

$$A^T A V = V \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_r & \\ & & & \end{bmatrix}$$

特别地,

$$A^T A \mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad 1 \leq i \leq r$$

$$A^T A \mathbf{v}_i = \mathbf{0}, \quad r+1 \leq i \leq n.$$

注意, 由于 $N(A^T A) = N(A)$ , 我们得到

$$A \mathbf{v}_i = \mathbf{0}, \quad r+1 \leq i \leq n.$$

令 $\sigma_i := \sqrt{\lambda_i}, 1 \leq i \leq r$ . 定义

$$\mathbf{u}_i := \frac{A \mathbf{v}_i}{\sigma_i}, \quad 1 \leq i \leq r.$$

验证 $\mathbf{u}_1, \dots, \mathbf{u}_r \in \mathbb{R}^m$ 为标准正交向量组, 即

$$(1) \|\mathbf{u}_i\| = 1. \text{ 这是因为 } \mathbf{u}_i^T \mathbf{u}_i = \frac{\mathbf{v}_i^T A^T A \mathbf{v}_i}{\sigma_i^2} = \frac{\mathbf{v}_i^T \sigma_i^2 \mathbf{v}_i}{\sigma_i^2} = 1.$$

$$(2) \mathbf{u}_i^T \mathbf{u}_j = 0, \quad i \neq j. \text{ 这是因为 } \mathbf{u}_i^T \mathbf{u}_j = \frac{\mathbf{v}_i^T A^T A \mathbf{v}_j}{\sigma_i \sigma_j} = \frac{\lambda_j \mathbf{v}_i^T \mathbf{v}_j}{\sigma_i \sigma_j} = 0.$$

将 $\mathbf{u}_1, \dots, \mathbf{u}_r$ 扩充为 $\mathbb{R}^m$ 一组标准正交基 $\mathbf{u}_1, \dots, \mathbf{u}_m$ , 令

$$U := [\mathbf{u}_1, \dots, \mathbf{u}_m].$$

我们有

$$A[\mathbf{v}_1, \dots, \mathbf{v}_n] = [\mathbf{u}_1, \dots, \mathbf{u}_m] \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix}.$$

即  $A = U\Sigma V^T$ . □

**注记2.3.** 考虑  $AA^T = U\Sigma V^T V \Sigma^T U^T = U\Sigma \Sigma^T U^T$ . 所以,  $\mathbf{u}$  是  $AA^T$  特征值为  $\sigma^2$  的特征向量;  $\mathbf{v}$  是  $A^T A$  特征值为  $\sigma^2$  的特征向量.

$\mathbf{v}_i, \mathbf{u}_i$  称为奇异向量 (*singular vector*). 由于  $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ , 称  $\mathbf{v}_i$  为  $A$  的从属于  $\sigma_i$  的左奇异向量; 由于  $A^T \mathbf{u}_i = \sigma_i \mathbf{v}_i$ , 称  $\mathbf{u}_i$  为  $A$  的从属于  $\sigma_i$  的右奇异向量.

**例2.4.**  $A = \mathbf{u}\mathbf{v}^T \in M_{n \times n}(\mathbb{R})$ . 其特征值(不计重数)为 0,  $\mathbf{u}^T \mathbf{v}$ . (见例 1.41).

$A^T A = \|\mathbf{u}\|^2 \mathbf{v}\mathbf{v}^T$  也是秩 1 矩阵.  $A^T A$  的特征值(不计重数)为 0,  $\|\mathbf{u}\|^2 \mathbf{v}^T \mathbf{v} = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2$ . 因此,  $A$  的奇异值为  $\|\mathbf{u}\| \|\mathbf{v}\|$ .

2.1.3. 奇异值分解的几何意义. 通过奇异值分解, 矩阵  $A$  写成(正交)  $\times$  (对角)  $\times$  (正交). 以  $n = 2$  为例, 为简单起见, 假定两个正交矩阵都为旋转矩阵  $U = [\mathbf{u}_1, \mathbf{u}_2], V = [\mathbf{v}_1, \mathbf{v}_2]$  分别为逆时针旋转  $\theta, \phi$  角. 那么

$$A = U\Sigma V^T = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix} \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix}.$$

$A$  在  $\mathbb{R}^2$  上的作用由它在一组标准正交基上的作用决定. 我们看  $A$  在标准正交基  $\{\mathbf{v}_1, \mathbf{v}_2\}$  上的作用

$$A[\mathbf{v}_1, \mathbf{v}_2] = U\Sigma V^T[\mathbf{v}_1, \mathbf{v}_2].$$

作用分为 3 步:

1.  $V^T$  将  $\mathbf{v}_1, \mathbf{v}_2$  顺时针旋转  $\phi$  角, 分别映为  $\mathbf{e}_1, \mathbf{e}_2$ . ( $V^T[\mathbf{v}_1, \mathbf{v}_2] = V^T V = I_2$ );
2.  $\Sigma$  将  $\mathbf{e}_1, \mathbf{e}_2$  分别拉伸  $\sigma_1, \sigma_2$  倍, 映为  $\sigma_1 \mathbf{e}_1, \sigma_2 \mathbf{e}_2$ ;
3.  $U$  将  $\sigma_1 \mathbf{e}_1, \sigma_2 \mathbf{e}_2$  逆时针旋转  $\theta$  角, 分别映为  $\sigma_1 \mathbf{u}_1, \sigma_2 \mathbf{u}_2$ .

图示如下:

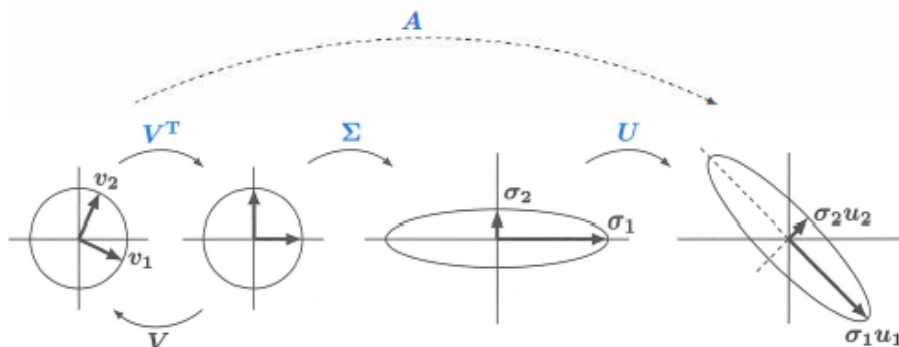


Figure 7.5:  $U$  and  $V$  are rotations and possible reflections.  $\Sigma$  stretches circle to ellipse.

#### 2.1.4. SVD与线性代数基本定理 (v.3.0) .

$A$ 的奇异值分解给出了四个子空间的一组标准正交基: 将等式 $A = U\Sigma V^T$ 写为

$$AV = U\Sigma.$$

$V = [\mathbf{v}_1, \dots, \mathbf{v}_n]$  ( $\mathbf{v}_1, \dots, \mathbf{v}_n$ 构成 $\mathbb{R}^n$ 一组标准正交基); 令 $U = [\mathbf{u}_1, \dots, \mathbf{u}_m]$  ( $\mathbf{u}_1, \dots, \mathbf{u}_m$ 构成 $\mathbb{R}^m$ 一组标准正交基).

**命题2.5.** (1)  $\mathbf{v}_1, \dots, \mathbf{v}_r$ 构成 $C(A^T)$ 的一组基。

(2)  $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ 构成 $N(A)$ 的一组基。

(3)  $\mathbf{u}_1, \dots, \mathbf{u}_r$ 构成 $C(A)$ 的一组基。

(4)  $\mathbf{u}_{r+1}, \dots, \mathbf{u}_m$ 构成 $N(A^T)$ 的一组基。

证明. 事实上,

$$A\mathbf{v}_{r+1} = 0, \dots, A\mathbf{v}_n = 0$$

表明 $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n \in N(A)$ , 且由于 $\dim N(A) = n - r$ ,

$$N(A) = \text{Span}(\mathbf{v}_{r+1}, \dots, \mathbf{v}_n)$$

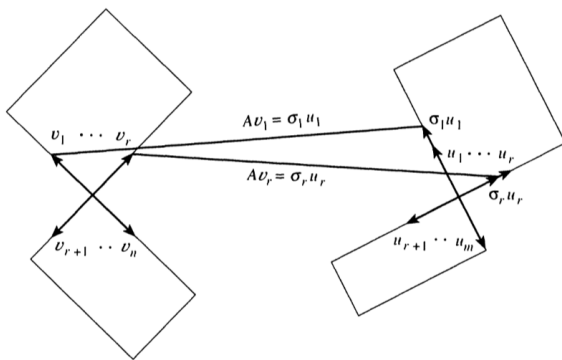
因此,  $C(A^T) = N(A)^\perp = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_r)$ .

$1 \leq i \leq r$ ,  $\mathbf{u}_i = \frac{A\mathbf{v}_i}{\sigma_i} \in C(A)$ . 由于 $\dim \text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_r) = r = \dim C(A)$ ,  $\mathbf{u}_1, \dots, \mathbf{u}_r$ 构成 $C(A)$ 一组基。  $\square$

**定理2.6** (线性代数基本定理 (v.3.0)).  $A \in M_{m \times n}(\mathbb{R})$ . 记 $r := r(A)$ . 我们有

- (1)  $r \leq \min(m, n)$ .
- (2)  $\dim C(A^T) = r$ . (一组基由  $rref(A)$  的非零行给出。)
- (3)  $\dim N(A) = n - r$ . (一组基由特殊零解给出。)
- (4)  $\dim C(A) = r$ . (一组基由  $A$  的主元列给出。)
- (5)  $\dim N(A^T) = m - r$ . (一组基由  $EA = rref(A)$  中  $E$  的后  $m - r(A)$  个行向量给出。)
- (6)  $N(A)$  与  $C(A^T)$  互为正交补.  $\mathbb{R}^n = N(A) \oplus C(A^T)$ .  $N(A^T)$  与  $C(A)$  互为正交补.  $\mathbb{R}^m = N(A^T) \oplus C(A)$ .
- (7)  $C(A^T), N(A), C(A), N(A^T)$  存在标准正交基  $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}, \{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}, \{\mathbf{u}_1, \dots, \mathbf{u}_r\}, \{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$  使得有  $\sigma_i > 0, 1 \leq i \leq r$  满足

$$A\mathbf{v}_1 = \sigma_1\mathbf{u}_1, A\mathbf{v}_2 = \sigma_2\mathbf{u}_2, \dots, A\mathbf{v}_r = \sigma_r\mathbf{u}_r, A\mathbf{v}_{r+1} = \mathbf{0}, \dots, A\mathbf{v}_n = \mathbf{0}.$$



2.2. 矩阵的低秩逼近. 设  $A \in M_{m \times n}(\mathbb{R})$  秩为  $r$ .

$$A = U\Sigma V^T = [\mathbf{u}_1, \dots, \mathbf{u}_m] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}$$

$$= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \dots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T.$$

$A$ 为 $r$ 个秩1矩阵的和。因此,  $U_r = [\mathbf{u}_1, \dots, \mathbf{u}_r]$ ,  $\Sigma_r = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$ ,

$V_r = [\mathbf{v}_1, \dots, \mathbf{v}_r]$ 的信息决定了 $A$ . 即

$$A = U_r \Sigma_r V_r^T.$$

这称为 $A$ 的简化奇异值分解。

**定义2.7.** 对 $k \leq r$ , 令

$$A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \dots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T = [\mathbf{u}_1, \dots, \mathbf{u}_k] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_k^T \end{bmatrix}.$$

$$\text{令 } U_k = [\mathbf{u}_1, \dots, \mathbf{u}_k], \Sigma_k = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix}, V_k = [\mathbf{v}_1, \dots, \mathbf{v}_k].$$

$$A_k = U_k \Sigma_k V_k^T = [\mathbf{u}_1, \dots, \mathbf{u}_m] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}.$$

由于 $U, V$ 可逆,  $A_k$ 的秩为 $k$ . 我们称 $A_k$ 为矩阵 $A$ 的秩为 $k$ 的逼近。

Eckart-Young-Mirsky定理:  $A_k$ 是所有秩 $k$ 矩阵中“距离” $A$ 最近的矩阵。为了使这句话有意义, 我们在下一小节定义矩阵距离的概念。

**例2.8.**  $A - A_k = \sigma_{k+1} \mathbf{u}_{k+1} \mathbf{v}_{k+1}^T + \dots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$ . 因此,  $A - A_k$ 的简化奇

异值分解为 $[\mathbf{u}_{k+1}, \dots, \mathbf{u}_r] \begin{bmatrix} \sigma_{k+1} & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_{k+1}^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix}$ . 因此,  $\sigma_{k+1} \geq$

$\dots \geq \sigma_r$ 为 $A - A_k$ 的奇异值。



将 $[\mathbf{u}_{k+1}, \dots, \mathbf{u}_r]$ 扩充为 $\mathbb{R}^m$ 的一组标准正交基; 将 $[\mathbf{v}_{k+1}, \dots, \mathbf{v}_r]$ 扩充为 $\mathbb{R}^n$ 的一组标准正交基; 将 $\Sigma' = \begin{bmatrix} \sigma_{k+1} & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$ 补为 $\begin{bmatrix} \Sigma' & O \\ O & O \end{bmatrix}$ , 可以得到 $A = A_k$ 的奇异值分解。

**注记2.9.** 分解 $A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \dots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$ 表明 $A$ 与零空间和左零空间标准正交基 $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ ,  $\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$ 的选取无关。

### 2.2.1. 矩阵的范数与矩阵空间上的距离.

**定义2.10.**  $M_{m \times n}(\mathbb{R})$ 上的函数 $f: M_{m \times n}(\mathbb{R}) \rightarrow \mathbb{R}$ 称为是一个范数(norm) (范数是绝对值的推广), 如果 $f$ 满足

- (1)  $f(A+B) \leq f(A) + f(B)$ ,
- (2)  $f(cA) = |c|f(A)$ ,  $c \in \mathbb{R}$
- (3)  $f(A) = 0$ 当且仅当 $A = 0_{m \times n}$ .

**定义2.11.** 定义函数 $\|\cdot\|: M_{m \times n}(\mathbb{R}) \rightarrow \mathbb{R}_{\geq 0}$ ,  $\|A\| := \max_{\mathbf{x} \neq 0} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\mathbf{x}, \|\mathbf{x}\|=1} \|A\mathbf{x}\|$ . ( $\{\mathbf{x} \mid \|\mathbf{x}\|=1\}$ 为 $\mathbb{R}^n$ 的单位实球)。这里 $\|A\mathbf{x}\|$ ,  $\|\mathbf{x}\|$ 分别为向量 $A\mathbf{x}$ ,  $\mathbf{x}$ 的长度。

**例2.12.** 设 $\mathbf{u} \in M_{n \times 1}(\mathbb{R})$ , 容易验证 $\|\mathbf{u}\|$ 作为矩阵范数等于 $\mathbf{u}$ 向量的长度。

**命题2.13.** (1)  $\|A+B\| \leq \|A\| + \|B\|$ ,

(2)  $\|cA\| = |c|\|A\|$ ,

(3)  $\|A\| = 0$ 当且仅当 $A = 0_{m \times n}$ .

(4)  $\|AB\| \leq \|A\|\|B\|$ .

**注记2.14.** (1),(2),(3)说明 $\|\cdot\|$ 是矩阵空间 $M_{m \times n}(\mathbb{R})$ 上的一个范数。称为矩阵的谱范数。

(4)说明矩阵乘法在这个范数下是连续的。

证明. (1)  $\frac{\|(A+B)\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{\|A\mathbf{x} + B\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\|A\mathbf{x}\| + \|B\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} + \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|}$ , 得到 $\|A+B\| \leq \|A\| + \|B\|$ .

(2)  $\frac{\|(cA)\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{\|cA\mathbf{x}\|}{\|\mathbf{x}\|} = |c| \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}$ , 得到 $\|cA\| = |c|\|A\|$ .

(3) 如果  $A = 0_{m \times n}$ ,  $\|A\| = 0$ . 反之, 如果  $\|A\| = 0$ , 那么  $\|A\mathbf{x}\| = 0, \forall \mathbf{x} \neq 0$ . 所以,  $N(A) = \mathbb{R}^n$ , 得到  $A = 0_{m \times n}$ .

(4)  $\|AB\| = \max_{\mathbf{x}} \frac{\|(AB)\mathbf{x}\|}{\|\mathbf{x}\|} = \max(\frac{\|AB\mathbf{x}\|}{\|B\mathbf{x}\|} \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|}) \leq \max(\frac{\|AB\mathbf{x}\|}{\|B\mathbf{x}\|}) \max(\frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|}) \leq \|A\|\|B\|.$   $\square$

就像有了绝对值, 我们定义两个数  $a, b$  之间的距离为  $|a - b|$ ; 我们有了矩阵范数, 就可以定义两个矩阵的距离: 对  $A, B \in M_{m \times n}(\mathbb{R})$ , 定义

$$d(A, B) := \|A - B\|$$

根据范数的性质我们容易得到:

- (1)  $d(A, B) = 0$  当且仅当  $A = B$ ;
- (2)  $d(A, B) = d(B, A)$ ;
- (3)  $d(A, C) \leq d(A, B) + d(B, C)$

这三条说明  $d$  定义了向量空间  $M_{m \times n}(\mathbb{R})$  上的一个距离。

**定理2.15** (Eckart-Young-Mirsky). 令  $A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$ . 对任意秩为  $k$  的矩阵  $B \in M_{m \times n}(\mathbb{R})$ , 我们有

$$d(A, A_k) = \|A - A_k\| \leq \|A - B\| = d(A, B).$$

为了证明这个定理, 我们需要建立矩阵范数与奇异值的关系。

**命题2.16.**  $A = U\Sigma V^T$  为奇异值分解, 那么  $\|A\| = \max_{\mathbf{x} \neq 0} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \sigma_1$ .

证明. 令  $S = A^T A = V \Lambda V^T$ ,  $\Lambda = \text{Diag}(\sigma_1^2, \dots, \sigma_r^2, 0, \dots, 0)$ ,  $V = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ .

由命题1.85,  $\sigma_1^2 = \max_{\mathbf{x} \neq 0} \frac{\mathbf{x}^T A^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \max_{\mathbf{x} \neq 0} \frac{\|A\mathbf{x}\|^2}{\|\mathbf{x}\|^2}$ . 于是  $\sigma_1 = \max_{\mathbf{x} \neq 0} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}$   $\square$

**注记2.17.** 由命题1.85, 我们同样可证:  $\sigma_k = \max_{\mathbf{x} \neq 0, \mathbf{v}_1^T \mathbf{x} = \dots = \mathbf{v}_{k-1}^T \mathbf{x} = 0} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}$ .

证明:  $2 \leq k \leq r$ ,

$$\sigma_k^2 = \max_{\mathbf{x} \neq 0, \mathbf{v}_1^T \mathbf{x} = \dots = \mathbf{v}_{k-1}^T \mathbf{x} = 0} \frac{\mathbf{x}^T A^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \max_{\mathbf{x} \neq 0, \mathbf{v}_1^T \mathbf{x} = \dots = \mathbf{v}_{k-1}^T \mathbf{x} = 0} \frac{\|A\mathbf{x}\|^2}{\|\mathbf{x}\|^2}$$

等价地  $\sigma_k = \max_{\mathbf{x} \neq 0, \mathbf{v}_1^T \mathbf{x} = \dots = \mathbf{v}_{k-1}^T \mathbf{x} = 0} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}$ .

**推论2.18.**  $\|A - A_k\| = \sigma_{k+1}$ .

证明. 由例2.8,  $A - A_k = \sigma_{k+1} \mathbf{u}_{k+1} \mathbf{v}_{k+1}^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$  的最大奇异值为  $\sigma_{k+1}$ .  $\square$

Eckart-Young-Mirsky定理的证明. 由于  $r(B) = k$ ,  $\dim N(B) = n - k$ . 因此,

$$\dim N(B) + \dim C(A_{k+1}^T) = n - k + k + 1 = n + 1.$$

由维数公式

$$\dim N(B) \cap C(A_{k+1}^T) = \dim N(B) + \dim C(A_{k+1}^T) - \dim(N(B) + C(A_{k+1}^T)) \geq n + 1 - n = 1.$$

所以, 存在非零向量  $\mathbf{x} \in N(B) \cap C(A_{k+1}^T)$ .

由命题2.5,  $\mathbf{v}_1, \dots, \mathbf{v}_{k+1}$  构成  $C(A_{k+1}^T)$  的一组标准正交基, 因此由  $\mathbf{x} \in C(A_{k+1}^T)$ , 存在  $c_1, \dots, c_{k+1} \in \mathbb{R}$  使得  $\mathbf{x} = c_1 \mathbf{v}_1 + \cdots + c_{k+1} \mathbf{v}_{k+1}$ .

$$\begin{aligned} \|A - B\| &\geq \frac{\|(A - B)\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{\|\sigma_1 c_1 \mathbf{v}_1 + \cdots + \sigma_{k+1} c_{k+1} \mathbf{v}_{k+1}\|}{\|\mathbf{x}\|} \\ &= \frac{\sqrt{|\sigma_1 c_1|^2 + \cdots + |\sigma_{k+1} c_{k+1}|^2}}{\sqrt{c_1^2 + \cdots + c_{k+1}^2}} \geq \frac{\sqrt{|\sigma_{k+1} c_1|^2 + \cdots + |\sigma_{k+1} c_{k+1}|^2}}{\sqrt{c_1^2 + \cdots + c_{k+1}^2}} \\ &= \sigma_{k+1} = \|A - A_k\|. \end{aligned}$$

$\square$

### 2.3. 奇异值分解的应用.

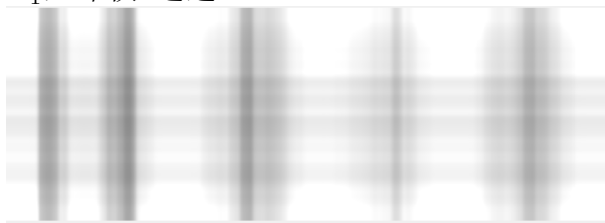
2.3.1. 数据压缩. 在一张图片上打上均匀的正方形网格, 对每个网格, 用一个数值记忆其灰度的深浅 (我们假定图片是黑白图片)。用此方法将图片转化为矩阵  $A \in M_{m \times n}(\mathbb{R})$ 。存储  $A$  的信息需要储存  $mn$  个数。如果对  $A$  做秩  $k$  逼近,  $A_k = U_k \Sigma_k V_k^T$ , 则只需要储存  $mk + k + kn = k(m + n + 1)$  个数。当  $m, n$  很大时, 后者所需储存远小于前者。

原图:

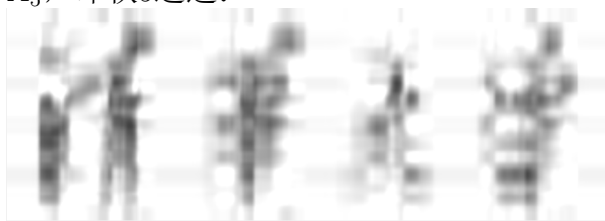


转化后的矩阵  $A$  为  $244 \times 691$  阶矩阵, 秩为 135

$A_1$ , 即秩1逼近:



$A_5$ , 即秩5逼近:



$A_{10}$ , 即秩10逼近:

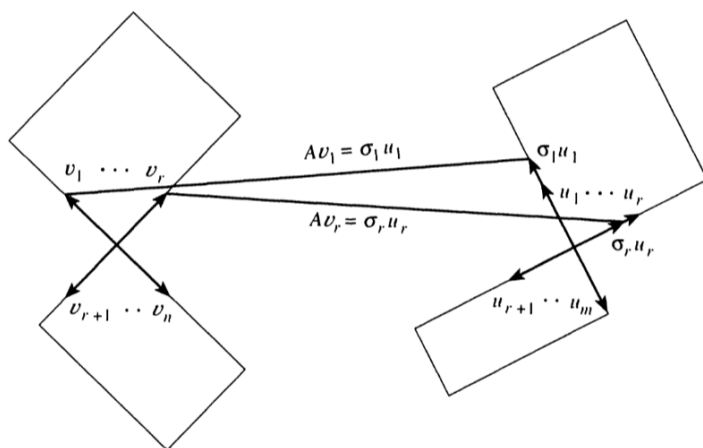


$A_{20}$ , 即秩20逼近:



2.3.2. 广义逆. 广义逆(pseudo-inverse)是矩阵逆在 $M_{m \times n}(\mathbb{R})$ 情形下的推广。

$$A \in M_{m \times n}(\mathbb{R}) \rightsquigarrow A^+ \in M_{n \times m}(\mathbb{R}).$$



$A^+ : \mathbb{R}^m \rightarrow \mathbb{R}^n$  应该满足

$$A^+(\sigma_1 \mathbf{u}_1) = \mathbf{v}_1, \dots, A^+(\sigma_r \mathbf{u}_r) = \mathbf{v}_r; A^+(\mathbf{u}_{r+1}) = \dots = A^+(\mathbf{u}_m) =$$

0.

或由线性性:

$$A^+(\mathbf{u}_1) = \frac{\mathbf{v}_1}{\sigma_1}, \dots, A^+(\mathbf{u}_r) = \frac{\mathbf{v}_r}{\sigma_r}; A^+(\mathbf{u}_{r+1}) = \dots = A^+(\mathbf{u}_m) = 0.$$

写成分块乘法

$$A^+[\mathbf{u}_1, \dots, \mathbf{u}_m] = [\mathbf{v}_1, \dots, \mathbf{v}_n] \begin{bmatrix} \sigma_1^{-1} & & & \\ & \ddots & & \\ & & \sigma_r^{-1} & \\ & & & 0 \end{bmatrix}.$$

**定义2.19.** 设  $\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix} \in M_{m \times n}(\mathbb{R})$ . 记  $\Sigma^+ := \begin{bmatrix} \sigma_1^{-1} & & & \\ & \ddots & & \\ & & \sigma_r^{-1} & \\ & & & \end{bmatrix} \in M_{n \times m}(\mathbb{R})$ .  $\Sigma^+$  满足

$$\Sigma \Sigma^+ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \in M_{m \times m}(\mathbb{R}), \quad \Sigma^+ \Sigma = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \in M_{n \times n}(\mathbb{R}).$$

**定义2.20.** (1) 如果  $A$  可逆, 定义其广义逆即为矩阵的逆  $A^+ = A^{-1}$ .

(2) 设  $A \in M_{m \times n}(\mathbb{R})$ . 根据奇异值分解  $A = U \Sigma V^T$ . 定义  $A$  的广义逆

$$A^+ = (U \Sigma V^T)^+ = (V^T)^+ \Sigma^+ U^+ = V \Sigma^+ U^T \in M_{n \times m}(\mathbb{R}).$$

**注记2.21.**  $A^+ = V \Sigma^+ U^T = \sigma_1^{-1} \mathbf{v}_1 \mathbf{u}_1^T + \cdots + \sigma_r^{-1} \mathbf{v}_r \mathbf{u}_r^T$ . 由于  $\sigma_r^{-1} \geq \cdots \geq \sigma_1^{-1} > 0$ ,  $A^+$  的奇异值分解为

$$A^+ = [\mathbf{v}_r, \dots, \mathbf{v}_1, \mathbf{v}_{r+1}, \dots, \mathbf{v}_n] \begin{bmatrix} \sigma_r^{-1} & & & \\ & \ddots & & \\ & & \sigma_1^{-1} & \\ & & & \end{bmatrix} \begin{bmatrix} \mathbf{u}_r^T \\ \vdots \\ \mathbf{u}_1^T \\ \mathbf{u}_{r+1}^T \\ \vdots \\ \mathbf{u}_m^T \end{bmatrix}.$$

**推论2.22.**  $\text{rank}(A^+) = r$ .  $A^+$  满足  $C(A^+) = C(A^T)$ ,  $N((A^+)^T) = N(A)$ ,  $C((A^+)^T) = C(A)$ ,  $N(A^+) = N(A^T)$ .

证明. 已知  $A^+$  的奇异值分解, 于是  $A^+$  的四个子空间的一组标准正交基由命题2.5应用到  $A^+$  给出. 与  $A$  的四个子空间做比较得到推论.  $\square$

**注记2.23.** 广义逆可以由一些代数性质唯一刻画。这些代数性质称为Penrose条件:  $A^+$ 为满足下列四条的唯一  $n \times m$  矩阵

- (1)  $AA^+A = A$ ,
- (2)  $A^+AA^+ = A^+$ ,
- (3)  $(AA^+)^T = AA^+$ ,
- (4)  $(A^+A)^T = A^+A$ .

**例2.24.** 求  $A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$  的奇异值分解。  $A^T A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix}$ ,

$\sigma_1 = 3 > \sigma_2 = 2 > \sigma_3 = 1$ . 求解  $A^T A \mathbf{v} = \sigma^2 \mathbf{v}$  得到

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

取  $\mathbf{v}_4 = [1, 0, 0, 0]^T$ . 求  $\mathbf{u} = \frac{A\mathbf{v}}{\sigma}$  得到

$$\mathbf{u}_1 = \frac{A\mathbf{v}_1}{\sigma_1} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \frac{A\mathbf{v}_2}{\sigma_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{u}_3 = \frac{A\mathbf{v}_3}{\sigma_3} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

$$A = U\Sigma V^T.$$

### 2.3.3. 广义逆与正交投影.

**命题2.25.**  $A \in M_{m \times n}(\mathbb{R})$ .

- (1)  $A^+A$  为  $\mathbb{R}^n$  向  $A$  的行空间的正交投影矩阵.
- (2)  $I - A^+A$  为  $\mathbb{R}^n$  向  $A$  的零空间的正交投影矩阵.
- (3)  $AA^+$  为  $\mathbb{R}^m$  向  $A$  的列空间的正交投影矩阵.
- (4)  $I - AA^+$  为  $\mathbb{R}^m$  向  $A$  的左零空间的正交投影矩阵.

证明. (2)由(1)得到, (4)由(3)得到. 只需证明(1)(3).

$$(1) A^+A = V\Sigma^+U^T U\Sigma V^T = V \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} V^T. A^+AV = V \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

此,  $1 \leq i \leq r$ ,  $A^+A\mathbf{v}_i = \mathbf{v}_i$ ;  $r+1 \leq i \leq n$ ,  $A^+A\mathbf{v}_i = \mathbf{0}$ . 由命

题2.5,  $A^+A$  在  $A$  的行空间上是恒等映射, 在零空间上是零映射。因此,  $A^+A$  为  $\mathbb{R}^n$  向  $A$  的行空间正交投影矩阵

$$AA^+ = U\Sigma V^T V\Sigma^+ U^T = U \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U^T. \quad AA^+U = U \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

由命题2.5,  $AA^+$  在  $A$  的列空间上是恒等映射, 在左零空间上是零映射。因此,  $AA^+$  为  $\mathbb{R}^m$  向  $A$  的列空间的正交投影。  $\square$

2.3.4. 最小二乘问题. 回顾最小二乘问题:  $\mathbf{b} \in \mathbb{R}^m$ , 找  $\mathbf{p} = A\hat{\mathbf{x}} \in C(A)$  满足  $\mathbf{p}$  为列空间中距离  $\mathbf{b}$  最近点。这等价于  $(\mathbf{b} - A\hat{\mathbf{x}}) \perp C(A)$ , 即

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0 \quad \text{或} \quad A^T A\hat{\mathbf{x}} = A^T \mathbf{b}.$$

当  $A$  列满秩时, 方程有唯一解; 当  $A$  不是列满秩时, 方程一定有解且有无穷多解。落在  $C(A^T)$  中的解  $\mathbf{x}^+$  为长度最短的最小二乘解, 称为**最优最小二乘解**。由勾股定理, 最小二乘解  $\hat{\mathbf{x}}$  为最优最小二乘解, 如果  $\hat{\mathbf{x}} \in C(A^T)$ . 利用广义逆, 我们可以直接给出最优最小二乘解。

**命题2.26.**  $\mathbf{x}^+ = A^+\mathbf{b}$  为方程  $A\mathbf{x} = \mathbf{b}$  最优最小二乘解。

证明. (1) 首先说明  $A^+\mathbf{b}$  为最小二乘解, 即  $A^+\mathbf{b}$  满足  $A(A^+\mathbf{b}) = \mathbf{p}$

$$A(A^+\mathbf{b}) = (AA^+)\mathbf{b}$$

由于  $AA^+$  为向  $A$  的列空间的正交投影矩阵,

$$A(A^+\mathbf{b}) = (AA^+)\mathbf{b} = \mathbf{p}.$$

(2) 其次, 我们验证  $A^+\mathbf{b} \in C(A^T)$ . 这由推论2.22立得。  $\square$