

高等代数 I

主讲教师：高 峡

xia@math.pku.edu.cn

智华楼 445

助教：苏本朝 王浩然 龚诚欣



大课：周二 3,4 节 周四 1,2 节

理教 102

习题课：周四 10, 11 节

三教 103, 107, 205



作业: 12月19日交

§ 5.7 1 (求 $X^T A X$ 在单位球面 $\|X\| = 1$ 上的最值, 在哪里取到), 4, 5

§ 6.1 1, 2 (求对称轴), 3(1)(3), 5,
6, 8, 13

补充题: 1, 2, 3, 4, 5

1. 求矩阵 $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$ 的 SVD 分解

$$A = P \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix} Q^T,$$

这里 P, Q 是正交矩阵, $\sigma_1 \geq \sigma_2$ 是 A 的奇异值. 写出 A 的最佳秩 1 逼近; 求 A^+ .

2. 证明 $A \in M_{m,n}(\mathbb{R})$ 的奇异值满足以下性质:

1) 若 P, Q 是正交矩阵, 则 PAQ 与 A 有相同的奇异值;

2) 若 A 是实对称矩阵, 则 A 的奇异值是 A 非零实特征值的绝对值 (从大到小排列);

3) $\max_{\|\beta\|=1, \|\gamma\|=1} |\beta^T A \gamma| = \sigma_1$ (A 最大奇异值).

3. 满足条件 $A^T = \bar{A}$ 的复方阵 A 称为埃尔米特 (Hermite) 矩阵; 满足条件 $A^T = -\bar{A}$ 的复方阵 A 称为反 Hermite 矩阵.

证明: Hermite 矩阵的复特征值都是实数;
反 Hermite 矩阵的复特征值都是 0 或纯虚数.

4. 设 n 级实矩阵 A 有 n 个实特征值 $\lambda_1, \dots, \lambda_n$,

又设 α 是属于 λ_1 的单位特征向量. 证明:

对任意 $\beta \in \mathbb{R}^n$, 矩阵 $A + \alpha \beta^T$ 的 n 个

特征值为

$$\lambda_1 + \beta^T \alpha, \lambda_2, \dots, \lambda_n;$$

且 $(A + \alpha \beta^T)\alpha = (\lambda_1 + \beta^T \alpha)\alpha.$

5. 某地大致由 A 区, B 区和 C 区三部分组成.

现向 A 区, B 区各投放共享单车 10 万辆.

经统计, 每周 A 区有 10% 的车流入 B 区,

10% 流入 C 区; B 区车有 10% 流入 A 区,

5% 流入 C 区; C 区车 5% 流入 A 区, 5%

流入 B 区. 求 k 周后 3 个区的车分布,

以及当 $k \rightarrow \infty$ 时车的分布.

提示：记第 k 周的车分布为 $X_k = \begin{bmatrix} a_k \\ b_k \\ c_k \end{bmatrix}$ (万辆)

则有 $X_0 = \begin{bmatrix} 10 \\ 10 \\ 0 \end{bmatrix}$, Markov 矩阵 A

$$X_{k+1} = \begin{bmatrix} a_{k+1} \\ b_{k+1} \\ c_{k+1} \end{bmatrix} = \begin{bmatrix} .8 & .1 & .05 \\ .1 & .85 & .05 \\ .1 & .05 & .9 \end{bmatrix} \begin{bmatrix} a_k \\ b_k \\ c_k \end{bmatrix}$$

将 $X_0 = \alpha_1 + \alpha_2 + \alpha_3$ 写成 A 特征向量的和

$$X_k = A^k X_0 = 1\alpha_1 + \lambda_2^k \alpha_2 + \lambda_3^k \alpha_3 \rightarrow \alpha_1$$

第 2 特征值(用 Matlab 计算)大小决定收敛的速度

用 Matlab 计算特征值与特征向量

```
>> A = [ 1, -2, -4;  
        -2, 4, -2;  
        -4, -2, 1 ]
```

```
A =  
    1    -2    -4  
   -2     4    -2  
   -4    -2     1
```

```
>> [ U, D ] = eig( A )
```

```
U =  
    0.6667   -0.3431   -0.6617  
    0.3333    0.9313   -0.1470  
    0.6667   -0.1225    0.7352
```

```
D =
```

```
   -4.0000         0         0  
         0    5.0000         0  
         0         0    5.0000
```

```
>> U * D * U^(-1)
```

```
    1.0000   -2.0000   -4.0000  
   -2.0000    4.0000   -2.0000  
   -4.0000   -2.0000    1.0000
```

```
>> U * U'
```

```
    1.0000   -0.0000         0  
   -0.0000    1.0000         0  
         0         0    1.0000
```

若 A 是非负实矩阵, 且 A 每一列元素的和都为 1, 则称 A 是 Markov 矩阵.

$$A = \begin{bmatrix} .8 & .1 & .05 \\ .1 & .85 & .05 \\ .1 & .05 & .9 \end{bmatrix}$$

定理:

若 A 是 Markov 矩阵, 则有

- 1) $\lambda = 1$ 是 A 的一个特征值;
- 2) A 的所有复特征值的模都 ≤ 1 .

证: 1) 由于 A^T 的每行和都为 1, 于是有

$$A^T [1 \ 1 \ \dots \ 1]^T = [1 \ 1 \ \dots \ 1]^T,$$

故 $\lambda = 1$ 是 A^T (因而也是 A) 的特征值;

2) 法一: 若 λ 是 A 的复特征值, 则 λ 落在

某个盖尔圆内:

$$|\lambda - a_{ii}| \leq \sum_{1 \leq j \leq n, j \neq i} |a_{ji}| = 1 - a_{ii}.$$

于是 $|\lambda| \leq 1$.

盖尔圆定理: 设 $A = [a_{ij}] \in M_n(\mathbb{C})$, 定义

$$r_i(A) = \sum_{1 \leq j \leq n, j \neq i} |a_{ij}|,$$

$$D_i(A) = \{z \in \mathbb{C} \mid |z - a_{ii}| \leq r_i(A)\}.$$

证明: A 的复特征值都至少在某一个圆盘

$$D_i(A) \ (1 \leq i \leq n) \text{ 中.}$$

证: 设 $A\beta = \lambda\beta$, $\beta = [b_1 \cdots b_n]^T \neq 0$.

$$\text{不妨设 } |b_k| = \max_{1 \leq j \leq n} |b_j|.$$

考察 $A\beta = \lambda\beta$ 的第 k 个分量, 我们有

$$(\lambda - a_{kk})b_k = \sum_{1 \leq j \leq n, j \neq k} a_{kj} b_j .$$

于是

$$|\lambda - a_{kk}| |b_k| \leq \sum_{1 \leq j \leq n, j \neq k} |a_{kj}| |b_j| = r_k(A) |b_k|$$

即
$$|\lambda - a_{kk}| \leq r_k(A) .$$

注: 由盖尔圆定理可推出对角优势矩阵可逆.

2) 法二：若 A, B 是 Markov 矩阵，则有

$$[1 \ 1 \ \dots \ 1] A B = [1 \ 1 \ \dots \ 1] B = [1 \ 1 \ \dots \ 1].$$

于是非负实矩阵 AB 也是 Markov 矩阵.

特别地, 对任意正整数 k , A^k 是 Markov

矩阵, 故 A^k 的元素总在区间 $[0, 1]$ 中变化.

设 λ 是 A 的复特征值, α 是 λ 的特征向量,

即
$$A \alpha = \lambda \alpha, \quad 0 \neq \alpha \in \mathbb{C}^n.$$

则
$$A^k \alpha = \lambda^k \alpha, \quad \forall k.$$

当 $k \rightarrow \infty$ 时, 由以上讨论, $A^k \alpha$ 的分量

都有界, 因而 $\lambda^k \alpha$ 的分量也有界,

由此推出 $|\lambda| \leq 1$.

元素都是正数的矩阵称为正矩阵, 元素都 ≥ 0 的矩阵称为非负矩阵. 这类矩阵在工程学, 经济学的稳定性理论中发挥着重要作用.

定理 (Perron-Frobenius) $A = \begin{bmatrix} .8 & .1 & .05 \\ .1 & .85 & .05 \\ .1 & .05 & .9 \end{bmatrix}$

对于正方阵 A , 以下命题成立:

- 1) A 有正特征值. 设 λ 是 A 的最大正特征值, 则 λ 的代数重数为 1, 且有正特征向量 α ;

2) 除 λ 以外, A 其它复特征值的绝对值 $< \lambda$;

3) 正方阵 A^T 与 A 有相同的最大正特征值.

注: A 的最大正特征值 λ 又称为 A 的

Frobenius 根, 它又等于 A 的谱半径

(即 A 所有复特征值的最大绝对值),

对正方阵取极限可推出非负方阵的结果.

定理. 对于非负方阵 A , 我们有

- 1) A 有非负特征值. 设 λ 是 A 的最大非负特征值, 则 λ 有非负的特征向量;
- 2) A 的复特征值的绝对值都不超过 λ ;
- 3) 非负方阵 A^T 与 A 有相同的最大非负特征值.

注: A 的最大正特征值 λ 是 A 的谱半径, 又称为 A 的 Frobenius 根.

例. 非负矩阵 $A = \begin{bmatrix} 0 & I_r & 0 \\ 0 & 0 & I_r \\ I_r & 0 & 0 \end{bmatrix}$ 的特征值为

$1, e^{2\pi i/3}, e^{-2\pi i/3}$ (代数, 几何重数都为 r)

$$\begin{bmatrix} 0 & I_r & 0 \\ 0 & 0 & I_r \\ I_r & 0 & 0 \end{bmatrix} \begin{bmatrix} I_r \\ I_r \\ I_r \end{bmatrix} = \begin{bmatrix} I_r \\ I_r \\ I_r \end{bmatrix}$$

第五章 矩阵的相似

1. 线性变换及其矩阵表示
2. 矩阵的相似
3. 矩阵的特征值与特征向量
4. 矩阵对角化的条件
5. 实对称矩阵正交对角化



实对称矩阵

不论在理论上，还是在实际问题里，
实对称矩阵都有大量应用。

实对称矩阵有实特征值，可用正交
矩阵对角化。

实对称矩阵的基本理论：

- 1) n 级实对称矩阵有 n 个实特征值；
- 2) 属于不同特征值的特征向量彼此正交；
- 3) 实对称矩阵都能正交对角化

$$A = P D P^{-1} = P D P^T$$

其中 P 是正交矩阵， D 实对角矩阵

定理 1:

实对称矩阵在复数域上的特征值
都是实数.

(特征多项式的根都分布在实轴上;
使得 $A - \lambda I$ 退化的 λ 一定是实数)

证：设 λ 是实对称矩阵 A 的复特征值，

α 是属于 λ 的复特征向量，即

$$A\alpha = \lambda\alpha, \quad \alpha \neq 0.$$

作转置

$$\alpha^T A = \lambda \alpha^T$$

求复共轭

$$A \bar{\alpha} = \bar{\lambda} \bar{\alpha}$$

$$\lambda \alpha^T \bar{\alpha} = \left[\alpha^T A \right] \bar{\alpha} = \alpha^T \left[A \bar{\alpha} \right] = \bar{\lambda} \alpha^T \bar{\alpha}$$

由 $\alpha \neq 0$ 知 $\alpha^T \bar{\alpha} \neq 0$ ，故 $\lambda = \bar{\lambda}$ 是实数。

用到代数学基本定理

定理 1:

n 级实对称矩阵 (算代数重数) 有
 n 个实特征值.

以下的讨论 (求特征向量, 对角化) 都
限制在实数域上, 在欧氏空间中进行

A 在内积里可以挪来挪去

引理：

若 A 是 n 阶实对称矩阵， $\alpha, \beta \in \mathbb{R}^n$ ，
则

$$(A\alpha, \beta) = (\alpha, A\beta)$$

||

||

$$(\alpha^T A^T) \beta = \alpha^T (A \beta)$$

定理 2： 实对称矩阵属于不同特征值的特征向量彼此正交.

证： 设 λ_1, λ_2 是 A 的不同特征值,

$$A\alpha_1 = \lambda_1\alpha_1, \quad A\alpha_2 = \lambda_2\alpha_2.$$

由引理 $(A\alpha_1, \alpha_2) = (\alpha_1, A\alpha_2)$

得
$$\lambda_1 (\alpha_1, \alpha_2) = \lambda_2 (\alpha_1, \alpha_2)$$

故
$$(\alpha_1, \alpha_2) = 0$$

定理 3：实对称矩阵 都可写成

$$\mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^{-1} = \mathbf{P} \mathbf{D} \mathbf{P}^T$$

其中 \mathbf{P} 是正交矩阵， \mathbf{D} 实对角矩阵

(我们称 \mathbf{A} 可正交对角化).

反之，能写成 $\mathbf{P} \mathbf{D} \mathbf{P}^T$ 的矩阵都实对称

归纳法： 假设 $n-1$ 阶实对称矩阵都可正交对角化， 考察 n 阶实对称矩阵 A ：
设 λ_1 是 A 的实特征值， α_1 是 λ_1 的单位特征向量. 将 α_1 扩充成 \mathbb{R}^n 的一组基，
再正交化、单位化得到标准正交基

$$\alpha_1, \alpha_2, \cdots, \alpha_n$$

及正交矩阵 $P = [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n]$.

$$A [\alpha_1 \alpha_2 \cdots \alpha_n] = [\lambda_1 \alpha_1 \quad A\alpha_2 \cdots A\alpha_n]$$

$$= \underbrace{[\alpha_1 \alpha_2 \cdots \alpha_n]}_{\text{正交矩阵 } P} \begin{bmatrix} \lambda_1 & * & \text{C} & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & \text{B} & \vdots \\ 0 & * & \cdots & * \end{bmatrix}$$

$$P^T A P = \begin{bmatrix} \lambda_1 & \text{C} \\ 0 & \text{B} \end{bmatrix}$$

实对称矩阵

$$A [\alpha_1 \alpha_2 \cdots \alpha_n] = [\lambda_1 \alpha_1 \quad A\alpha_2 \cdots A\alpha_n]$$

$$= \underbrace{[\alpha_1 \alpha_2 \cdots \alpha_n]}_{\text{正交矩阵 } P} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & * & \cdots & * \\ \vdots & \vdots & \mathbf{B} & \vdots \\ 0 & * & \cdots & * \end{bmatrix}$$

$$P^T A P = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \mathbf{B} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & P_1 D_1 P_1^T \end{bmatrix}$$

实对称矩阵

归纳假设

$$\mathbf{A} = \mathbf{P} \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_1 \mathbf{D}_1 \mathbf{P}_1^T \end{bmatrix} \mathbf{P}^T$$

$$= \left(\mathbf{P} \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_1 \end{bmatrix} \right) \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_1 \end{bmatrix} \left(\begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_1^T \end{bmatrix} \mathbf{P}^T \right)$$

正交矩阵

实对角

正交的逆

$$\mathbf{P}^T \mathbf{A} \mathbf{P} = \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_1 \mathbf{D}_1 \mathbf{P}_1^T \end{bmatrix}$$

归纳假设

实对称矩阵

- 每个特征子空间中取基，正交化、单位化后得到标准正交基；
- 将各个特征子空间的标准正交基并在一起，得到 \mathbf{R}^n 的标准正交基及正交矩阵 \mathbf{P} ，则有

$$\mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^{-1} = \mathbf{P} \mathbf{D} \mathbf{P}^T$$

\mathbf{D} 是实对角矩阵，对角线上依次是 \mathbf{P} 的列向量的特征值

例. 求满足条件

$$\mathbf{A}^5 + 3\mathbf{A} - \mathbf{I} = \mathbf{0}$$

的所有 n 级实对称矩阵 \mathbf{A} .

证: 设实对称矩阵 \mathbf{A} 与实对角矩阵 \mathbf{D}

相似, 则 $\mathbf{D}^5 + 3\mathbf{D} - \mathbf{I} = \mathbf{0}$.

故 \mathbf{D} 的对角元 λ 都是

$$f(x) = x^5 + 3x - 1$$

的实根.

由于 $f'(x) = 5x^4 + 3 > 0$ 知

$f(x)$ 在 \mathbf{R} 上递增, 有唯一实根 c .

故 $\mathbf{D} = c\mathbf{I}$ 是数量矩阵, 于是 $\mathbf{A} = c\mathbf{I}$.

实对称矩阵的基本理论：

- 1) n 级实对称矩阵有 n 个实特征值；
- 2) 属于不同特征值的特征向量彼此正交；
- 3) 实对称矩阵都能正交对角化

$$A = P D P^{-1} = P D P^T$$

其中 P 是正交矩阵， D 实对角矩阵

例：将 n 级实对称矩阵对角化

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

特征向量的分布也应该呈一定规律

A 的特征向量为

$$\text{记 } \theta = \frac{\pi}{n+1}$$

$$\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \sin j\theta \\ \sin 2j\theta \\ \sin 3j\theta \\ \vdots \\ \sin nj\theta \end{bmatrix}$$

$$1 \leq j \leq n$$

$$= \begin{bmatrix} \sin 2j\theta \\ \sin j\theta + \sin 3j\theta \\ \sin 2j\theta + \sin 4j\theta \\ \vdots \\ \sin(n-1)j\theta + \sin(n+1)j\theta \\ \text{= } \sin j\pi = 0 \end{bmatrix} = 2 \cos j\theta \begin{bmatrix} \sin j\theta \\ \sin 2j\theta \\ \sin 3j\theta \\ \vdots \\ \sin nj\theta \end{bmatrix}$$

记 $\theta = \pi/(n+1)$

正弦正交矩阵

$$\mathbf{P} = \frac{\sqrt{2}}{\sqrt{n+1}} \begin{bmatrix} \sin \theta & \sin 2\theta & \cdots & \sin n\theta \\ \sin 2\theta & \sin 4\theta & \cdots & \sin 2n\theta \\ \vdots & \vdots & & \vdots \\ \sin n\theta & \sin 2n\theta & \cdots & \sin n^2\theta \end{bmatrix}$$

$$\mathbf{A} = \mathbf{P} \begin{bmatrix} 2\cos \theta & & & \\ & 2\cos 2\theta & & \\ & & \ddots & \\ & & & 2\cos n\theta \end{bmatrix} \mathbf{P}^{-1}$$

记 $\theta = \pi/(n+1)$

$$a \mathbf{I} + \mathbf{A} = \begin{bmatrix} a & 1 & 0 & \cdots & 0 \\ 1 & a & 1 & & 0 \\ 0 & 1 & a & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & 1 & a \end{bmatrix}$$

$$= \mathbf{P} \begin{bmatrix} a + 2\cos \theta & & & \\ & a + 2\cos 2\theta & & \\ & & \ddots & \\ & & & a + 2\cos n\theta \end{bmatrix} \mathbf{P}^T$$

记 $\theta = \pi/(n+1)$

$$|aI + A| = \begin{vmatrix} a & 1 & 0 & \cdots & 0 \\ 1 & a & 1 & & 0 \\ 0 & 1 & a & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & 1 & a \end{vmatrix}$$

$$= (a + 2\cos \theta)(a + 2\cos 2\theta) \cdots (a + 2\cos n\theta)$$

$$= \prod_{j=1}^n \left(a + 2\cos \frac{j\pi}{n+1} \right)$$

考试题型

例：设

$$\mathbf{A} = \begin{bmatrix} 1 & -2 & -4 \\ -2 & 4 & -2 \\ -4 & -2 & 1 \end{bmatrix}$$

求正交矩阵 \mathbf{P} 及实对角矩阵 \mathbf{D} ，使得

$$\mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^{-1} = \mathbf{P} \mathbf{D} \mathbf{P}^T.$$

解：先求 \mathbf{A} 的特征多项式

$$\begin{aligned}
 |\lambda \mathbf{I} - \mathbf{A}| &= \begin{vmatrix} \lambda-1 & 2 & 4 \\ 2 & \lambda-4 & 2 \\ 4 & 2 & \lambda-1 \end{vmatrix} \\
 &= \begin{vmatrix} \lambda-1 & 2 & 4 \\ 2 & \lambda-4 & 2 \\ 0 & -2\lambda+10 & \lambda-5 \end{vmatrix}
 \end{aligned}$$

Diagram illustrating the calculation of the characteristic polynomial $|\lambda \mathbf{I} - \mathbf{A}|$ for a 3x3 matrix \mathbf{A} . The matrix is shown as $\begin{bmatrix} \lambda-1 & 2 & 4 \\ 2 & \lambda-4 & 2 \\ 4 & 2 & \lambda-1 \end{bmatrix}$. The determinant is calculated by expanding along the first row, with the minor for the element 4 (row 1, column 3) highlighted in orange and labeled $\leftarrow -2$. The resulting 2x2 determinant is $\begin{vmatrix} 2 & 2 \\ 4 & \lambda-1 \end{vmatrix}$, which is also highlighted in orange. The final 3x3 determinant is shown as $\begin{vmatrix} \lambda-1 & 2 & 4 \\ 2 & \lambda-4 & 2 \\ 0 & -2\lambda+10 & \lambda-5 \end{vmatrix}$, with the last row and column highlighted in yellow.

解：先求 \mathbf{A} 的特征多项式

$$|\lambda \mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda - 1 & 2 & 4 \\ 2 & \lambda - 4 & 2 \\ 4 & 2 & \lambda - 1 \end{vmatrix}$$

$$= \begin{vmatrix} \lambda - 1 & 10 & 4 \\ 2 & \lambda & 2 \\ 0 & 0 & \lambda - 5 \end{vmatrix}$$

解：先求 \mathbf{A} 的特征多项式

$$|\lambda \mathbf{I} - \mathbf{A}| = (\lambda - 5) \begin{vmatrix} \lambda - 1 & 10 \\ 2 & \lambda \end{vmatrix}$$

$$= \begin{vmatrix} \lambda - 1 & 10 & 4 \\ 2 & \lambda & 2 \\ 0 & 0 & \lambda - 5 \end{vmatrix}$$

解:先求 \mathbf{A} 的特征多项式

$$\begin{aligned} |\lambda \mathbf{I} - \mathbf{A}| &= (\lambda - 5) \begin{vmatrix} \lambda - 1 & 10 \\ 2 & \lambda \end{vmatrix} \\ &= (\lambda - 5) (\lambda^2 - \lambda - 20) \\ &= (\lambda - 5)^2 (\lambda + 4) \end{aligned}$$

\mathbf{A} 的特征值为 $\lambda = 5$ (代数二重), -4 .

对 $\lambda = 5$, 解 $(5\mathbf{I} - \mathbf{A})\mathbf{X} = \mathbf{0}$:

$$\begin{bmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1/2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

得 $x_1 = -1/2 x_2 - x_3$, x_2, x_3 自由变量

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

$\lambda = 5$ 特征子空间的基为

$$\alpha_1 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \alpha_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

解 $\mathbf{A} \mathbf{X} = -4\mathbf{X}$, 得 $\lambda = -4$

的基

$$\alpha_3 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

不同特征子空间的
向量彼此正交



对 $\alpha_1 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \alpha_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ 正交化:

令 $\beta_1 = \alpha_1,$

$$\beta_2 = \alpha_2 - \frac{(\alpha_2, \beta_1)}{(\beta_1, \beta_1)} \beta_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} -4 \\ -2 \\ 5 \end{bmatrix}$$

再单位化:

$$\eta_1 = \frac{1}{\|\beta_1\|} \beta_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix},$$

$$\eta_2 = \frac{1}{\|\beta_2\|} \beta_2 = \frac{1}{\sqrt{45}} \begin{bmatrix} -4 \\ -2 \\ 5 \end{bmatrix}$$

$$\eta_3 = \frac{1}{\|\alpha_3\|} \alpha_3 = \frac{1}{3} \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -2 & -4 \\ -2 & 4 & -2 \\ -4 & -2 & 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -4 \end{bmatrix} \mathbf{P}^T$$

$$\mathbf{P} = [\eta_1 \ \eta_2 \ \eta_3] = \begin{bmatrix} -\frac{1}{\sqrt{5}} & -\frac{4}{\sqrt{45}} & \frac{2}{3} \\ \frac{2}{\sqrt{5}} & -\frac{2}{\sqrt{45}} & \frac{1}{3} \\ 0 & \frac{5}{\sqrt{45}} & \frac{2}{3} \end{bmatrix}$$

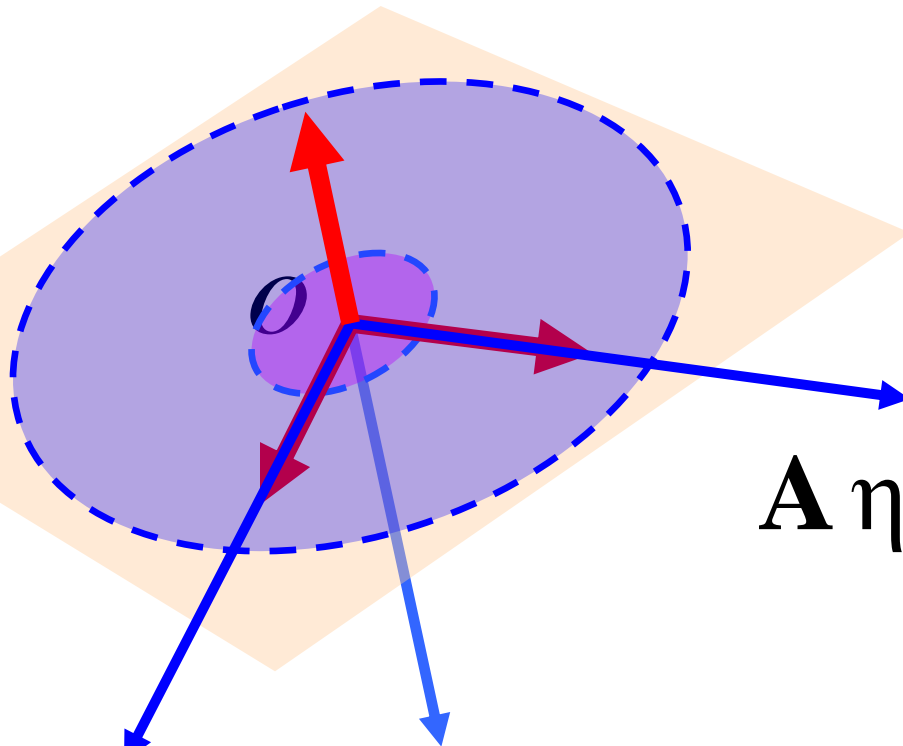
$$\mathbf{A} \mathbf{X} = -4 \mathbf{X}$$

$$\mathbf{A} \mathbf{X} = 5 \mathbf{X}$$

$$\mathbf{A} \boldsymbol{\eta}_1 = 5 \boldsymbol{\eta}_1$$

$$\mathbf{A} \boldsymbol{\eta}_2 = 5 \boldsymbol{\eta}_2$$

$$\mathbf{A} \boldsymbol{\eta}_3 = -4 \boldsymbol{\eta}_3$$



例: 1) 求二次齐次函数(二次型)

$$f(x_1, x_2, x_3) = x_1^2 + 4x_2^2 + x_3^2 \\ - 4x_1x_2 - 8x_1x_3 - 4x_2x_3$$

在单位球面 $x_1^2 + x_2^2 + x_3^2 = 1$ 上取到的最大、最小值, 在何处取到?

2) 求二次曲面 $f(x_1, x_2, x_3) = \pm 1$ 的分类, 是椭球面? 抛物面? 单(双)叶双曲面?

$$f = x_1^2 + 4x_2^2 + x_3^2 - 4x_1x_2 - 8x_1x_3 - 4x_2x_3$$

$$= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 1 & -2 & -4 \\ -2 & 4 & -2 \\ -4 & -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

A 实对称矩阵

$$= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \mathbf{P} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -4 \end{bmatrix} \mathbf{P}^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

令 $\begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}$

作变量(正交)替换 $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = P \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ 后,

$$\begin{aligned} f &= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} P \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -4 \end{bmatrix} P^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \\ &= 5y_1^2 + 5y_2^2 - 4y_3^2 \end{aligned}$$

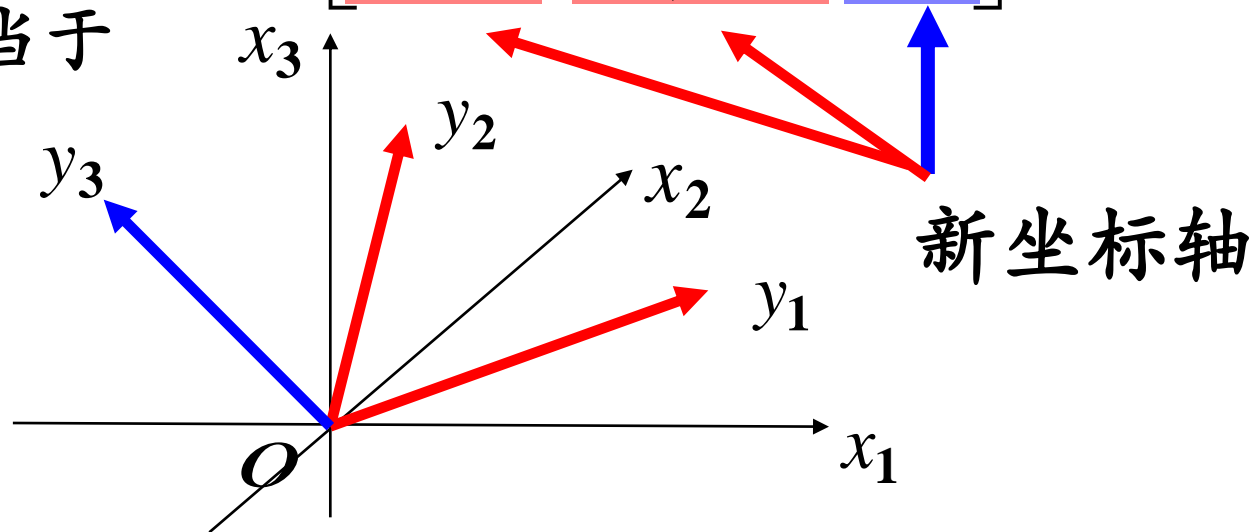
做正交替换相当于以 P 的列向量为轴建立

新直角坐标系 Y

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = P \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{4}{\sqrt{45}} & \frac{2}{3} \\ \frac{2}{\sqrt{5}} & -\frac{2}{\sqrt{45}} & \frac{1}{3} \\ 0 & \frac{5}{\sqrt{45}} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Y 坐标 $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ 相当于

X 坐标 $\begin{bmatrix} 2/3 \\ 1/3 \\ 2/3 \end{bmatrix}$



由于 \mathbf{P} 是正交矩阵,

$$\begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \mathbf{P} \qquad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \mathbf{P}^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

我们有

$$\begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \mathbf{P} \mathbf{P}^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$y_1^2 + y_2^2 + y_3^2 = x_1^2 + x_2^2 + x_3^2$$

$$\text{故 } y_1^2 + y_2^2 + y_3^2 = 1 \Leftrightarrow x_1^2 + x_2^2 + x_3^2 = 1$$

$$f = x_1^2 + 4x_2^2 + x_3^2 - 4x_1x_2 - 8x_1x_3 - 4x_2x_3$$

$$= 5y_1^2 + 5y_2^2 - 4y_3^2$$

$$= 5(y_1^2 + y_2^2 + y_3^2) - 9y_3^2$$

$$\leq 5$$

等号成立当且仅当 $y_3 = 0$

等号成立当且仅当 $y_3 = 0$, 即

$$\mathbf{P} = [\eta_1 \ \eta_2 \ \eta_3]$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{5}} & -\frac{4}{\sqrt{45}} & \frac{2}{3} \\ \frac{2}{\sqrt{5}} & -\frac{2}{\sqrt{45}} & \frac{1}{3} \\ 0 & \frac{5}{\sqrt{45}} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ 0 \end{bmatrix} = y_1 \eta_1 + y_2 \eta_2$$

在 $\lambda = 5$ 的特征子空间与单位球面
的交集上取到最大值 5

当 $y_1^2 + y_2^2 + y_3^2 = x_1^2 + x_2^2 + x_3^2 = 1$ 时,

$$\begin{aligned} & x_1^2 + 4x_2^2 + x_3^2 - 4x_1x_2 - 8x_1x_3 - 4x_2x_3 \\ &= 5y_1^2 + 5y_2^2 - 4y_3^2 \\ &= -4(y_1^2 + y_2^2 + y_3^2) + 9y_1^2 + 9y_2^2 \\ &\geq -4 \end{aligned}$$

等号成立当且仅当 $y_1 = y_2 = 0$

等号成立当且仅当 $y_1 = y_2 = 0$

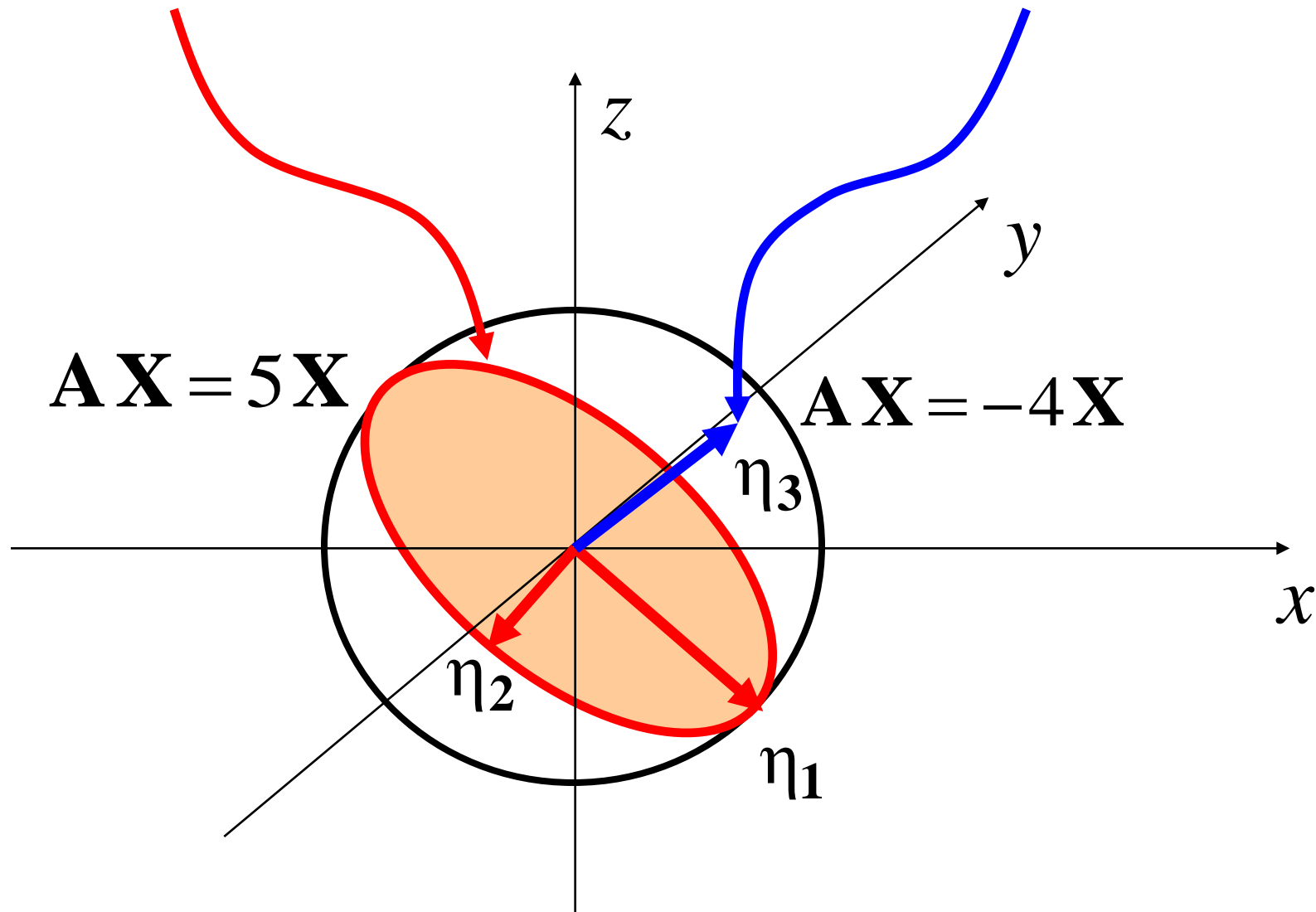
$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{5}} & -\frac{4}{\sqrt{45}} & \frac{2}{3} \\ \frac{2}{\sqrt{5}} & -\frac{2}{\sqrt{45}} & \frac{1}{3} \\ 0 & \frac{5}{\sqrt{45}} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ y_3 \end{bmatrix} = y_3 \eta_3$$

在 $\lambda = -4$ 的特征子空间与单位球面

的交集上取到最小值 -4

f 取到最大值 5

f 取到最小值 -4



设 A 是实对称矩阵, 且有正交分解

$$A = \underbrace{[\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n]}_{\text{正交矩阵 } P} \underbrace{\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}}_{\text{实对角矩阵 } D} \underbrace{\begin{bmatrix} \alpha_1^T \\ \alpha_2^T \\ \vdots \\ \alpha_n^T \end{bmatrix}}_{\substack{P^{-1} \\ = P^T}}$$

$$\lambda_1 = \cdots = \lambda_s > \lambda_{s+1} \geq \cdots \geq \lambda_n$$

$$\lambda_1 = \cdots = \lambda_s > \lambda_{s+1} \geq \cdots \geq \lambda_n$$

令 $\mathbf{X} = \mathbf{P} \mathbf{Y} = y_1 \boldsymbol{\alpha}_1 + \cdots + y_n \boldsymbol{\alpha}_n$, 则

$$\mathbf{X}^T \mathbf{A} \mathbf{X} = \mathbf{X}^T \mathbf{P} \mathbf{D} \mathbf{P}^T \mathbf{X} = \mathbf{Y}^T \mathbf{D} \mathbf{Y}$$

$$= \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$$

$$= \lambda_1 (y_1^2 + \cdots + y_n^2) - (\lambda_1 - \lambda_2) y_2^2 - \cdots$$

$$- (\lambda_1 - \lambda_{s+1}) y_{s+1}^2 - \cdots - (\lambda_1 - \lambda_n) y_n^2$$

$$\leq \lambda_1 (y_1^2 + \cdots + y_n^2) = \lambda_1 \|\mathbf{X}\|^2$$

$$\text{等号成立} \iff y_{s+1} = \cdots = y_n = 0$$

若 $\mathbf{X} = y_1 \alpha_1 + \cdots + y_n \alpha_n = \mathbf{P} \mathbf{Y}$, 则有

$$\mathbf{X}^T \mathbf{A} \mathbf{X} \leq \lambda_1 \|\mathbf{X}\|^2$$

等号成立 $\Leftrightarrow \mathbf{X} = y_1 \alpha_1 + \cdots + y_s \alpha_s$

即 \mathbf{X} 落在最大特征值 λ_1 的特征子空间内,

$$\mathbf{A} \mathbf{X} = \lambda_1 \mathbf{X}.$$

$$\mathbf{A} \mathbf{X} = \lambda_n \mathbf{X}$$

$$\mathbf{A} \mathbf{X} = \lambda_1 \mathbf{X}$$

定理:

对任意列向量 $\mathbf{X} \in \mathbf{R}^n$, 有

$$\lambda_n \|\mathbf{X}\|^2 \leq \mathbf{X}^T \mathbf{A} \mathbf{X} \leq \lambda_1 \|\mathbf{X}\|^2$$

等号成立 当且仅当 \mathbf{X} 落在最大, 最小特征值 λ_1, λ_n 的特征子空间内.

$$\mathbf{f}(\mathbf{X}) = \mathbf{f}\left(\|\mathbf{X}\| \frac{\mathbf{X}}{\|\mathbf{X}\|}\right) = \mathbf{f}\left(\frac{\mathbf{X}}{\|\mathbf{X}\|}\right) \|\mathbf{X}\|^2$$

例：设二次曲面 S 在 X -坐标下的方程为

$$x_1^2 + 4x_2^2 + x_3^2 - 4x_1x_2 - 8x_1x_3 - 4x_2x_3 = 1$$

这是一个什么曲面？椭球面，抛物面，

圆锥面还是单(双)叶双曲面？

思路：建立新直角坐标系

$$x_1^2 + 4x_2^2 + x_3^2 - 4x_1x_2 - 8x_1x_3 - 4x_2x_3$$

$$= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 1 & -2 & -4 \\ -2 & 4 & -2 \\ -4 & -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \mathbf{P} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -4 \end{bmatrix} \mathbf{P}^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

令 $\begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

作正交变量替换 $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{P} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$

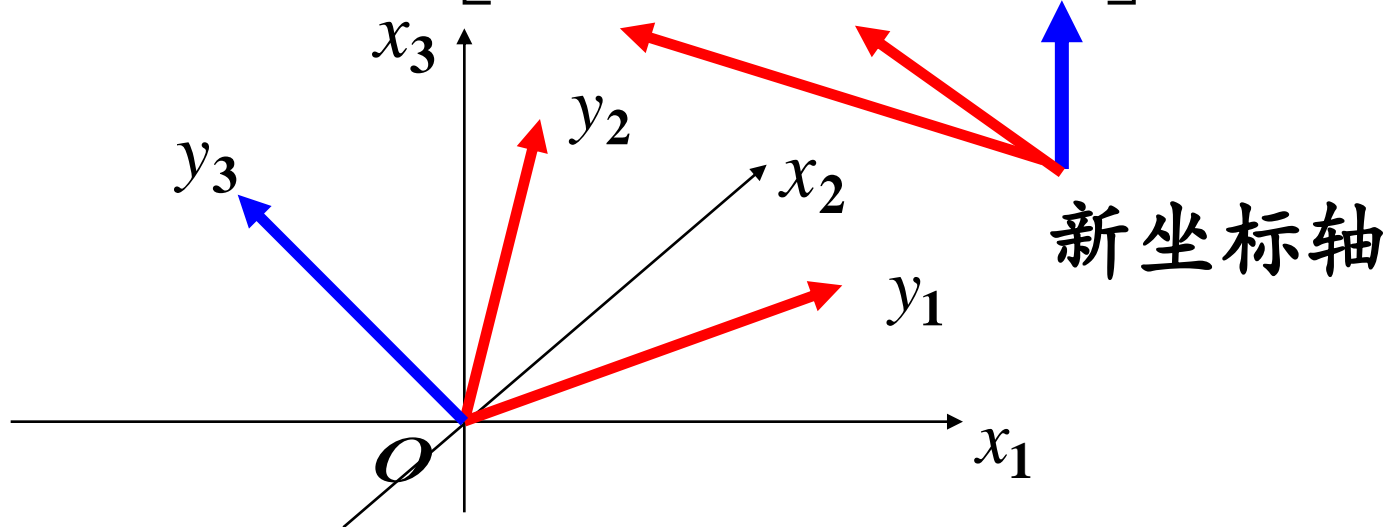
$$\begin{aligned} x_1^2 + 4x_2^2 + x_3^2 - 4x_1x_2 - 8x_1x_3 - 4x_2x_3 \\ = 5y_1^2 + 5y_2^2 - 4y_3^2 \end{aligned}$$

在新坐标 $\mathbf{Y} = \mathbf{P}^T \mathbf{X}$ 下, 二次曲面 S 的方程为

$$5y_1^2 + 5y_2^2 - 4y_3^2 = 1$$

以 \mathbf{P} 的列向量为坐标轴建立新直角坐标系 \mathbf{Y}

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{P} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{4}{\sqrt{45}} & \frac{2}{3} \\ \frac{2}{\sqrt{5}} & -\frac{2}{\sqrt{45}} & \frac{1}{3} \\ 0 & \frac{5}{\sqrt{45}} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$



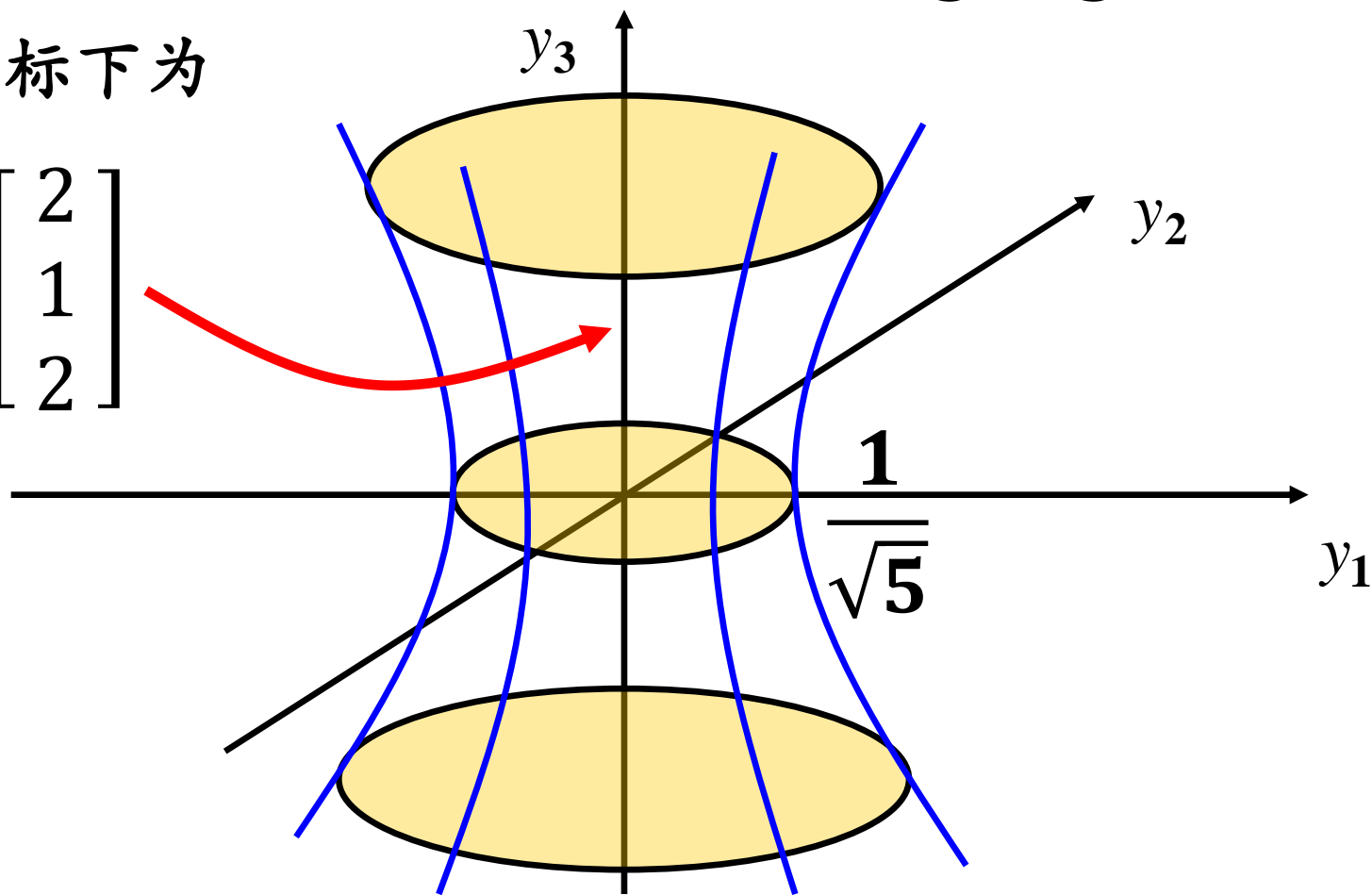
二次曲面S是单叶双曲面

曲面对称轴

在 X 坐标下为

$$\frac{1}{3} \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

$$y_1^2 + y_2^2 = \frac{1}{5} + \frac{4}{5} y_3^2$$



实对称矩阵正交对角化应用

1. 实矩阵的奇异值分解 (SVD)
2. 主成分分析 (PCA)



例:

$$A = \begin{bmatrix} 2 & -1 & 5 & 2 & -1 \\ 4 & -1 & 9 & 3 & 4 \\ 3 & -2 & 8 & -2 & 1 \\ 1 & 1 & 1 & 4 & 4 \\ 3 & 1 & 5 & 8 & 6 \end{bmatrix}$$

满秩分解

主元列向量

简化阶梯形的
的非零行

$$= \begin{bmatrix} 2 & -1 & 2 \\ 4 & -1 & 3 \\ 3 & -2 & -2 \\ 1 & 1 & 4 \\ 3 & 1 & 8 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 2 & 0 & 3 \\ 0 & 1 & -1 & 0 & 5 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 5 & 2 & -1 \\ 4 & -1 & 9 & 3 & 4 \\ 3 & -2 & 8 & -2 & 1 \\ 1 & 1 & 1 & 4 & 4 \\ 3 & 1 & 5 & 8 & 6 \end{bmatrix}$$

分块乘法

$$= \begin{bmatrix} 2 \\ 4 \\ 3 \\ 1 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 0 & 3 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \\ -2 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & -1 & 0 & 5 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \\ -2 \\ 4 \\ 8 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

引理 1. 若 A 是实矩阵, 则实对称矩阵 $A^T A$ 的特征值都 ≥ 0 (即 $A^T A$ 半正定).

证: 设 $\lambda \in \mathbf{R}$ 是 $A^T A$ 的一个特征值, α 是属于 λ 的一个实特征向量, 即

$$A^T A \alpha = \lambda \alpha, \quad 0 \neq \alpha \in \mathbf{R}^n.$$

则 $\alpha^T A^T A \alpha = (A \alpha)^T (A \alpha) = \|A \alpha\|^2 = \lambda \alpha^T \alpha,$

于是有

$$\lambda = \frac{\|A \alpha\|^2}{\|\alpha\|^2} \geq 0.$$

定义. 设 A 是一个秩为 r 的 $m \times n$ 实矩阵, 则
实对称矩阵 $A^T A$ 与 AA^T 有相同的正特征值

$$\lambda_1 \geq \cdots \geq \lambda_r > 0.$$

这些正特征值的算术平方根(从大到小排列)

$$\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_r}$$

称为矩阵 A 的奇异值, 通常用 $\sigma_1 \geq \cdots \geq \sigma_r$ 表示.

性质: 左乘, 右边乘正交矩阵, 矩阵的奇异值不变.

$$((PAQ)^T(PAQ) = Q^T A^T A Q \text{ 与 } A^T A \text{ 相似})$$

Singular Value Decomposition

定理：每个 $m \times n$ 实矩阵 A 都可分解为

$$A = P S Q^T$$

的形式，其中 P, Q 分别是 m 级, n 级正交矩阵，

$$S = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix}_{m,n}.$$

这里 $r = A$ 秩, $\sigma_1 \geq \dots \geq \sigma_r > 0$ 是 A 的奇异值.

矩阵的奇异值分解(SVD)

定理：每个 $m \times n$ 实矩阵 A 都可写成

$$A = \underbrace{[\beta_1 \ \beta_2 \ \cdots \ \beta_m]}_{\mathbf{P} \text{ 正交}} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \mathbf{0} \end{bmatrix} \underbrace{\begin{bmatrix} \gamma_1^T \\ \gamma_2^T \\ \vdots \\ \gamma_n^T \end{bmatrix}}_{\mathbf{Q}^T \text{ 正交}}$$

$$= \sigma_1 \beta_1 \gamma_1^T + \sigma_2 \beta_2 \gamma_2^T + \cdots + \sigma_r \beta_r \gamma_r^T$$

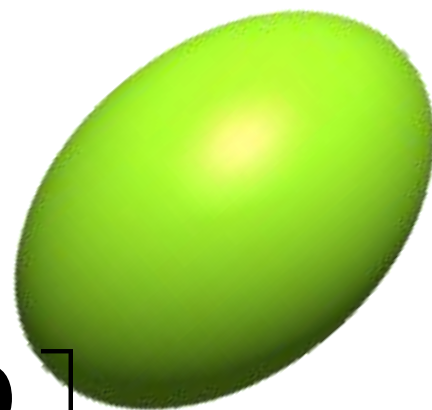
$$A \gamma_i = \sigma_i \beta_i, \quad A^T \beta_i = \sigma_i \gamma_i, \quad i = 1, \cdots, r$$

SVD 的几何意义

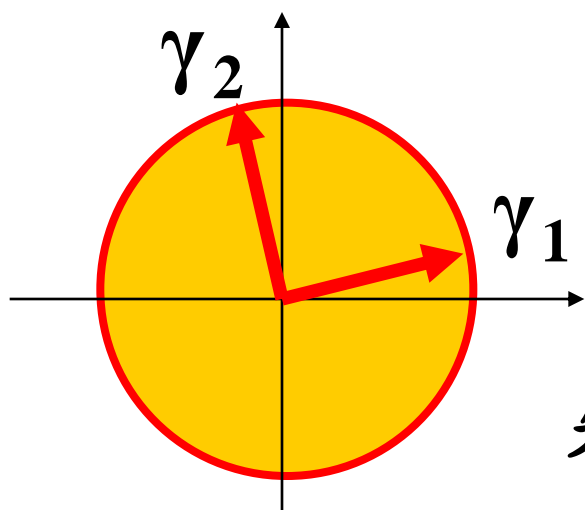
$$\mathbf{X} \mapsto \mathbf{A} \mathbf{X}$$



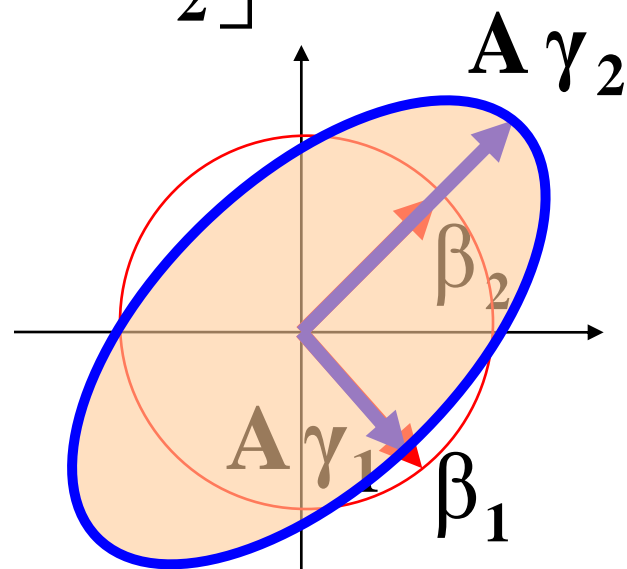
\mathbf{A} 实矩阵



$$\mathbf{A}[\gamma_1 \ \gamma_2] = [\beta_1 \ \beta_2] \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$



先扭再拉伸



$$(\gamma_1, \gamma_2) \rightarrow (\beta_1, \beta_2) \rightarrow (\sigma_1 \beta_1, \sigma_2 \beta_2)$$

$\mathbf{A}^T \mathbf{A}$ 特征向量

$\mathbf{A} \mathbf{A}^T$ 特征向量

极分解定理:

每个实方阵都是一个(半正定)实对称矩阵
与一个正交矩阵的乘积.

$$\mathbf{A} = \mathbf{P} \mathbf{S} \mathbf{Q}^T = \underbrace{(\mathbf{P} \mathbf{S} \mathbf{P}^T)}_{\text{半正定}} \underbrace{(\mathbf{P} \mathbf{Q}^T)}_{\text{正交}}$$

$$(\gamma_1, \dots, \gamma_n) \rightarrow (\beta_1, \dots, \beta_n) \rightarrow (\sigma_1 \beta_1, \dots, \sigma_n \beta_n)$$

正交变换

对称变换

奇异值 $\sigma_1 \geq \cdots \geq \sigma_r > 0$ 的几何意义

设 $A = P S Q^T = [\beta_1 \cdots \beta_m] S [\gamma_1 \cdots \gamma_n]^T$, 则有

$$AQ = [A\gamma_1 \cdots A\gamma_n] = P S = [\sigma_1 \beta_1 \cdots \sigma_r \beta_r \ 0 \cdots 0].$$

线性映射 $X \mapsto AX$ 将 \mathbb{R}^n 中的 n 维单位球体

$\|X\| \leq 1$ 映成 \mathbb{R}^m 中的 r 维椭球体. 椭球的长短

半轴由 $\sigma_1 \beta_1 = A\gamma_1, \cdots, \sigma_r \beta_r = A\gamma_r$ 给出. 故

A 的 r 个奇异值就是椭球半轴长度, 反映映射的奇异程度.

SVD 证明: 设 A 是秩 r 的 $m \times n$ 实矩阵.

由于 $A^T A$ 是实对称矩阵, 故存在正交矩阵 Q , 使得

$$A^T A = \underbrace{[\gamma_1 \cdots \gamma_r \cdots \gamma_n]}_{Q \text{ 正交}} \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_r & \\ & & & 0 \end{bmatrix} Q^T$$

由引理 1, $A^T A$ 的特征值非负

$$\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0.$$

将 \mathbf{Q} 与 \mathbf{Q}^T 移到等号左边, 得到对角矩阵

$$\mathbf{Q}^T \mathbf{A}^T \mathbf{A} \mathbf{Q} = (\mathbf{A}\mathbf{Q})^T \mathbf{A}\mathbf{Q} = \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_r & \\ & & & \mathbf{0} \end{bmatrix}$$

注意左边是 $\mathbf{A}\mathbf{Q} = [\mathbf{A}\boldsymbol{\gamma}_1 \cdots \mathbf{A}\boldsymbol{\gamma}_n]$ 的 Gram 矩阵:

$$[\mathbf{A}\boldsymbol{\gamma}_1 \cdots \mathbf{A}\boldsymbol{\gamma}_n]^T [\mathbf{A}\boldsymbol{\gamma}_1 \cdots \mathbf{A}\boldsymbol{\gamma}_n] = [(\mathbf{A}\boldsymbol{\gamma}_i, \mathbf{A}\boldsymbol{\gamma}_j)]_{1 \leq i, j \leq n}.$$

这说明 $\mathbf{A}\mathbf{Q}$ 的列向量 $\mathbf{A}\boldsymbol{\gamma}_1, \cdots, \mathbf{A}\boldsymbol{\gamma}_n \in \mathbf{R}^m$ 两两正交, 且

$$\|\mathbf{A}\boldsymbol{\gamma}_1\| = \sqrt{\lambda_1}, \cdots, \|\mathbf{A}\boldsymbol{\gamma}_r\| = \sqrt{\lambda_r}, \mathbf{A}\boldsymbol{\gamma}_{r+1}, \cdots, \mathbf{A}\boldsymbol{\gamma}_n = \mathbf{0}$$

令

$$\beta_i = \frac{1}{\sqrt{\lambda_i}} A \gamma_i = \frac{1}{\sigma_i} A \gamma_i, \quad i = 1, \dots, r;$$

由于 β_1, \dots, β_r 是单位正交向量组, 故可扩充成 \mathbf{R}^m 的标准正交基 $\beta_1, \dots, \beta_r, \dots, \beta_m$. 于是

$$\begin{aligned} A Q &= [A \gamma_1 \cdots A \gamma_r \ 0 \cdots 0] \\ &= [\sigma_1 \beta_1 \cdots \sigma_r \beta_r \ 0 \cdots 0] \\ &= [\beta_1 \cdots \beta_r \cdots \beta_m] \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix}_{m,n} \end{aligned}$$

记正交矩阵 $\mathbf{P} = [\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_m]$, 则有

$$\begin{aligned}
 \mathbf{A} &= \mathbf{P} \begin{bmatrix} \boldsymbol{\sigma}_1 & & & \\ & \ddots & & \\ & & \boldsymbol{\sigma}_r & \\ & & & \mathbf{0} \end{bmatrix}_{m,n} \mathbf{Q}^T \\
 &= [\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_r \cdots \boldsymbol{\beta}_m] \begin{bmatrix} \boldsymbol{\sigma}_1 & & & \\ & \ddots & & \\ & & \boldsymbol{\sigma}_r & \\ & & & \mathbf{0} \end{bmatrix}_{m,n} \begin{bmatrix} \boldsymbol{\gamma}_1^T \\ \boldsymbol{\gamma}_2^T \\ \vdots \\ \boldsymbol{\gamma}_n^T \end{bmatrix} \\
 &= \boldsymbol{\sigma}_1 \boldsymbol{\beta}_1 \boldsymbol{\gamma}_1^T + \boldsymbol{\sigma}_2 \boldsymbol{\beta}_2 \boldsymbol{\gamma}_2^T + \cdots + \boldsymbol{\sigma}_r \boldsymbol{\beta}_r \boldsymbol{\gamma}_r^T
 \end{aligned}$$

注:

在矩阵的 **SVD** 分解 $A = P S Q^T$ 中, 正交矩阵 P, Q 取法不唯一.

一个极端的例子是 A 为单位矩阵的情况, 此时

$A = P I P^T$, $P (= Q)$ 可以取任意正交矩阵.

即使在 A 可逆且奇异值互异的情况, P 的每个列向量也可以相差一个正负号.

若 $A = P S Q^T$ 是 A 的奇异值分解, $P = [\beta_1 \cdots \beta_m]$,

$Q = [\gamma_1 \cdots \gamma_n]$ 是正交矩阵, S 是奇异值矩阵,

则有 $AA^T = P S S^T P^T$, $A^T A = Q S^T S Q^T$.

故 β_1, \cdots, β_m 与 $\gamma_1, \cdots, \gamma_n$ 分别是 AA^T 与 $A^T A$ 的特征向量, 称为 A 的左, 右奇异向量.

这里起关键作用的是 β_1, \cdots, β_r 和 $\gamma_1, \cdots, \gamma_r$.

由 $AQ = P S$ 知它们之间有绑定关系

$$A\gamma_i = \sigma_i \beta_i, A^T \beta_i = \sigma_i \gamma_i, i = 1, \cdots, r.$$

进一步比较维数知, β_1, \dots, β_r 与 $\gamma_1, \dots, \gamma_r$ 分别构成 A 列空间与行空间的标准正交基. 而且由 PCA 定理知它们是最能拟合 A 的列向量, 行向量 (看成空间里的点) 分布的正交基: β_1 是 A 列向量分布的主方向, β_1, β_2 生成 A 列向量分布的主平面...; $\gamma_1, \gamma_2, \dots$ 也有类似的性质.

最后 $\beta_{r+1}, \dots, \beta_m$ 与 $\gamma_{r+1}, \dots, \gamma_n$ 分别构成 $\text{Ker}A^T$ 与 $\text{Ker}A$ 的标准正交基 (这个基可独立取, 不唯一).

A 的奇异值 $\sigma_1 \geq \dots \geq \sigma_r > 0$ 是 AA^T (或 $A^T A$) 正特征值的平方根. 一般来说, 排在后面的奇异值会越来越接近 0. 省略这些小值, 只保留前 k 个奇异值, 就得到矩阵 A 在秩 $\leq k$ 限制下的最佳逼近

$$A_k = \sigma_1 \beta_1 \gamma_1^T + \dots + \sigma_k \beta_k \gamma_k^T \rightsquigarrow A$$

用 $\| \cdot \|_F$ 表示矩阵的 Frobenius 范数, 我们有
定理: 对任意秩 $\leq k$ 的矩阵 $M \in M_{m,n}(\mathbb{R})$, 有

$$\| A - M \|_F^2 \geq \| A - A_k \|_F^2 = \sigma_{k+1}^2 + \dots + \sigma_r^2.$$

$$\begin{aligned} \mathbf{A}_k &= \sigma_1 \boldsymbol{\beta}_1 \boldsymbol{\gamma}_1^T + \cdots + \sigma_k \boldsymbol{\beta}_k \boldsymbol{\gamma}_k^T \\ &= \boldsymbol{\beta}_1 \boldsymbol{\beta}_1^T \mathbf{A} + \cdots + \boldsymbol{\beta}_k \boldsymbol{\beta}_k^T \mathbf{A} = \mathbf{B}_k \mathbf{B}_k^T \mathbf{A}. \end{aligned}$$

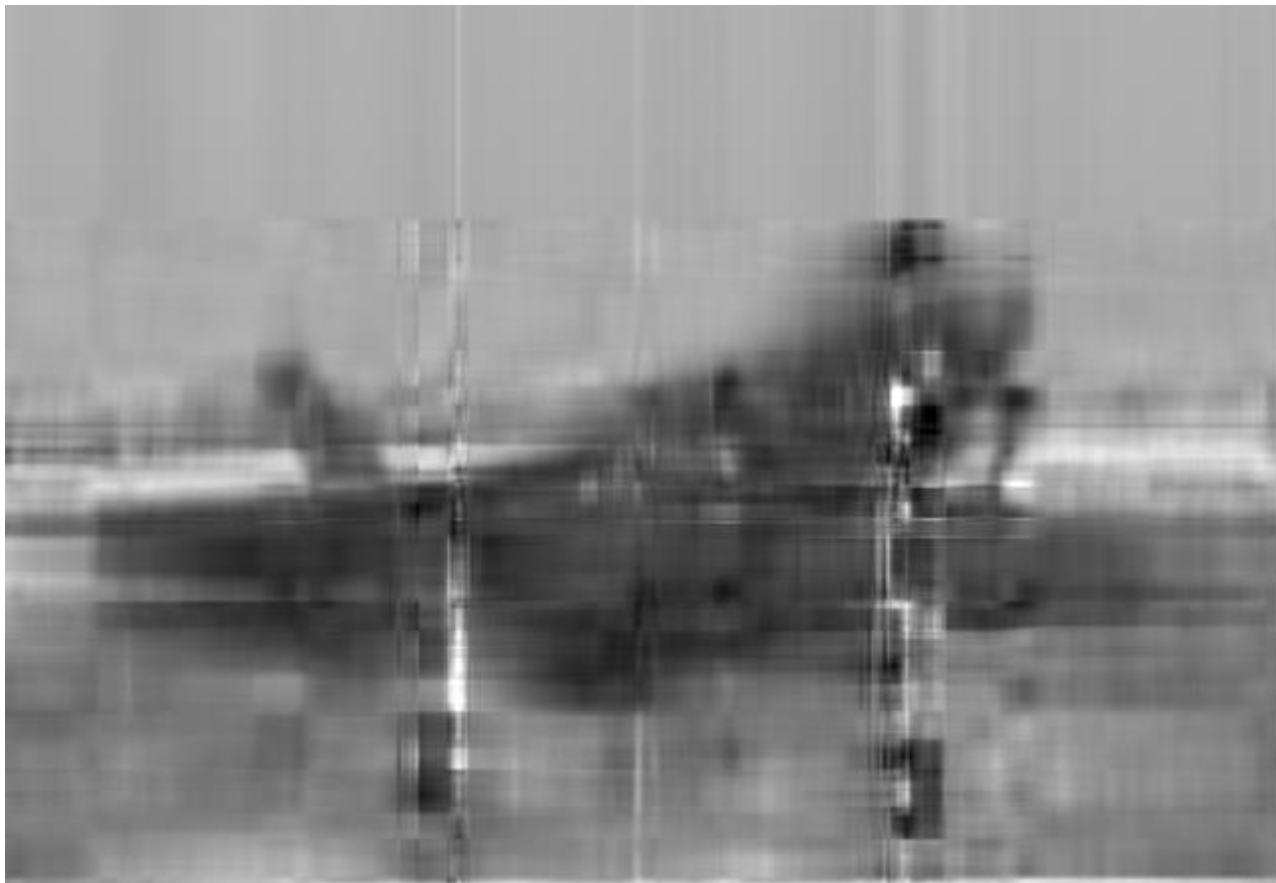
这里记 $\mathbf{B}_k = [\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_k]$. 注意 $\mathbf{B}_k \mathbf{B}_k^T$ 是从 \mathbf{R}^m 向 \mathbf{B}_k 的列空间 $\langle \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_k \rangle$ 作正交投影的矩阵.

图像压缩实验



A : 440×640 灰度矩阵

秩 K=10 逼近



$$\sigma_1 \beta_1 \gamma_1^T + \dots + \sigma_{10} \beta_{10} \gamma_{10}^T$$

秩 $K=30$ 逼近



$$\sigma_1 \beta_1 \gamma_1^T + \dots + \sigma_{30} \beta_{30} \gamma_{30}^T$$

秩 K=50 逼近



$$\sigma_1 \alpha_1 \beta_1^T + \dots + \sigma_{50} \alpha_{50} \beta_{50}^T$$

秩 $K=100$ 逼近



$$\sigma_1 \beta_1 \gamma_1^T + \dots + \sigma_{100} \beta_{100} \gamma_{100}^T$$

作业: 12 月 26 日 交

§ 6.2 6, 8

§ 6.3 4, 6, 7, 9, 10, 11, 19

补充题: 1, 2, 3, 4

1. 设 A 是 n 级实对称矩阵, V 是 R^n 的一个 k -维子空间. 取 V 的一组标准正交基 $\gamma_1, \dots, \gamma_k$, 并记 $B = [\gamma_1 \cdots \gamma_k]$. 设实对称矩阵 $B^T A B$ 的特征值为 $\mu_1 \geq \dots \geq \mu_k$.

证明:

1) $\mu_1 \geq \dots \geq \mu_k$ 只依赖于 V , 与标准正交基 $\gamma_1, \dots, \gamma_k$ 的选取无关;

2) 对于 $0 \neq \alpha \in V$, Rayleigh 商满足不等式

$$\mu_k \leq \frac{\alpha^T A \alpha}{\alpha^T \alpha} \leq \mu_1 ,$$

确定最大, 最小值分别在 V 的哪些地方取到.

注: $B^T A B$ 与 $B B^T A$ 不仅有相同的迹, 还有相同的非零特征值.

3) 设实对称矩阵 A 的特征值为

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n .$$

则有 $\lambda_1 \geq \mu_1 \geq \lambda_{n-k+1} ,$

$$\lambda_2 \geq \mu_2 \geq \lambda_{n-k+2} ,$$

$$\cdots \quad \cdots ,$$

$$\lambda_k \geq \mu_k \geq \lambda_n .$$

2. 设 n 级实对称矩阵 A 的特征值为

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n .$$

设 B 是 A 的一个 $n-1$ 级主子阵, 特征值为

$$\mu_1 \geq \mu_2 \cdots \geq \mu_{n-1} .$$

证明: $\lambda_k \geq \mu_k \geq \lambda_{k+1}$, $\forall 1 \leq k < n$.

Cauchy Interlacing 定理

3. (~~Cauchy Interlacing~~ 定理)

设 n 级实对称矩阵 A 的特征值为

$$\lambda_1 \geq \cdots \geq \lambda_n .$$

又设 α 是一个单位向量, $t \geq 0$, 矩阵

$A + t \alpha \alpha^T$ 的特征值记为 $\mu_1 \geq \cdots \geq \mu_n$.

证明:

$$\lambda_1 + t \geq \mu_1 \geq \lambda_1 \geq \mu_2 \geq \lambda_2 \geq \cdots \geq \mu_n \geq \lambda_n$$

4. (广义瑞利商) 设 A, B 是 n 级实对称矩阵且 A 正定. 证明:

$$\max_{0 \neq \alpha \in \mathbb{R}^n} \frac{\alpha^T B \alpha}{\alpha^T A \alpha} = \mu,$$

这里 μ 是 $B A^{-1}$ 的最大实特征值. 确定这个最大值在哪里取到.

(容易看出, $B A^{-1}$ 的特征值都是实数).

Linear Discriminant Analysis

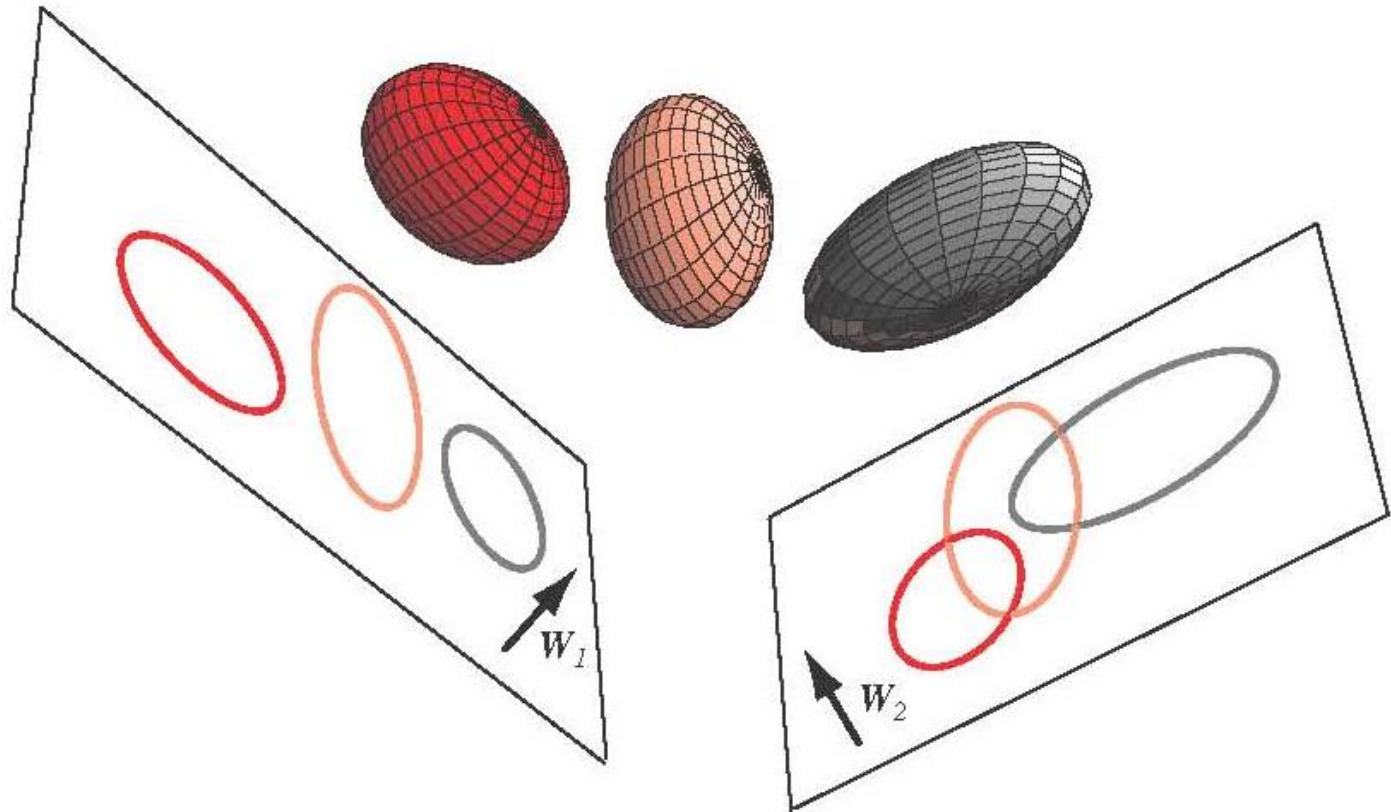


FIGURE 3.6. Three three-dimensional distributions are projected onto two-dimensional subspaces, described by a normal vectors \mathbf{W}_1 and \mathbf{W}_2 . Informally, multiple discriminant methods seek the optimum such subspace, that is, the one with the greatest separation of the projected distributions for a given total within-scatter matrix, here as associated with \mathbf{W}_1 . From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

例(线性判别分析LDA).

设 $\alpha_1, \dots, \alpha_s$ 与 β_1, \dots, β_t 是 \mathbb{R}^m 中的两组点. 记

$$\tilde{\alpha} = \frac{1}{s}(\alpha_1 + \dots + \alpha_s), \tilde{\beta} = \frac{1}{t}(\beta_1 + \dots + \beta_t).$$

求向量 $0 \neq \gamma \in \mathbb{R}^m$, 使得向直线 $\langle \gamma \rangle$ 作正交投影后, 两组点 $\alpha_1, \dots, \alpha_s$ 与 β_1, \dots, β_t 的像之间能尽可能分离. 具体来说, 希望 $\tilde{\alpha}$ 的像与 $\tilde{\beta}$ 的像之间距离尽可能大, 而 $\alpha_i - \tilde{\alpha}$ 与 $\beta_j - \tilde{\beta}$ 像的长度平方和(散度)尽可能小.

问题可阐述为: 求单位向量 $\gamma \in \mathbb{R}^m$, 使得

$$D(\gamma) = \frac{\|\gamma^T \tilde{\alpha} - \gamma^T \tilde{\beta}\|^2}{\sum_{1 \leq i \leq s} \|\gamma^T (\alpha_i - \tilde{\alpha})\|^2 + \sum_{1 \leq j \leq t} \|\gamma^T (\beta_j - \tilde{\beta})\|^2}$$

取最大值. 记矩阵 $S_b = (\tilde{\alpha} - \tilde{\beta})(\tilde{\alpha} - \tilde{\beta})^T$,

$$S_w = \sum_{1 \leq i \leq s} (\alpha_i - \tilde{\alpha})(\alpha_i - \tilde{\alpha})^T + \sum_{1 \leq j \leq t} (\beta_j - \tilde{\beta})(\beta_j - \tilde{\beta})^T.$$

问题最后归结为

求向量 $0 \neq \gamma \in \mathbb{R}^m$, 使得广义瑞利商

$$D(\gamma) = \frac{\gamma^T S_b \gamma}{\gamma^T S_w \gamma}$$

取最大值.

实对称矩阵正交对角化应用

1. 实矩阵的奇异值分解 (SVD)
2. 主成分分析 (PCA)



例. 从乳腺肿瘤的观察数据和最终诊断结果,
学习判断 哪些肿瘤是良性的, 哪些是
恶性的.



worst texture	worst perimeter	worst area	worst smoothness	worst compactness	worst concavity	worst concave points	worst symmetry	worst fractal dimension	cancer
30.88	186.80	2398.0	0.1512	0.3150	0.53720	0.23880	0.2768	0.07615	0
19.26	99.70	711.2	0.1440	0.1773	0.23900	0.12880	0.2977	0.07259	1
20.49	96.09	630.5	0.1312	0.2776	0.18900	0.07283	0.3184	0.08183	1
15.66	65.13	314.9	0.1324	0.1148	0.08867	0.06227	0.2450	0.07773	1
19.08	125.10	980.9	0.1390	0.5954	0.63050	0.23930	0.4667	0.09946	0

例. 图片的灰度矩阵转化成列向量作分类识别



$M_{640,480}$



R^{307200}



人脸子空间

predicted: George W Bush
true: George W Bush



predicted: George W Bush
true: George W Bush



predicted: Tony Blair
true: Tony Blair



predicted: George W Bush
true: George W Bush



predicted: George W Bush
true: George W Bush



predicted: George W Bush
true: George W Bush



predicted: Gerhard Schroeder
true: Gerhard Schroeder



predicted: Colin Powell
true: Colin Powell



predicted: George W Bush
true: George W Bush



predicted: George W Bush
true: George W Bush



predicted: George W Bush
true: George W Bush



predicted: George W Bush
true: George W Bush





文本 \Rightarrow 向量

在 ChatGPT 出现之前，计算机读不懂句子，只能把文章看成一些关键词的线性组合，闻闻文章的“味道”。词典中选定10万个词作关键词。对每一篇文章，计算关键词出现的频率（TF/IDF），得到一个10万维的向量。研究这些向量的空间位置关系作分类推送。

Index Words	Titles								
	T1	T2	T3	T4	T5	T6	T7	T8	T9
book			1	1					
dads						1			1
dummies		1						1	
estate							1		1
guide	1					1			
investing	1	1	1	1	1	1	1	1	1
market	1		1						
real							1		1
rich						2			1
stock	1		1					1	
value				1	1				

在欧氏空间中研究一组样本点

$\alpha_1, \alpha_2, \dots, \alpha_n$ 的分布.

第一步：向量中心化

记
$$\tilde{\alpha} = \frac{1}{n} (\alpha_1 + \alpha_2 + \dots + \alpha_n)$$

$$\widetilde{\alpha}_i = \alpha_i - \tilde{\alpha}, \quad i = 1, \dots, n$$

下面我们假定 $\alpha_1 + \dots + \alpha_n = 0$.



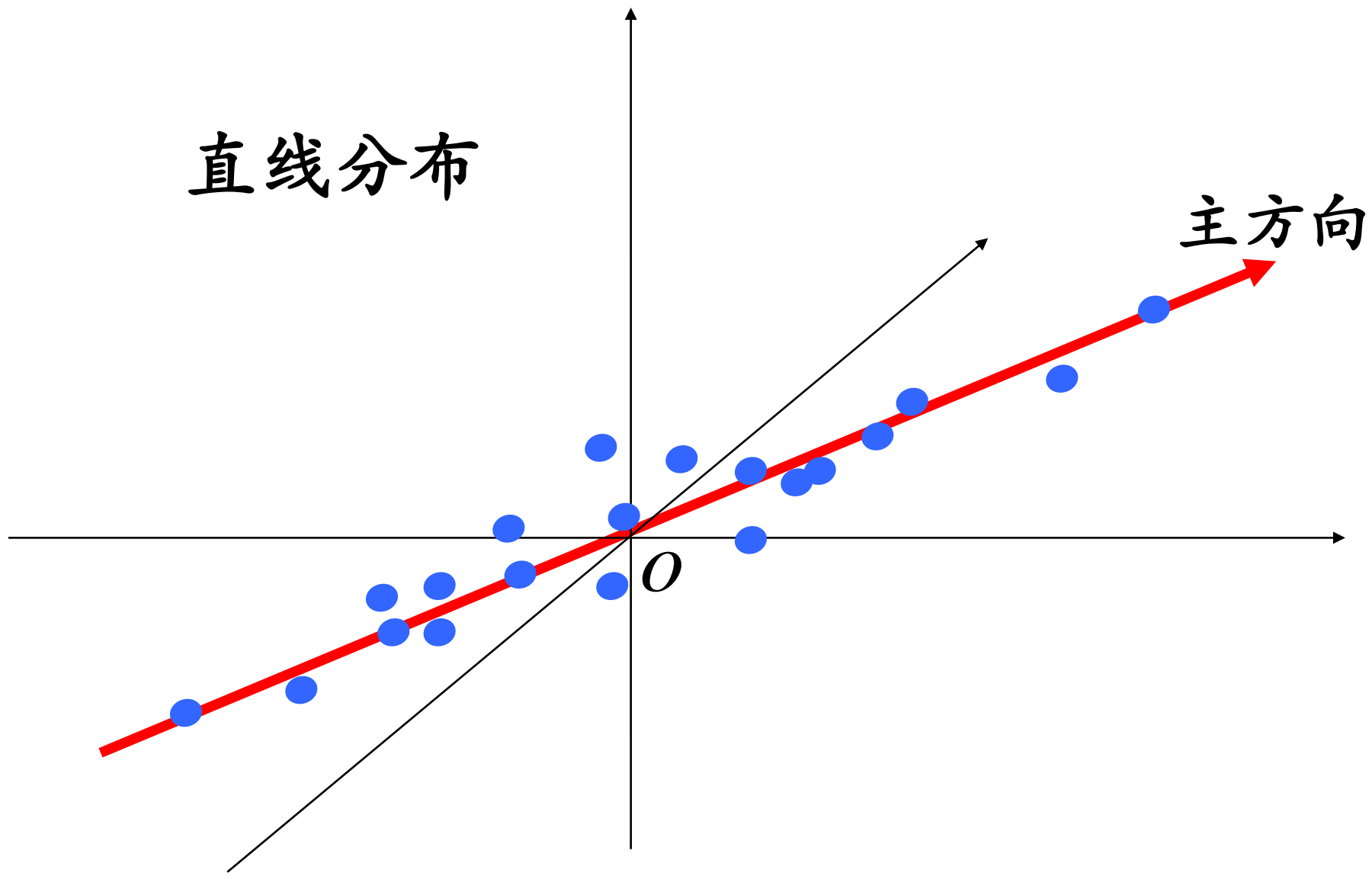
分布的主方向, 主平面

在欧氏空间中给定一组中心化的样本点

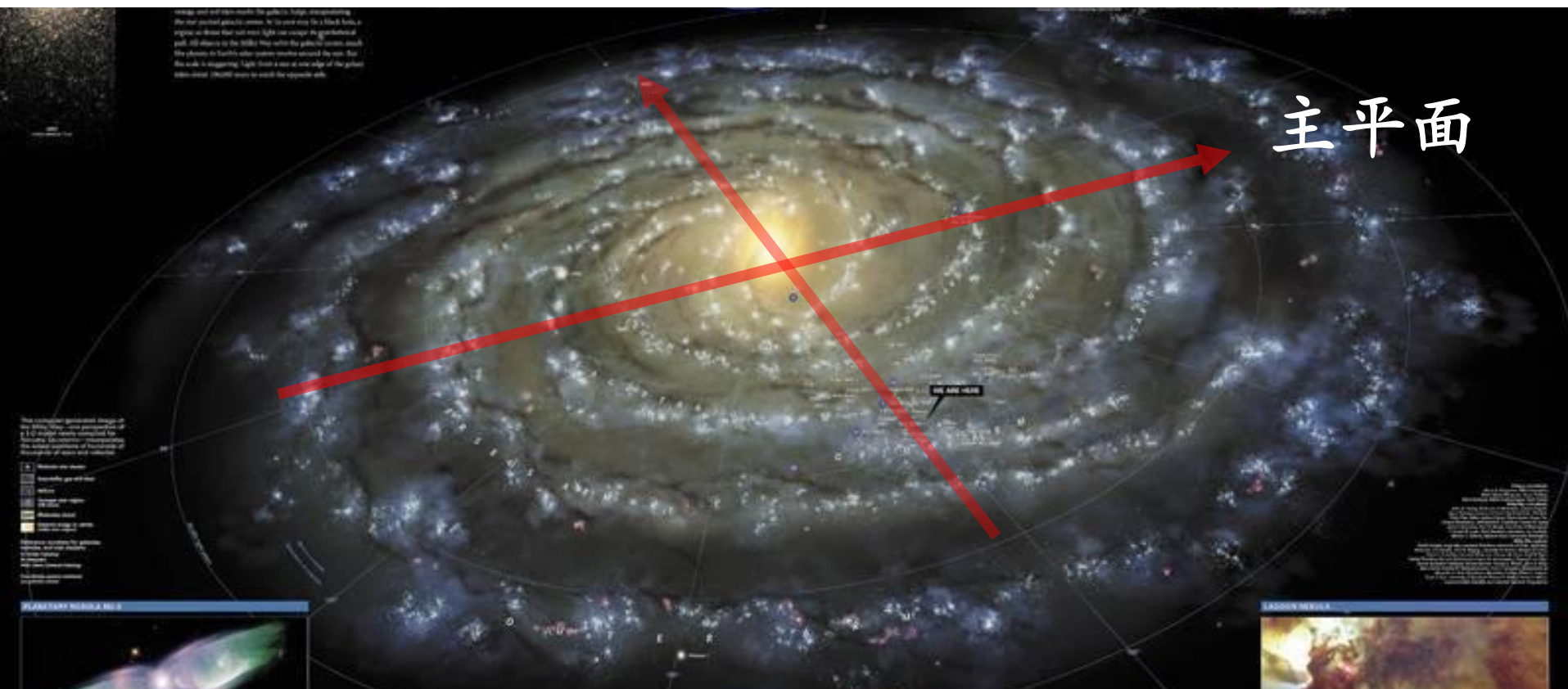
$$\alpha_1, \dots, \alpha_n.$$

则存在一条过原点直线 β , 使得在所有过原点直线中, 点 $\alpha_1, \dots, \alpha_n$ 到 β 的距离的平方和是最小的. 这样的直线 β 称为点 $\alpha_1, \dots, \alpha_n$ 分布的主方向.

直线分布



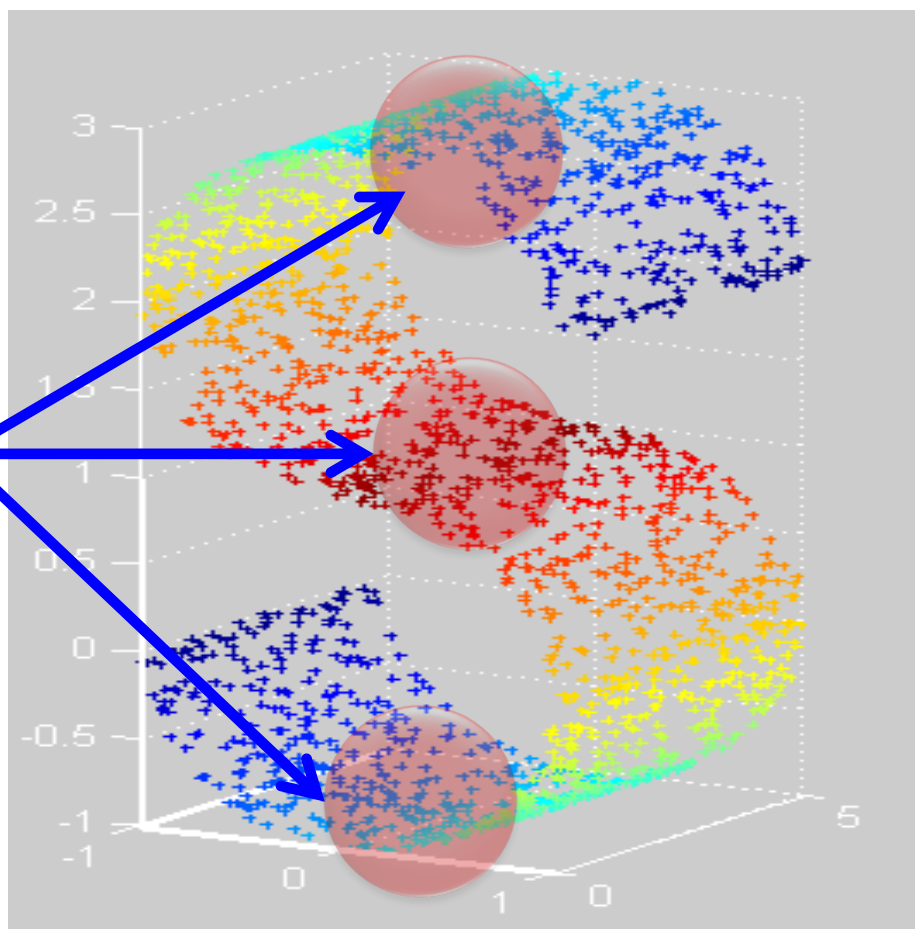
饼状分布



非线性分布



局部近似
平面分布



主成分分析(PCA)

设 $A = [\alpha_1 \alpha_2 \cdots \alpha_n]$ 是 $m \times n$ 实矩阵,

A 列分布的主方向, 主...

$$AA^T = \underbrace{[\beta_1 \cdots \beta_r \cdots \beta_m]}_{\text{正交矩阵 } P} \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_r & \\ & & & \mathbf{0} \end{bmatrix} P^T$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_m = 0$$

若 V 是 \mathbf{R}^m 的子空间, 记

$$Q(A, V) = \sum_{i=1}^n d(\alpha_i, V)^2 .$$

这里 $d(\alpha, V)$ 表示 α 到 V 的最短欧氏距离.

对于 $1 \leq k \leq m$, 记 $V_k = \langle \beta_1, \dots, \beta_k \rangle$.

$$\min_{\dim V=k} Q(A, V) = ?$$

定理 (Principal Component Analysis):

在欧氏空间 \mathbb{R}^m 的所有 k 维子空间中,
点 $\alpha_1, \dots, \alpha_n$ 到子空间 $V_k = \langle \beta_1, \dots, \beta_k \rangle$
的距离的平方和最小:

$$\min_{\dim V=k} Q(A, V) = Q(A, V_k) = \lambda_{k+1} + \dots + \lambda_m$$

证：设 V 是 \mathbf{R}^m 的一个 k 维子空间，

$\omega_1, \dots, \omega_k$ 是 V 的一组标准正交基。

记 $\mathbf{B} = [\omega_1 \dots \omega_k]$ ，则有 $\mathbf{B}^T \mathbf{B} = \mathbf{I}_k$ ，

$\mathbf{B} \mathbf{B}^T \alpha_i$ 是 α_i 在子空间 V 上的正交投影。



$$\begin{aligned} & (\alpha_i, \omega_1) \omega_1 + \dots + (\alpha_i, \omega_k) \omega_k \\ &= \omega_1 (\omega_1^T \alpha_i) + \dots + \omega_k (\omega_k^T \alpha_i) \\ &= (\omega_1 \omega_1^T + \dots + \omega_k \omega_k^T) \alpha_i = \mathbf{B} \mathbf{B}^T \alpha_i \end{aligned}$$

对 α_i ($i = 1, \dots, n$) 应用勾股定理后求和

$$\sum \|\alpha_i\|^2 = \sum \|\mathbf{B} \mathbf{B}^T \alpha_i\|^2 + \sum \|\alpha_i - \mathbf{B} \mathbf{B}^T \alpha_i\|^2$$

 $\underbrace{\qquad\qquad\qquad}_{\text{正交投影平方和}} + \underbrace{\qquad\qquad\qquad}_{\text{垂线平方和}}$

$$\text{tr}(\mathbf{A}^T \mathbf{A}) = \text{tr}(\mathbf{A} \mathbf{A}^T) = \lambda_1 + \dots + \lambda_m$$



欲使垂线平方和取最小值，只需让

正交投影的平方和取到最大，即最大化

$$\|\mathbf{B} \mathbf{B}^T \mathbf{A}\|_F^2 = \text{tr}\left((\mathbf{B} \mathbf{B}^T \mathbf{A})^T (\mathbf{B} \mathbf{B}^T \mathbf{A})\right) = \text{tr}(\mathbf{A}^T \mathbf{B} \mathbf{B}^T \mathbf{A})$$

$\alpha_1, \dots, \alpha_n$ 到 V 的正交投影长度的平方和

$$= \text{tr}((\mathbf{B}\mathbf{B}^T\mathbf{A})^T(\mathbf{B}\mathbf{B}^T\mathbf{A}))$$

$$= \text{tr}(\mathbf{A}^T\mathbf{B}\mathbf{B}^T\mathbf{B}\mathbf{B}^T\mathbf{A})$$

$$\mathbf{B}^T\mathbf{B} = \mathbf{I}_k$$

$$= \text{tr}(\mathbf{A}^T\mathbf{B}\mathbf{B}^T\mathbf{A})$$

$$= \text{tr}(\mathbf{B}\mathbf{B}^T\mathbf{A}\mathbf{A}^T)$$

$$= \text{tr}(\mathbf{B}\mathbf{B}^T\mathbf{P}\mathbf{D}\mathbf{P}^T)$$

存在正交矩阵 \mathbf{P} ,
 $\mathbf{D} = \text{diag}\{\lambda_1, \dots, \lambda_m\}$

$$= \text{tr}(\mathbf{P}^T\mathbf{B}\mathbf{B}^T\mathbf{P}\mathbf{D})$$

$$= \text{tr}(\mathbf{C}\mathbf{C}^T\mathbf{D})$$

记 $\mathbf{C} = \mathbf{P}^T\mathbf{B}$





记 $m \times k$ 矩阵 $C = P^T B$, 由

$$C^T C = B^T P P^T B = B^T B = I_k$$

知 C 的 k 列也由两两正交的单位向量组成.

故 C 的列组可扩充成 R^m 的标准正交基,

排成一个正交矩阵. 由此知 C 第 i 行元素

的平方和 c_i 满足条件 $0 \leq c_i \leq 1$, 且

$$c_1 + \dots + c_m = \text{tr}(C C^T) = \text{tr}(C^T C) = k .$$

设 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ 及排序不等式, 得
 $\text{tr}(\mathbf{C} \mathbf{C}^T \mathbf{D}) = c_1 \lambda_1 + \dots + c_m \lambda_m$
 $\leq \lambda_1 + \lambda_2 + \dots + \lambda_k$
 等号可在 $\mathbf{B} = [\beta_1 \dots \beta_k]$, $\mathbf{C} = \mathbf{P}^T \mathbf{B} =$
 $\mathbf{V}_1 \dots \mathbf{V}_k$ 时取到.
 若 $\lambda_k > \lambda_{k+1}$ 时, 等号仅在 $\mathbf{V}_k = \langle \beta_1, \dots, \beta_k \rangle$
 取到, 此时 k 维主子空间唯一.

由 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ 及排序不等式, 得

$$\text{tr}(\mathbf{C} \mathbf{C}^T \mathbf{D}) = c_1 \lambda_1 + \dots + c_m \lambda_m$$

$$\leq \lambda_1 + \lambda_2 + \dots + \lambda_k$$

等号可在 $\mathbf{B} = \mathbf{P} \mathbf{C} = [\beta_1 \dots \beta_k]$, $\mathbf{C} = \begin{bmatrix} \mathbf{I}_k \\ \mathbf{0} \end{bmatrix}$,

$\mathbf{V}_k = \langle \beta_1, \dots, \beta_k \rangle$ 时取到.

当 $\lambda_k > \lambda_{k+1}$ 时, 等号仅在 $\mathbf{V}_k = \langle \beta_1, \dots, \beta_k \rangle$

取到, 此时 k 维主子空间唯一.

$$\mathbf{A} = [\alpha_1 \alpha_2 \cdots \alpha_n]$$

$$\mathbf{A}\mathbf{A}^T = [\beta_1 \cdots \beta_k \cdots \beta_m]$$

$$\mathbf{B}_k = [\beta_1 \cdots \beta_k]$$

投影平方和最大

垂线平方和最小

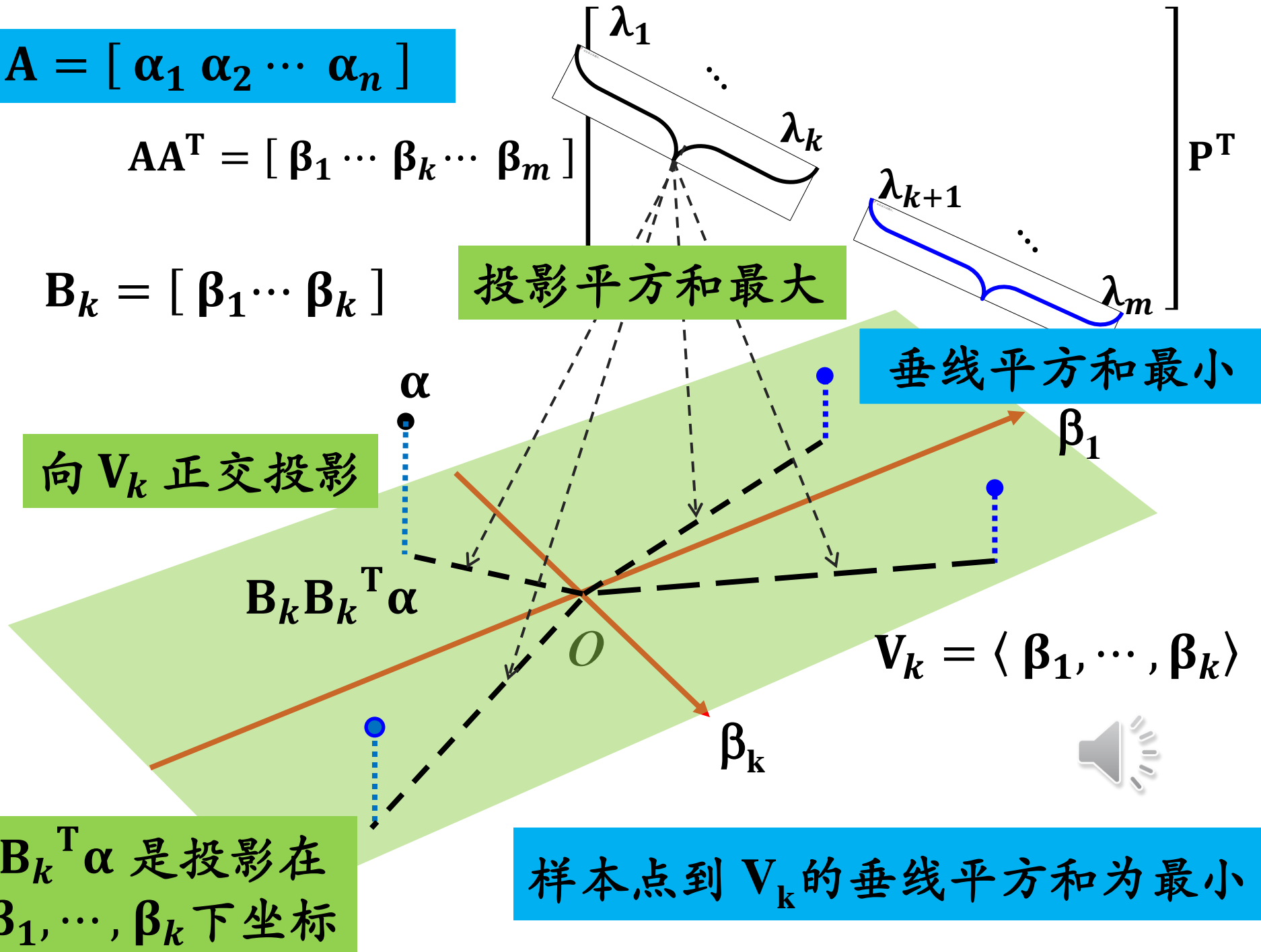
向 V_k 正交投影

$$\mathbf{B}_k \mathbf{B}_k^T \alpha$$

$$V_k = \langle \beta_1, \cdots, \beta_k \rangle$$

$\mathbf{B}_k^T \alpha$ 是投影在 β_1, \cdots, β_k 下坐标

样本点到 V_k 的垂线平方和为最小



定理 5. 设 $A = P \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix} Q^T$ 是实矩阵

$A \in M_{m,n}(\mathbb{R})$ 的 SVD 分解. 这里 P, Q 分别是 m 级, n 级正交矩阵, $r = A$ 秩, $\sigma_1 \geq \dots \geq \sigma_r > 0$ 是 A 的奇异值. 记

$$A_k = P \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_k & \\ & & & 0 \end{bmatrix} Q^T, \quad 1 \leq k \leq r$$

矩阵的奇异值分解(SVD)

定理：每个 $m \times n$ 实矩阵 A 都可写成

$$A = \underbrace{[\beta_1 \ \beta_2 \ \cdots \ \beta_m]}_{\mathbf{P} \text{ 正交}} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \mathbf{0} \end{bmatrix} \underbrace{\begin{bmatrix} \gamma_1^T \\ \gamma_2^T \\ \vdots \\ \gamma_n^T \end{bmatrix}}_{\mathbf{Q}^T \text{ 正交}}$$

$$= \sigma_1 \beta_1 \gamma_1^T + \sigma_2 \beta_2 \gamma_2^T + \cdots + \sigma_r \beta_r \gamma_r^T$$

$$A \gamma_i = \sigma_i \beta_i, \quad A^T \beta_i = \sigma_i \gamma_i, \quad i = 1, \cdots, r$$

则 \mathbf{A}_k 是 \mathbf{A} 在秩 k 下的最佳逼近：

即对任意秩为 k 的矩阵 $\mathbf{B} \in \mathbf{M}_{m,n}(\mathbf{R})$, 有

$$\|\mathbf{A} - \mathbf{B}\|_{\mathbf{F}}^2 \geq \|\mathbf{A} - \mathbf{A}_k\|_{\mathbf{F}}^2 = \sigma_{k+1}^2 + \cdots + \sigma_r^2.$$

这里

$$\begin{aligned} \|\mathbf{A}\|_{\mathbf{F}} &= \sqrt{\text{tr}(\mathbf{A}^T \mathbf{A})} = \sqrt{\text{tr}(\mathbf{A} \mathbf{A}^T)} = \sqrt{\sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} a_{ij}^2} \\ &= \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} \end{aligned}$$

表示实矩阵 $\mathbf{A} = [a_{ij}]$ 的 Frobenius 范数.

证: 记 $A = [\alpha_1 \cdots \alpha_n], B = [\xi_1 \cdots \xi_n] \in M_{m,n}(\mathbb{R})$.

$V = \langle \xi_1, \cdots, \xi_n \rangle$ ($\dim V = k$) 为的 B 列空间.

则有

$$\|\alpha_i - \xi_i\|^2 \geq d(\alpha_i, V)^2, 1 \leq i \leq n.$$

这里 $d(\alpha, V)$ 表示 α 到 V 的最短欧氏距离. 故

$$\begin{aligned} \|A - B\|_F^2 &\geq Q(A, V) = \sum_{i=1}^n d(\alpha_i, V)^2 \\ &\geq Q(A, V_k) \quad (\text{PCA定理}) \end{aligned}$$

上式中的等号可在 $\mathbf{B} = \mathbf{A}_k = \mathbf{B}_k \mathbf{B}_k^T \mathbf{A}$ 时取到，
这里 $\mathbf{B}_k = [\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_k]$ 。注意 $\mathbf{B}_k \mathbf{B}_k^T$ 是 \mathbf{R}^m 向
 \mathbf{B}_k 的列空间 V_k 作正交投影的矩阵。此时 \mathbf{B} 的
列向量 $\mathbf{B}_k \mathbf{B}_k^T \boldsymbol{\alpha}_i$ 刚好是 $\boldsymbol{\alpha}_i$ 在 V_k 上的正交投影，
 $1 \leq i \leq n$ ，故等号成立。

Moore–Penrose 广义逆

若 \mathbf{A} 是复矩阵, 则 *存在唯一的* 复矩阵 \mathbf{X} ,
满足

$$\mathbf{A} \mathbf{X} \mathbf{A} = \mathbf{A}, \quad \mathbf{X} \mathbf{A} \mathbf{X} = \mathbf{X}$$

$$\overline{(\mathbf{A} \mathbf{X})}^T = \mathbf{A} \mathbf{X}, \quad \overline{(\mathbf{X} \mathbf{A})}^T = \mathbf{X} \mathbf{A}.$$

\mathbf{X} 称为 \mathbf{A} 的 Moore–Penrose 广义逆,
记为 \mathbf{A}^+ . 我们主要研究实矩阵的情况.

设 A 是 $m \times n$ 实矩阵. 满足条件

$$A X A = A, \quad X A X = X,$$

$$(A X)^T = A X, \quad (X A)^T = X A$$

的实矩阵 X 存在且唯一, 称为 A 的 **Moore–Penrose** 广义逆, 记为 A^+ . 容易验证:

若 B 列满秩, 则 $B^+ = (B^T B)^{-1} B^T$ ($B^+ B = I$);

若 C 行满秩, 则 $C^+ = C^T (C C^T)^{-1}$ ($C C^+ = I$);

若 $A = BC$ 是 $A \in M_{m,n}(R)$ 的满秩分解,
即 B 列满秩, C 行满秩 (通常 B 可取成 A 的
主元列向量组排成的矩阵, C 则由 A 的行
简化阶梯型的非零行构成), 则有

$$\begin{aligned} A^+ &= C^+ B^+ \\ &= C^T (C C^T)^{-1} (B^T B)^{-1} B^T \end{aligned}$$

实矩阵的 **M – Penrose** 广义逆也可以用
SVD 分解来计算. 容易验证, 若 $A = P S Q^T$,

则 $A^+ = Q S^+ P^T$

$$= Q \begin{bmatrix} \sigma_1^{-1} & & & \\ & \ddots & & \\ & & \sigma_r^{-1} & \\ & & & \mathbf{0} \end{bmatrix}_{n \times m} P^T$$

Moore – Penrose 广义逆的性质

设 $A \in M_{m,n}(R)$.

- 1) 若方程组 $AX = \beta$ 无解, 则 $X = A^+\beta$ 是方程组的最小二乘解(满足 $A^T A X = A^T \beta$)
- 2) 若方程组 $AX = \beta$ 有解, 则 $X = A^+\beta$ 是方程组的最小范数解 ($A^+\beta \perp \text{Ker } A$).

用 Matlab 计算特征值与特征向量



```
>> A = [ 1, -2, -4;  
        -2, 4, -2;  
        -4, -2, 1 ]
```

```
A =  
    1    -2    -4  
   -2     4    -2  
   -4    -2     1
```

```
>> [ U, D ] = eig(A)
```

```
U =  
    0.6667   -0.3431   -0.6617  
    0.3333    0.9313   -0.1470  
    0.6667   -0.1225    0.7352
```

```
D =
```

```
   -4.0000         0         0  
         0    5.0000         0  
         0         0    5.0000
```

```
>> U * D * U^(-1)
```

```
    1.0000   -2.0000   -4.0000  
   -2.0000    4.0000   -2.0000  
   -4.0000   -2.0000    1.0000
```

```
>> U * U'
```

```
    1.0000   -0.0000         0  
   -0.0000    1.0000         0  
         0         0    1.0000
```


Eigenvectors and Eigenvalues in Matlab

A =

1	-2	-4
-2	4	-2
-4	-2	1

```
>> [ P, S, Q ] = svd(A)
```

P =

-0.0000	-0.7454	-0.6667
-0.8944	0.2981	-0.3333
0.4472	0.5963	-0.6667

S =

5.0000	0	0
0	5.0000	0
0	0	4.0000

Q =

0	-0.7454	0.6667
-0.8944	0.2981	0.3333
0.4472	0.5963	0.6667

```
>> P*S*Q'
```

ans =

1.0000	-2.0000	-4.0000
-2.0000	4.0000	-2.0000
-4.0000	-2.0000	1.0000



```
>> B = [ 2,4,1,5 ; 3,2,6,9 ;  
         3,7,1,8 ]
```

```
B =  
     2     4     1     5  
     3     2     6     9  
     3     7     1     8
```

```
>> rref( B ) %简化阶梯型  
ans =  
1.0000     0     0    -0.6000  
     0 1.0000     0     1.2000  
     0     0 1.0000     1.4000
```

```
>> [Q,R] = qr( B ) % QR分解  
Q =  
-0.4264    0.2242   -0.8763  
-0.6396   -0.7598    0.1168  
-0.6396    0.6103    0.4674
```

```
R =  
-4.6904   -7.4620   -4.9036  -13.0052  
     0    3.6494   -3.7241   -0.8345  
     0         0    0.2921    0.4089
```



```
>> Q*Q', Q*R
```

```
ans =  
1.0000         0    0.0000  
         0    1.0000   -0.0000  
0.0000   -0.0000    1.0000
```

```
ans =  
2.0000    4.0000    1.0000    5.0000  
3.0000    2.0000    6.0000    9.0000  
3.0000    7.0000    1.0000    8.0000
```

```
>> B = [ 2,4,1,5 ; 3,2,6,9 ;  
         3,7,1,8 ]
```

```
B =  
     2     4     1     5  
     3     2     6     9  
     3     7     1     8
```

```
>> [ P, S, Q ] = svd( B )
```

```
P =  
    -0.4026    0.2710   -0.8743  
    -0.6483   -0.7588    0.0633  
    -0.6462    0.5923    0.4812
```

```
S =  
    16.4635     0     0     0  
     0     5.2856     0     0  
     0     0    0.1254     0
```

```
Q =
```

```
   -0.2848    0.0081   -0.9183    0.2750  
   -0.4513    0.7024   -0.0186   -0.5500  
   -0.3000   -0.6980   -0.1053   -0.6417  
   -0.7907   -0.1391    0.3813    0.4583
```



```
>> S(3,3) = 0
```

```
S =  
    16.4635     0     0     0  
     0     5.2856     0     0  
     0     0     0     0
```

```
>> B - P*S*Q'
```

```
ans =  
    0.1007    0.0020    0.0115   -0.0418  
   -0.0073   -0.0001   -0.0008    0.0030  
   -0.0554   -0.0011   -0.0064    0.0230
```

用 Matlab 计算主方向

```
% 随机产生一个2×12矩阵
>> A = 10*rand(2,12);

% 列向量中心化
>> rm = sum( A' ) / 12;
    A = A - diag(rm)*ones(2,12)
    A =
•   -1.6650    2.8442   -3.1978
      0.7599    3.6580   -2.1712
      0.0801   -2.4130   -1.9733
      1.7668    2.2753    0.0361
•   -0.3648   -4.4312   -3.9954
      2.0140    1.0397   -2.4991
      2.8466   -4.3883    3.0497
      4.7125   -0.8095    2.8257
```

```
>> [ U,D ] = eig( A*A' )
    U =
      -0.9430    0.3329
       0.3329    0.9430
    D =
      49.6889           0
           0   122.7180
% 主特征向量，长度正规化
>> E=U*(D/12)^(1/2)
    E =
      -1.9188    1.0647
       0.6775    3.0155
```

A =

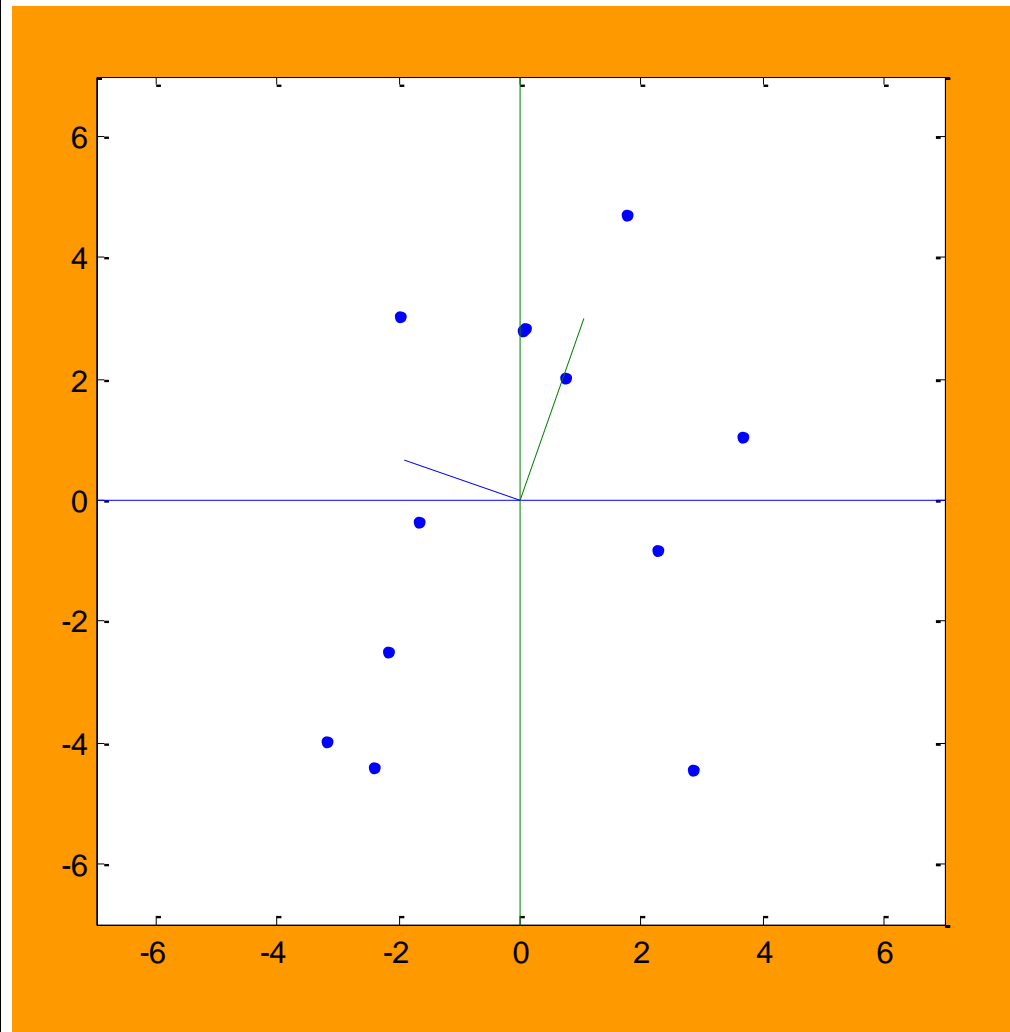
- | | | |
|---------|---------|---------|
| -1.6650 | 2.8442 | -3.1978 |
| 0.7599 | 3.6580 | -2.1712 |
| 0.0801 | -2.4130 | -1.9733 |
| 1.7668 | 2.2753 | 0.0361 |
- | | | |
|---------|---------|---------|
| -0.3648 | -4.4312 | -3.9954 |
| 2.0140 | 1.0397 | -2.4991 |
| 2.8466 | -4.3883 | 3.0497 |
| 4.7125 | -0.8095 | 2.8257 |

E =

-1.9188	1.0647
0.6775	3.0155

```
>> n = 1:12 ;
```

```
plot( A(1,n), A(2,n), '.'),  
line([-10,0;10,0], [0,-10;0,10]),  
line([0,0;E(1,: )], [0,0;E(2,: )]),  
axis([-7 7 -7 7]);
```



作业: PCA 应用---BioMotion

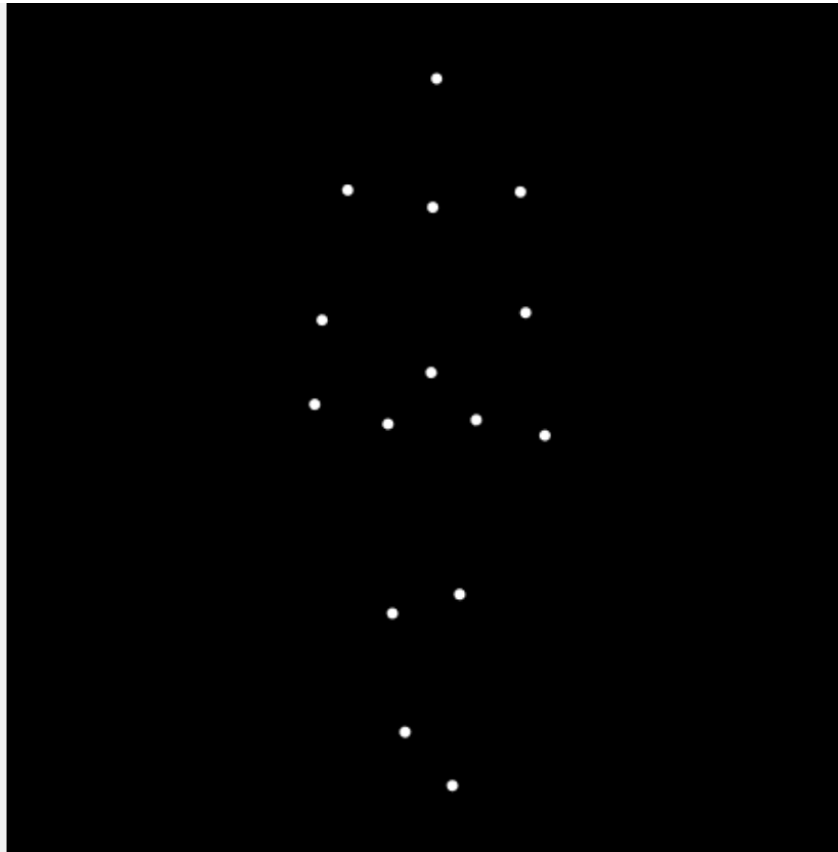
阅读以下链接文章, 学习如何用主成份分析的方法来优美地(滤波)表示周期性运动.

<https://www.biomotionlab.ca/html5-bml-walker/>

<http://jov.arvojournals.org/article.aspx?articleid=2192503>

<https://blogs.mathworks.com/cleve/2016/04/11/the-eigenwalker-model-of-the-human-gait/>

作业: PCA 应用---BioMotion



Nervousness <

Happiness <

Rotate <

Lines



Pause



Info

Reset



练习：如何作正交投影

1. 若实矩阵 A 列满秩，证明映射：

$$X \mapsto A (A^T A)^{-1} A^T X$$

是欧氏空间到 A 列空间的正交投影。

2. 若实矩阵 A 满足 $A^T A = I$ ，则映射：

$$X \mapsto A A^T X$$

是欧氏空间到 A 列空间的正交投影。

练习：如何作投影

3. 若实对称矩阵 A 满足 $A^2 = A$ ，则映射：

$$X \mapsto AX$$

是欧氏空间到 A 列空间的正交投影。

4. 若矩阵 A 满足 $A^2 = A$ ，则映射：

$$X \mapsto AX$$

是 K^n 沿着 A 解空间到 A 列空间的投影变换。

练习 5 . 设实矩阵 $A = [a_{ij}] \in M_{m,n}(\mathbb{R})$ 的奇异值
(即 AA^T 正特征值的平方根) 为

$$\sigma_1 \geq \cdots \geq \sigma_r > 0, \quad r = A \text{ 秩}.$$

证明: $|a_{ij}| \leq \sigma_1, \quad \forall i, j.$

Review

1. 已知 3 阶矩阵 \mathbf{A} 特征值为 0, -1, 1, 对应的特征向量为 $(1, 0, 0)^T$, $(0, 1, 0)^T$, $(-1, 0, 1)^T$, 求 \mathbf{A}^m .

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

$$\mathbf{A}^{\mathbf{m}} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & & \\ & (-1)^{\mathbf{m}} & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 0 & 0 & -1 \\ 0 & (-1)^{\mathbf{m}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

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

$$\begin{aligned}
\mathbf{A}^{\mathbf{m}} &= \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & & \\ & (-1)^{\mathbf{m}} & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} 0 & 0 & -1 \\ 0 & (-1)^{\mathbf{m}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & -1 \\ 0 & (-1)^{\mathbf{m}} & 0 \\ 0 & 0 & 1 \end{bmatrix}
\end{aligned}$$

2. 证明： 对角分块方阵

$$\mathbf{C} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{B} \end{array} \right] \left. \vphantom{\begin{array}{c|c} \mathbf{A} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{B} \end{array}} \right\} \mathbf{n}$$

$\underbrace{\hspace{1.5cm}}_{\mathbf{m}}$

可对角化当且仅当 \mathbf{A} 、 \mathbf{B} 可对角化 .

若

$$\mathbf{A}_1 = \mathbf{U}_1 \mathbf{D}_1 \mathbf{U}_1^{-1} \qquad \mathbf{A}_2 = \mathbf{U}_2 \mathbf{D}_2 \mathbf{U}_2^{-1}$$

则

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix} \sim \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{U}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix} \begin{bmatrix} \mathbf{U}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2^{-1} \end{bmatrix}$$

反之，若 $\mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}$ 可对角化，则存在 \mathbf{n} 个线性无关的特征向量 $\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}$ ， $1 \leq i \leq \mathbf{n}$ 。

$$\left\{ \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \lambda_i \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} \right\}^{\mathbf{m}}$$

于是 $\mathbf{A} \alpha_i = \lambda_i \alpha_i$ ， $\mathbf{B} \beta_i = \lambda_i \beta_i$

注意到 $\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \beta_1 & \beta_2 & \cdots & \beta_n \end{bmatrix}$ 行向量线性无关,

$\Rightarrow [\alpha_1, \alpha_2, \dots, \alpha_n]$ 的秩 = 行数 = m

$\Rightarrow A$ 有 m 个线性无关的特征向量,

可对角化

类似的, 可以证明 B 也可对角化.

Review

3. 若方阵 **A** 、 **B** 都可对角化, **AB** 是否一定能对角化?

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

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

Review

4. 若方阵 A 、 B 都可对角化, 且 $AB = BA$.

问 AB 是否一定能对角化?

结论: AB 可对角化.

A 可写成 $U D U^{-1}$, U 可逆, D 对角.

记 $C = U^{-1} B U$. 则 $AB = U D C U^{-1}$,

且 $C D = U^{-1} B A U = U^{-1} A B U = D C$.

问题归结为证明 $C D$ 可对角化.

将对角矩阵 **D** 写成对角分块形式， λ_i 互异

$$\mathbf{D} = \begin{bmatrix} \lambda_1 \mathbf{I}_1 & & & \\ & \lambda_2 \mathbf{I}_2 & & \\ & & \ddots & \\ & & & \lambda_s \mathbf{I}_s \end{bmatrix}$$

将 **C** 也写成相应的分块矩阵形式

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1s} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \cdots & \mathbf{C}_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{s1} & \mathbf{C}_{s2} & \cdots & \mathbf{C}_{ss} \end{bmatrix}$$

$$\mathbf{DC} = \begin{bmatrix} \lambda_1 \mathbf{C}_{11} & \lambda_1 \mathbf{C}_{12} & \cdots & \lambda_1 \mathbf{C}_{1s} \\ \lambda_2 \mathbf{C}_{21} & \lambda_2 \mathbf{C}_{22} & \cdots & \lambda_2 \mathbf{C}_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_s \mathbf{C}_{s1} & \lambda_s \mathbf{C}_{s2} & \cdots & \lambda_s \mathbf{C}_{ss} \end{bmatrix}$$

$$\mathbf{CD} = \begin{bmatrix} \lambda_1 \mathbf{C}_{11} & \lambda_2 \mathbf{C}_{12} & \cdots & \lambda_s \mathbf{C}_{1s} \\ \lambda_1 \mathbf{C}_{21} & \lambda_2 \mathbf{C}_{22} & \cdots & \lambda_s \mathbf{C}_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 \mathbf{C}_{s1} & \lambda_2 \mathbf{C}_{s2} & \cdots & \lambda_s \mathbf{C}_{ss} \end{bmatrix}$$

$$\mathbf{DC} = \begin{bmatrix} \lambda_1 \mathbf{C}_{11} & & & \\ & \lambda_2 \mathbf{C}_{22} & & \\ & & \ddots & \\ & & & \lambda_s \mathbf{C}_{ss} \end{bmatrix}$$

||

$$\mathbf{CD} = \begin{bmatrix} \lambda_1 \mathbf{C}_{11} & & & \\ & \lambda_2 \mathbf{C}_{22} & & \\ & & \ddots & \\ & & & \lambda_s \mathbf{C}_{ss} \end{bmatrix}$$

由 $\mathbf{C} \mathbf{D} = \mathbf{D} \mathbf{C}$ 可推出 \mathbf{C} 是对角分块矩阵,
对角块为 $\mathbf{C}_{11}, \mathbf{C}_{22}, \dots, \mathbf{C}_{ss}$.

利用第 2 题的结论,

\mathbf{C} 可对角化

\Rightarrow 每个对角块 \mathbf{C}_{ii} 可对角化

\Rightarrow 每个 $\lambda_i \mathbf{C}_{ii}$ 可对角化

$\Rightarrow \mathbf{C} \mathbf{D}$ 可对角化

*See you
next
time*



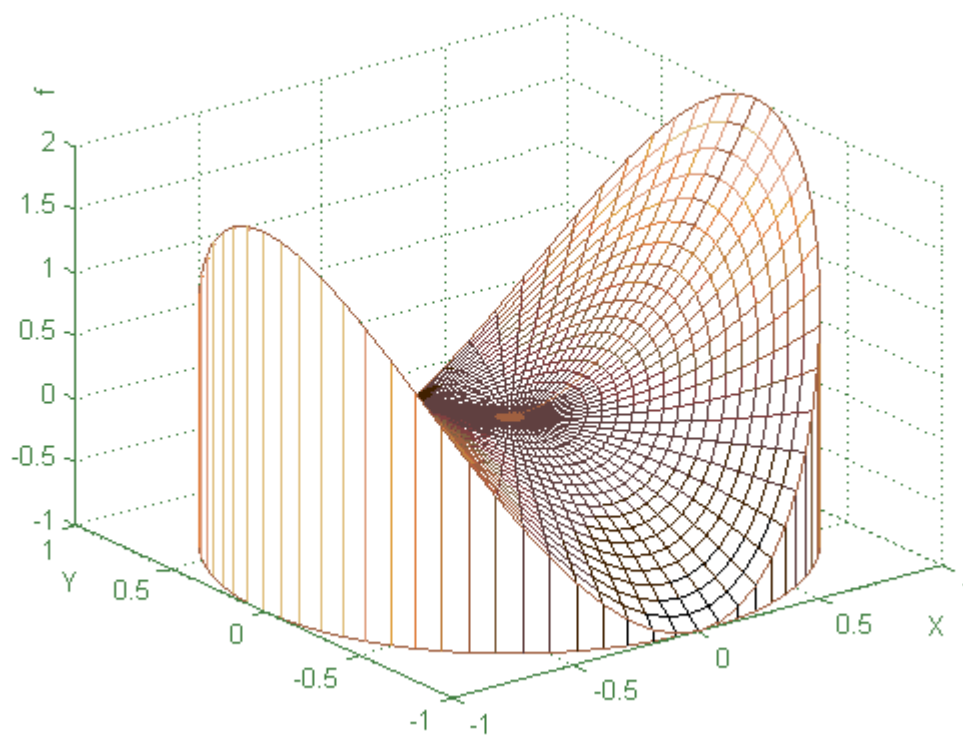
4. 设 A 是一个 n 级复矩阵, $\lambda_1, \dots, \lambda_n$ 是 A 的复特征值, α 是属于 λ_1 的一个复特征向量, 且 $\bar{\alpha}^T \alpha = 1$. 证明: 对任意 $\beta \in \mathbb{C}^n$, 矩阵 $A + \alpha \bar{\beta}^T$ 的特征值为

$$\lambda_1 + \bar{\beta}^T \alpha, \lambda_2, \dots, \lambda_n;$$

且有 $(A + \alpha \bar{\beta}^T)\alpha = (\lambda_1 + \bar{\beta}^T \alpha)\alpha.$

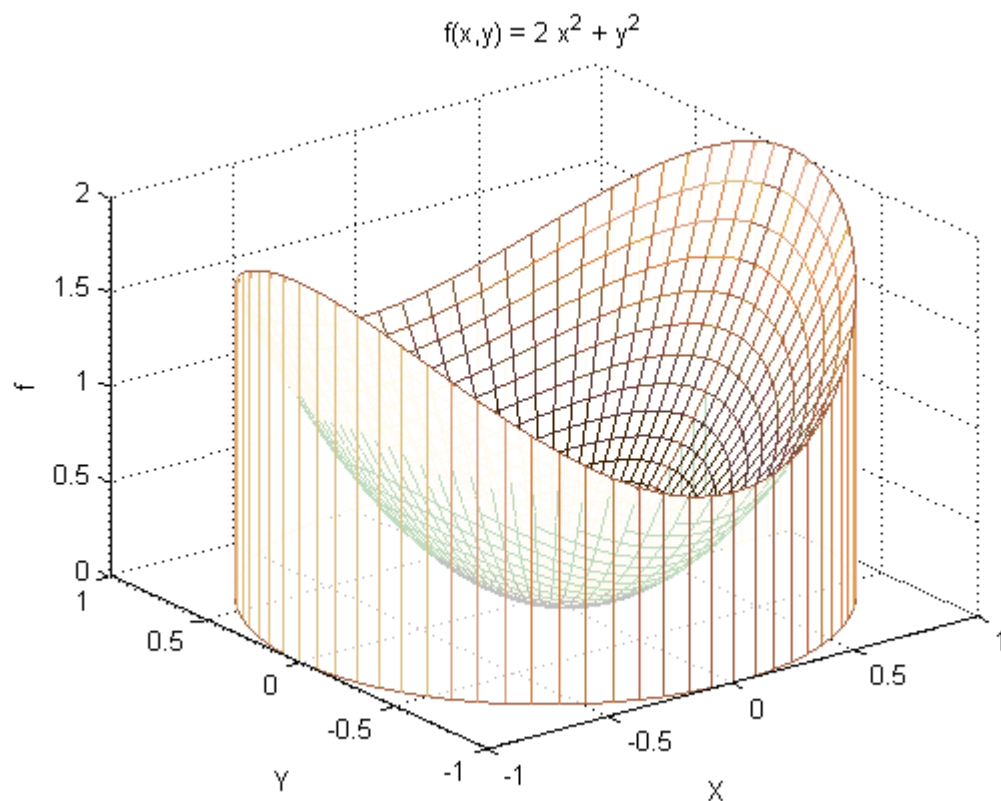
例：实二次型 $f(x, y) = 2x^2 - y^2$

在单位圆 $x^2 + y^2 = 1$ 上的最大, 最小值

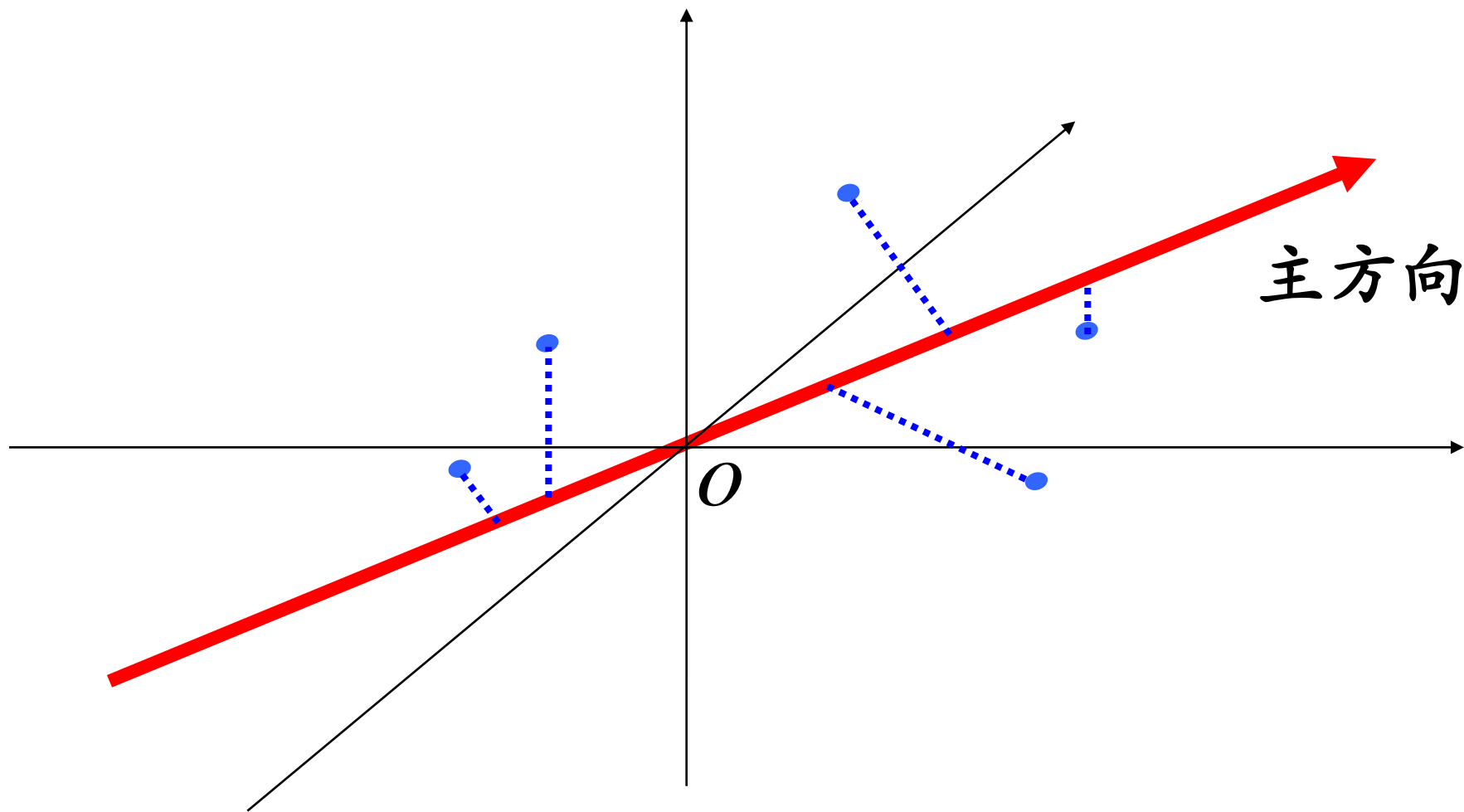


例：实二次型 $f(\mathbf{x}, \mathbf{y}) = 2\mathbf{x}^2 + \mathbf{y}^2$

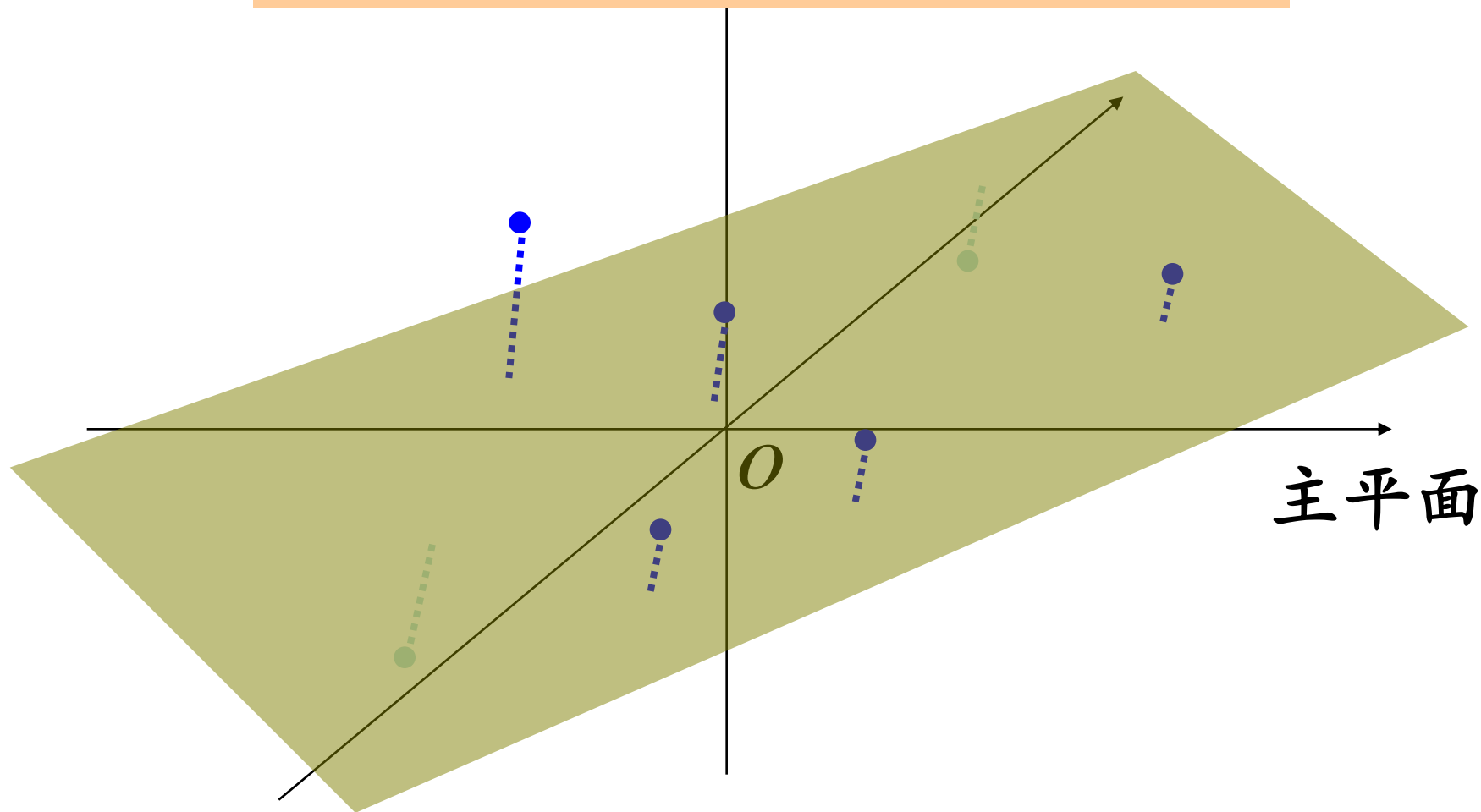
在单位圆 $\mathbf{x}^2 + \mathbf{y}^2 = 1$ 上的最大, 最小值



使距离平方和最小的直线



使垂线距离的平方和取最小的
平面称为分布的主平面



定理 1：若 A 是实矩阵，则有

1) $A^T A$ 解空间 = A 解空间

2) $A^T A$ 秩 = A 秩 = $A A^T$ 秩

3) A 列空间 = $A A^T$ 列空间

3) 证： A 列空间 $\supseteq A A^T$ 列空间；

又 A 秩 = $A A^T$ 秩，推得 3)。

$\beta_1 \cdots \beta_r$ 是 A 列空间里最好的标准正交基

设 $A = [\alpha_1 \alpha_2 \cdots \alpha_n]$ 是 $m \times n$ 实矩阵,

则 AA^T 是 m 级实对称矩阵, 于是存在

正交矩阵 P , 使得

A 列分布的主方向, 主...

$$AA^T = \underbrace{[\beta_1 \cdots \beta_r \cdots \beta_m]}_{\text{正交矩阵 } P} \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_r & \\ & & & 0 \end{bmatrix} P^T$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_m = 0$$

主方向的计算

定理 2: 设 $A = [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n]$.

向量 β 是 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 的主方向

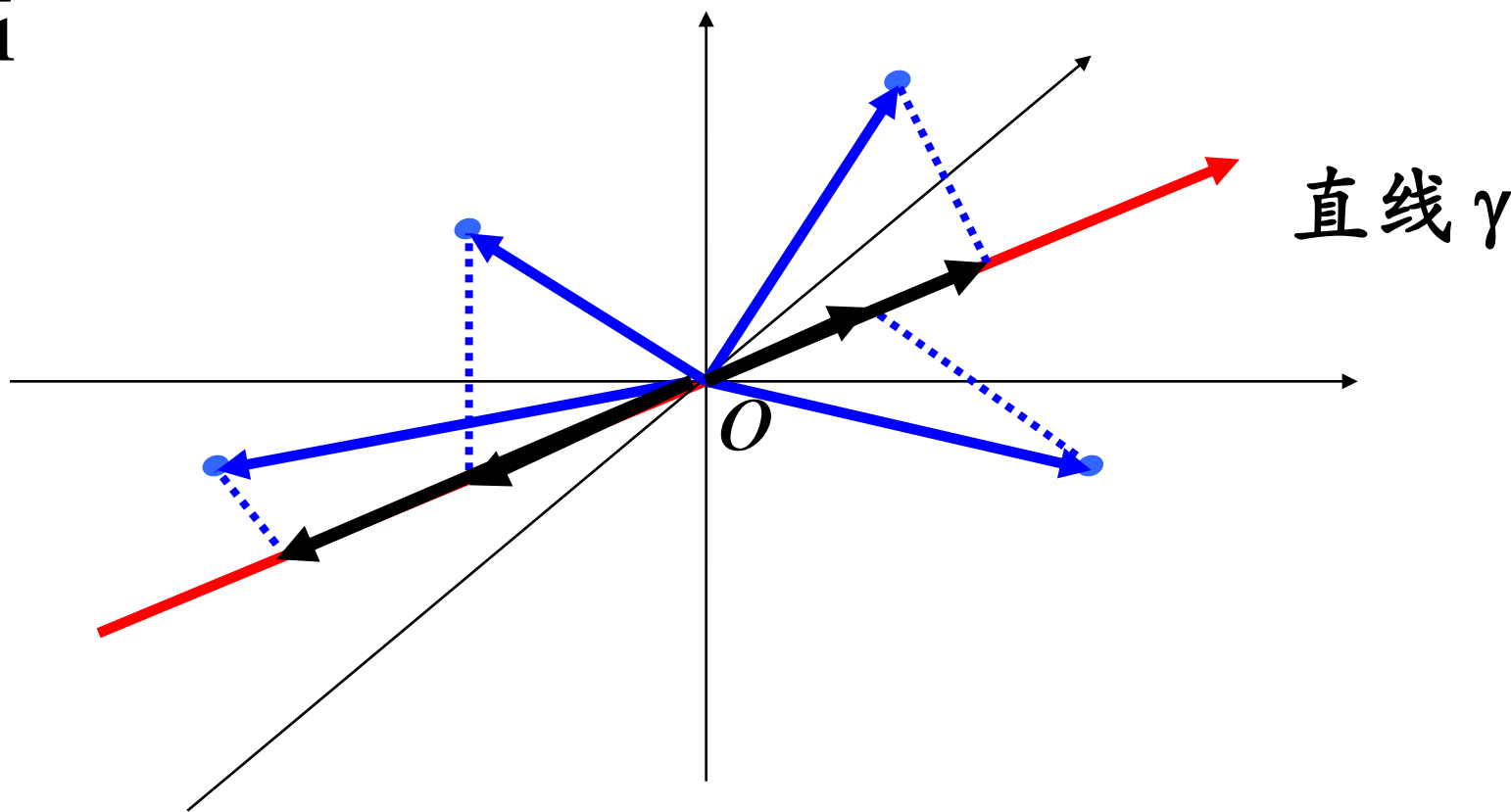
$\Leftrightarrow \beta$ 是属于 AA^T 最大特征值 λ_1 的

特征向量

由勾股定理，距离平方和最小

\Leftrightarrow 投影长度平方和最大

$$\sum_{i=1}^n \|\alpha_i\|^2 = \text{tr}(\mathbf{A}^T \mathbf{A}) = \lambda_1 + \cdots + \lambda_m \quad \leftarrow \text{固定}$$



$$\underbrace{\sum_{i=1}^n \|\alpha_i - (\alpha_i, \beta) \beta\|^2}_{\text{垂线平方和}} + \underbrace{\sum_{i=1}^n (\alpha_i, \beta)^2}_{\text{(正交)投影平方和}} = \underbrace{\sum_{i=1}^n \|\alpha_i\|^2}_{\text{tr}(\mathbf{A}^T \mathbf{A})}$$

勾股定理

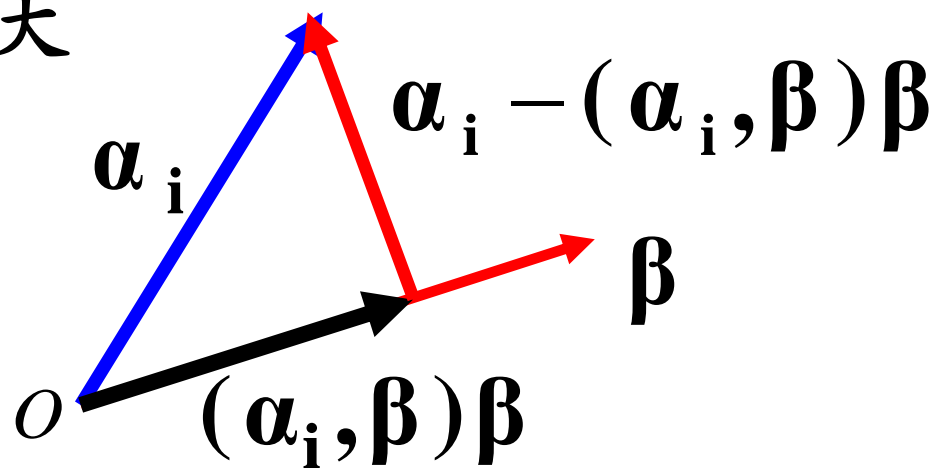
固定

欲使垂线部分的平方和取最小，只需让

正交投影的平方和取最大

不妨设 $\|\beta\| = 1$,

β 在单位球面上变化



$$\underbrace{\sum_{i=1}^n \| \alpha_i - (\alpha_i, \beta) \beta \|^2}_{\text{垂线平方和}} + \underbrace{\sum_{i=1}^n (\alpha_i, \beta)^2}_{\text{正交投影平方和}} = \underbrace{\sum_{i=1}^n \| \alpha_i \|^2}_{\text{tr}(\mathbf{A}^T \mathbf{A})}$$

固定

$$\begin{aligned} \sum_{1 \leq i \leq n} (\alpha_i, \beta)^2 &= \| \mathbf{A}^T \beta \|^2 = (\mathbf{A}^T \beta)^T \mathbf{A}^T \beta \\ &= \beta^T \mathbf{A} \mathbf{A}^T \beta \leq \lambda_1 \| \beta \|^2 = \lambda_1 \end{aligned}$$

取等号 $\Leftrightarrow \beta$ 是 $\mathbf{A} \mathbf{A}^T$ 最大特征值 λ_1 的单位特征向量

$$\mathbf{A} \mathbf{X} = \lambda_n \mathbf{X}$$

$$\mathbf{A} \mathbf{X} = \lambda_1 \mathbf{X}$$

二次型的取值

对任意列向量 $\mathbf{X} \in \mathbf{R}^n$, 有

$$\lambda_n \|\mathbf{X}\|^2 \leq \mathbf{X}^T \mathbf{A} \mathbf{X} \leq \lambda_1 \|\mathbf{X}\|^2$$

这里 λ_1, λ_n 是实对称矩阵 \mathbf{A} 的最大, 最小特征值, 且等号成立当且仅当 \mathbf{X} 落在相应特征值 λ_1, λ_n 的特征子空间内.



在没有度量的情况下,满秩分解 $A = BC$ 是一般矩阵最好的分解: B 的列向量是 A 最靠左的极大无关组,用此无关组线性表出 A 的任一系列向量,表出系数就在 C 的对应列向量的分量里.

C 由行简化阶梯型 $\text{rref}(A)$ 的非零行部分构成.

C 的行向量是 A 行空间一组非常好的基, A 行空间中的向量 y^T 在这组基下的坐标就是 y^T 在主元位置的分量(在这个例子里就是 y^T 的第 1, 2, 4 分量).

以下证明:

$\alpha_1, \dots, \alpha_n$ 在 \mathbf{R}^m 的任一 k 维子空间 V 上的
正交投影长度的平方和

$$\leq \lambda_1 + \lambda_2 + \dots + \lambda_k$$

且等号可在 $V_k = \langle \beta_1, \beta_2, \dots, \beta_k \rangle$ 取到.

(若 $\lambda_k > \lambda_{k+1}$, 则等号只能在 V_k 取到)



另一方面, 由前面 PCA 定理知, 在 \mathbf{R}^m 的所有 k 维子空间中, 向量 $\alpha_1, \dots, \alpha_n$ 到其 k 维主子空间

$$V_k = \langle \beta_1, \dots, \beta_k \rangle$$

(这里 $AA^T = PDP^T$, D 是对角元从大到小排列的对角矩阵, β_1, \dots, β_k 是正交矩阵 P 的前 k 个列向量)

距离的平方和取最小, 此最小值为矩阵 AA^T 的最后 $m - k$ 个特征值之和, 即 $\sigma_{k+1}^2 + \dots + \sigma_m^2$.

故
$$\|A - B\|_F^2 \geq \sigma_{k+1}^2 + \dots + \sigma_r^2.$$