A Short Excursion into Matrix Algebra

矩阵代数简要回顾

2026 年 3 月 12 日

概述 (Outline)

- 矩阵代数简要回顾
  - 基本运算 (Elementary Operations)
  - 谱分解 (Spectral Decompositions)
  - 二次型 (Quadratic Forms)
  - 导数 (Derivatives)
  - 分块矩阵 (Partitioned Matrices)
  - 几何直观 (Geometrical Aspects)

基本运算

- 矩阵 $\mathscr{A}$ 是一个有 n 行 p 列的数表.

$$\mathscr{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{pmatrix} = \left(a_{ij}\right)_{n\times p} = \mathscr{A}_{n\times p}$$

行向量

列向量

- 特殊矩阵：

名称	定义	记号	例子
标量	$p=n=1$	a	3
列向量	$p=1$	$\boldsymbol{a}$	$\begin{pmatrix}1\\3\end{pmatrix}$
行向量	$n=1$	$\boldsymbol{a}^{\mathrm{T}}$	$(1\quad 3)$
1 向量	$\underbrace{(1,\ldots,1)^{\mathrm{T}}}_{n}$	$\mathbf{1}_n$	$\begin{pmatrix}1\\1\end{pmatrix}$
零向量	$\underbrace{(0,\ldots,0)^{\mathrm{T}}}_{n}$	$\mathbf{0}_n$	$\begin{pmatrix}0\\0\end{pmatrix}$

基本运算

- 特殊矩阵：

名称	定义	记号	例子
方阵	$n=p$	$\mathscr{A}(p\times p)$	$\begin{pmatrix}1&2\\3&4\end{pmatrix}$
对角阵	$a_{ij}=0, i\neq j, n=p$	$\operatorname{diag}\left(a_{11}, \ldots, a_{pp}\right)$	$\begin{pmatrix}1&0\\0&2\end{pmatrix}$
单位阵	$\operatorname{diag}(\underbrace{1, \ldots, 1}_{p})$	$\mathscr{I}_p$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$
1 矩阵	$a_{ij}=1, n=p$	$\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$	$\begin{pmatrix}1&1\\1&1\end{pmatrix}$
对称矩阵	$a_{ij}=a_{ji}$	$\mathscr{A}^{\mathrm{T}}=\mathscr{A}$	$\begin{pmatrix}1&2\\2&3\end{pmatrix}$
零矩阵	$a_{ij}=0$	$\mathbf{O}$	$\begin{pmatrix}0&0\\0&0\end{pmatrix}$

基本运算

- 特殊矩阵：

名称	定义	记号	例子
上三角阵	$a_{ij}=0\,,i<j$		$\begin{pmatrix}1&2&4\\0&1&3\\0&0&1\end{pmatrix}$
幂等阵	$\mathscr{A}\mathscr{A}=\mathscr{A}$		$\begin{pmatrix}1&0&0\\0&\frac{1}{2}&\frac{1}{2}\\0&\frac{1}{2}&\frac{1}{2}\end{pmatrix}$
正交阵	$\mathscr{A}^{\mathrm{T}}\mathscr{A}=\mathscr{I}=\mathscr{A}\mathscr{A}^{\mathrm{T}}$		$\begin{pmatrix}\frac{1}{2}&\frac{1}{2}\\\frac{1}{2}&-\frac{1}{2}\end{pmatrix}$

基本运算

- 矩阵运算：$\mathscr{A} = \left(a_{ij}\right)_{n\times p}, \mathscr{B} = \left(b_{ij}\right)$

转置: $\mathscr{A}^{\mathrm{T}} = \left(a_{ji}\right)_{p\times n}$

加法: $\mathscr{A} + \mathscr{B}_{n\times p} = \left(a_{ij} + b_{ij}\right)$

减法: $\mathscr{A} - \mathscr{B}_{n\times p} = \left(a_{ij} - b_{ij}\right)$

数乘: $c \cdot \mathscr{A} = \left(c \cdot a_{ij}\right)$

乘法: $\mathscr{A} \cdot \mathscr{B} = \mathscr{A}_{n\times p}\mathscr{B}_{p\times m} = \mathscr{C}_{n\times m} = \left(c_{ij}\right)_{n\times m} = \left(\sum_{k=1}^{p} a_{ik}b_{kj}\right)$

- 矩阵运算的性质：

$$\mathscr{A} + \mathscr{B} = \mathscr{B} + \mathscr{A}$$
$$\mathscr{A}\left(\mathscr{B} + \mathscr{C}\right) = \mathscr{A}\mathscr{B} + \mathscr{A}\mathscr{C}$$
$$\mathscr{A}\left(\mathscr{B}\mathscr{C}\right) = \left(\mathscr{A}\mathscr{B}\right)\mathscr{C}$$
$$\left(\mathscr{A}^{\mathrm{T}}\right)^{\mathrm{T}} = \mathscr{A}$$
$$\left(\mathscr{A}\mathscr{B}\right)^{\mathrm{T}} = \mathscr{B}^{\mathrm{T}}\mathscr{A}^{\mathrm{T}}$$

基本运算

- 矩阵的特征

  - 秩 (rank)：矩阵 $\mathscr{A}_{n\times p}$ 的秩是其行(列)向量组的最大线性无关组中向量的个数.

  - 线性无关 (linearly independent): 矩阵 $\mathscr{A}_{n\times p}$ 的 k 个行向量 $\boldsymbol{a}_j$ 构成的向量组称为线性无关是指 $\sum_{j=1}^{k} c_j\boldsymbol{a}_j = \boldsymbol{0}_p$ 意味着 $c_j = 0,\ j = 1, 2, \ldots, k$，其中 $c_1, c_2, \ldots, c_k$ 为标量. 换句话说，该向量组的任何一个向量都无法表示为其余 $(k-1)$ 个向量的线性组合.

基本运算

- 矩阵的特征

  - 迹 (trace): 方阵$\mathscr{A}_{p\times p}$ 的迹是其主对角线元素之和 $\mathrm{tr}\,(\mathscr{A}) = \sum_{i=1}^{p} a_{ii}$.

  - 行列式 (determinant): 方阵$\mathscr{A}_{p\times p}$ 的行列式定义为

$$\det(\mathscr{A}) = \left|\mathscr{A}\right| = \sum (-1)^{|\tau|} a_{1\tau_1} a_{2\tau_2} \cdots a_{p\tau_p}$$

$$|\tau| = \begin{cases} 0, & \tau_1\tau_2\cdots\tau_p \text{ 是偶排列} \\ 1, & \tau_1\tau_2\cdots\tau_p \text{ 是奇排列} \end{cases}$$

对 $\{1, 2, \ldots, p\}$ 的全部排列 $\tau_1\tau_2\cdots\tau_p$ 求和.

$$\left|\mathscr{A}^{\mathrm{T}}\right| = \left|\mathscr{A}\right|$$

$$\left|\mathscr{A}\mathscr{B}\right| = \left|\mathscr{A}\right| \cdot \left|\mathscr{B}\right|$$

$$\left|c\mathscr{A}\right| = c^{p}\left|\mathscr{A}\right|$$

基本运算

- 矩阵的特征

  - 逆 (inverse): 方阵 $\mathscr{A}_{p\times p}$ 的行列式 $\left|\mathscr{A}\right| \neq 0$, 则 $\mathscr{A}_{p\times p}$ 的逆矩阵 $\mathscr{A}^{-1}$ 存在

$$\mathscr{A}\mathscr{A}^{-1} = \mathscr{A}^{-1}\mathscr{A} = \mathscr{I}_p$$

对于较小的矩阵而言，$\mathscr{A}^{-1} = \dfrac{\mathscr{C}}{\left|\mathscr{A}\right|}$，其中 $\mathscr{C} = \left(c_{ij}\right)$ 是 $\mathscr{A}$ 的伴随矩阵.

$$c_{ji} = (-1)^{i+j} \begin{vmatrix} a_{11} & \cdots & a_{1(j-1)} & a_{1(j+1)} & \cdots & a_{1p} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{(i-1)1} & \cdots & a_{(i-1)(j-1)} & a_{(i-1)(j+1)} & \cdots & a_{(i-1)p} \\ a_{(i+1)1} & \cdots & a_{(i+1)(j-1)} & a_{(i+1)(j+1)} & \cdots & a_{(i+1)p} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{p1} & \cdots & a_{p(j-1)} & a_{p(j+1)} & \cdots & a_{pp} \end{vmatrix}$$

矩阵 $\mathscr{A}$ 的行列式与其逆矩阵的行列式之间满足关系 $\left|\mathscr{A}^{-1}\right| = \left|\mathscr{A}\right|^{-1}$.

基本运算

- 矩阵的特征

  - 逆 (inverse): 方阵 $\mathscr{A}_{p\times p}$ 的行列式 $\left|\mathscr{A}\right| \neq 0$, 则 $\mathscr{A}_{p\times p}$ 的逆矩阵 $\mathscr{A}^{-1}$ 存在

$$\mathscr{A}\mathscr{A}^{-1} = \mathscr{A}^{-1}\mathscr{A} = \mathscr{I}_p$$

Hilbert 矩阵: $\mathscr{H}_n = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{n+1} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \cdots & \frac{1}{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \frac{1}{n+2} & \cdots & \frac{1}{2n-1} \end{pmatrix}$. 求 $\mathscr{H}_n$ 的逆矩阵 .

基本运算

- 矩阵的特征
  - 逆 (inverse): 方阵 $\mathscr{A}_{p\times p}$ 的行列式 $\left|\mathscr{A}\right| \neq 0$, 则 $\mathscr{A}_{p\times p}$ 的逆矩阵 $\mathscr{A}^{-1}$ 存在

$$\mathscr{A}\mathscr{A}^{-1} = \mathscr{A}^{-1}\mathscr{A} = \mathscr{I}_p$$

```
hilbert = function(n) { i = 1:n; 1 / outer(i - 1, i, `+`) }
h8 = hilbert(8)
h8
```

```
> h8
          [,1]      [,2]      [,3]       [,4]       [,5]       [,6]       [,7]       [,8]
[1,] 1.0000000 0.5000000 0.3333333 0.25000000 0.20000000 0.16666667 0.14285714 0.12500000
[2,] 0.5000000 0.3333333 0.2500000 0.20000000 0.16666667 0.14285714 0.12500000 0.11111111
[3,] 0.3333333 0.2500000 0.2000000 0.16666667 0.14285714 0.12500000 0.11111111 0.10000000
[4,] 0.2500000 0.2000000 0.1666667 0.14285714 0.12500000 0.11111111 0.10000000 0.09090909
[5,] 0.2000000 0.1666667 0.1428571 0.12500000 0.11111111 0.10000000 0.09090909 0.08333333
[6,] 0.1666667 0.1428571 0.1250000 0.11111111 0.10000000 0.09090909 0.08333333 0.07692308
[7,] 0.1428571 0.1250000 0.1111111 0.10000000 0.09090909 0.08333333 0.07692308 0.07142857
[8,] 0.1250000 0.1111111 0.1000000 0.09090909 0.08333333 0.07692308 0.07142857 0.06666667
```

基本运算

- 矩阵的特征
  - 逆 (inverse): 方阵 $\mathscr{A}_{p\times p}$ 的行列式 $\left|\mathscr{A}\right| \neq 0$, 则 $\mathscr{A}_{p\times p}$ 的逆矩阵 $\mathscr{A}^{-1}$ 存在

$$\mathscr{A}\mathscr{A}^{-1} = \mathscr{A}^{-1}\mathscr{A} = \mathscr{I}_p$$

```
inverse_h8 = solve(h8)
inverse_h8
```

```
> inverse_h8
        [,1]      [,2]       [,3]       [,4]        [,5]        [,6]        [,7]       [,8]
[1,]      64     -2016      20160     -92400      221760     -288288      192192     -51480
[2,]   -2016     84672    -952560    4656960   -11642400    15567552   -10594584    2882880
[3,]   20160   -952560   11430720  -58212000   149688000  -204324119   141261119  -38918880
[4,]  -92400   4656960  -58212000  304919999  -800414996  1109908794  -776936155  216215998
[5,]  221760 -11642400  149688000 -800414996  2134439987 -2996753738  2118916783 -594593995
[6,] -288288  15567552 -204324119 1109908793 -2996753738  4249941661 -3030050996  856215352
[7,]  192192 -10594584  141261119 -776936154  2118916782 -3030050996  2175421226 -618377753
[8,]  -51480   2882880  -38918880  216215998  -594593995   856215351  -618377753  176679358
```

基本运算

- 矩阵的特征
  - 逆 (inverse): 方阵 $\mathscr{A}_{p\times p}$ 的行列式 $\left|\mathscr{A}\right| \neq 0$, 则 $\mathscr{A}_{p\times p}$ 的逆矩阵 $\mathscr{A}^{-1}$ 存在

$$\mathscr{A}\mathscr{A}^{-1} = \mathscr{A}^{-1}\mathscr{A} = \mathscr{I}_p$$

```
round(inverse_h8 %*% h8, 3)
```

```
> round(inverse_h8 %*% h8, 3)
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    1    0    0    0    0    0    0    0
[2,]    0    1    0    0    0    0    0    0
[3,]    0    0    1    0    0    0    0    0
[4,]    0    0    0    1    0    0    0    0
[5,]    0    0    0    0    1    0    0    0
[6,]    0    0    0    0    0    1    0    0
[7,]    0    0    0    0    0    0    1    0
[8,]    0    0    0    0    0    0    0    1
```

基本运算

- 矩阵的特征
  - 广义逆 (Generalized Inverse): 更为普遍的一个概念是广义逆 $\mathscr{A}^-$， 它满足

$$\mathscr{A}\mathscr{A}^-\mathscr{A} = \mathscr{A}$$

```
library(MASS)
g_inverse_h8 = ginv(h8)
g_inverse_h8
```

```
> g_inverse_h8
            [,1]        [,2]        [,3]        [,4]        [,5]        [,6]        [,7]        [,8]
[1,]    28.73978   -393.0194    1485.847   -1692.903   -504.9095    1096.219    1019.066   -1044.712
[2,]  -393.01944   7211.2813  -30930.652   38410.109  10112.8407  -25961.633  -23182.639   24889.963
[3,]  1485.84738 -30930.6524  142748.292 -187830.270 -44618.2163  131217.715  113742.530 -126725.011
[4,] -1692.90267  38410.1087 -187830.270  259805.119  55517.2024 -187227.591 -157885.910  182423.576
[5,]  -504.90948  10112.8407  -44618.216   55517.202  15016.8164  -37016.192  -33448.235   35089.987
[6,]  1096.21940 -25961.6326  131217.715 -187227.591 -37016.1922  137785.205  114058.380 -135222.084
[7,]  1019.06577 -23182.6392  113742.530 -157885.910 -33448.2351  114058.380   95980.054 -111215.714
[8,] -1044.71211  24889.9625 -126725.011  182423.576  35089.9867 -135222.084 -111215.714  133139.864
```

基本运算

- 矩阵的特征

  - 广义逆 (Generalized Inverse): 更为普遍的一个概念是广义逆 $\mathscr{A}^-$， 它满足

$$\mathscr{A}\mathscr{A}^-\mathscr{A} = \mathscr{A}$$

```
round(h8 %*% g_inverse_h8 %*% h8, 5)
round(h8, 5)
```

```
> round(h8, 5)
        [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]    [,8]
[1,] 1.00000 0.50000 0.33333 0.25000 0.20000 0.16667 0.14286 0.12500
[2,] 0.50000 0.33333 0.25000 0.20000 0.16667 0.14286 0.12500 0.11111
[3,] 0.33333 0.25000 0.20000 0.16667 0.14286 0.12500 0.11111 0.10000
[4,] 0.25000 0.20000 0.16667 0.14286 0.12500 0.11111 0.10000 0.09091
[5,] 0.20000 0.16667 0.14286 0.12500 0.11111 0.10000 0.09091 0.08333
[6,] 0.16667 0.14286 0.12500 0.11111 0.10000 0.09091 0.08333 0.07692
[7,] 0.14286 0.12500 0.11111 0.10000 0.09091 0.08333 0.07692 0.07143
[8,] 0.12500 0.11111 0.10000 0.09091 0.08333 0.07692 0.07143 0.06667
```

```
> round(h8 %*% g_inverse_h8 %*% h8, 5)
        [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]    [,8]
[1,] 1.00000 0.50000 0.33333 0.25000 0.20000 0.16667 0.14286 0.12500
[2,] 0.50000 0.33333 0.25000 0.20000 0.16667 0.14286 0.12500 0.11111
[3,] 0.33333 0.25000 0.20000 0.16667 0.14286 0.12500 0.11111 0.10000
[4,] 0.25000 0.20000 0.16667 0.14286 0.12500 0.11111 0.10000 0.09091
[5,] 0.20000 0.16667 0.14286 0.12500 0.11111 0.10000 0.09091 0.08333
[6,] 0.16667 0.14286 0.12500 0.11111 0.10000 0.09091 0.08333 0.07692
[7,] 0.14286 0.12500 0.11111 0.10000 0.09091 0.08333 0.07692 0.07143
[8,] 0.12500 0.11111 0.10000 0.09091 0.08333 0.07692 0.07143 0.06667
```

$\mathscr{A}$ 可逆时，广义逆就是逆矩阵.

基本运算

- 矩阵的特征
  - 广义逆 (Generalized Inverse): 更为普遍的一个概念是广义逆 $\mathscr{A}^-$ ， 它满足

$$\mathscr{A}\mathscr{A}^-\mathscr{A} = \mathscr{A}$$

```
x = c(1, 0, 0, 0)
A = matrix(x, nrow=2)
A
g_inverse_A = ginv(A)
g_inverse_A
A %*% g_inverse_A %*% A
```

```
> A
     [,1] [,2]
[1,]    1    0
[2,]    0    0
```

```
> g_inverse_A
     [,1] [,2]
[1,]    1    0
[2,]    0    0
```

```
> A %*% g_inverse_A %*% A
     [,1] [,2]
[1,]    1    0
[2,]    0    0
```

矩阵 $\mathscr{A}$ 不可逆.

矩阵 $\mathscr{A}$ 存在广义逆.

基本运算

- 矩阵的特征
  - 特征值 (eigenvalue) 与特征向量 (eigenvector): 考虑方阵 $\mathscr{A}_{p\times p}$. 若存在数 λ 与向量 $\boldsymbol{\gamma}$ 满足

$\lambda \neq 0$, $\boldsymbol{\gamma} \neq \mathbf{0}$

$$\mathscr{A}\boldsymbol{\gamma} = \lambda\boldsymbol{\gamma}$$

特征值　　特征向量

特征值 $\lambda_1, \lambda_2, \ldots, \lambda_p$ 是 $\left|\mathscr{A} - \lambda \boldsymbol{I}_p\right| = 0$ 的根.

对应于特征值 λ_j 的特征向量 $\boldsymbol{\gamma}_j$ 是 $\left(\mathscr{A} - \lambda_j \boldsymbol{I}_p\right)\boldsymbol{x} = \mathbf{0}$ 的解向量.

记 $\Lambda = \mathrm{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_p\right)$.

基本运算

- 矩阵的特征
  - 特征值 (eigenvalue) 与特征向量 (eigenvector): 考虑方阵 $\mathscr{A}_{p\times p}$. 若存在数 λ 与向量 $\boldsymbol{\gamma}$ 满足

$\lambda \neq 0$, $\boldsymbol{\gamma} \neq \mathbf{0}$

$$\mathscr{A}\boldsymbol{\gamma} = \lambda\boldsymbol{\gamma}$$

特征值

特征向量

$$\left|\mathscr{A}\right| = \left|\Lambda\right| = \prod_{j=1}^{p} \lambda_j$$

$$\mathrm{tr}\,(\mathscr{A}) = \mathrm{tr}\,(\Lambda) = \sum_{j=1}^{p} \lambda_j$$

$\mathscr{A}\mathscr{A} = \mathscr{A}$

幂等阵 $\mathscr{A}$ 的特征值只取 $\{0, 1\}$，因此

$$\mathrm{tr}(\mathscr{A}) = \mathrm{rank}(\mathscr{A}) = \text{非零特征值的个数}$$

基本运算

- 矩阵的特征
  - 特征值 (eigenvalue) 与特征向量 (eigenvector): 考虑方阵 $\mathscr{A}_{p\times p}$. 若存在数 λ 与向量 γ 满足

$$\mathscr{A}\gamma = \lambda\gamma$$

$\lambda \neq 0 , \boldsymbol{\gamma} \neq \mathbf{0}$

特征值

特征向量

```
x = c(1, 0, 0, 0, 1/2, 1/2, 0, 1/2, 1/2)
A = matrix(x, nrow=3, byrow = TRUE)
A
A %*% A
options(digits = 4)
eigen(A)
```

```
> A
     [,1] [,2] [,3]
[1,]    1  0.0  0.0
[2,]    0  0.5  0.5
[3,]    0  0.5  0.5
```

```
> A %*% A
     [,1] [,2] [,3]
[1,]    1  0.0  0.0
[2,]    0  0.5  0.5
[3,]    0  0.5  0.5
```

幂等阵

```
> eigen(A)
eigen() decomposition
$values
[1] 1.000e+00 1.000e+00 5.551e-16

$vectors
       [,1] [,2]    [,3]
[1,] 0.0000    1  0.0000
[2,] 0.7071    0  0.7071
[3,] 0.7071    0 -0.7071
```

$$\lambda_1 = 1,\ \lambda_2 = 1,\ \lambda_3 = 0$$

$$\mathrm{tr}(\mathscr{A}) = \lambda_1 + \lambda_2 + \lambda_3 = 2$$

$$|\mathscr{A}| = \lambda_1\lambda_2\lambda_3 = 0$$

$$\mathrm{rank}(\mathscr{A}) = 2$$

$$\gamma_1 = \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix},\ \gamma_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},\ \gamma_3 = \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}$$

基本运算

- 矩阵特征的性质

$\mathscr{A}_{n\times n}$, $\mathscr{B}_{n\times n}$, $c \in \mathbb{R}$

$$\operatorname{tr}\left(\mathscr{A}+\mathscr{B}\right) = \operatorname{tr}(\mathscr{A}) + \operatorname{tr}\left(\mathscr{B}\right)$$

$$\operatorname{tr}(c\mathscr{A}) = c\operatorname{tr}(\mathscr{A})$$

$$\left|c\mathscr{A}\right| = c^n\left|\mathscr{A}\right|$$

$$\left|\mathscr{A}\mathscr{B}\right| = \left|\mathscr{B}\mathscr{A}\right| = \left|\mathscr{A}\right|\left|\mathscr{B}\right|$$

基本运算

- 矩阵特征的性质

$\mathscr{A}_{n\times p}$, $\mathscr{B}_{p\times n}$

$$\operatorname{tr}\left(\mathscr{A}\cdot\mathscr{B}\right)=\operatorname{tr}\left(\mathscr{B}\cdot\mathscr{A}\right)$$

$$\operatorname{rank}(\mathscr{A})\leq\min(n,p)$$

$$\operatorname{rank}(\mathscr{A})\geq 0$$

$$\operatorname{rank}(\mathscr{A})=\operatorname{rank}\left(\mathscr{A}^{\mathrm{T}}\right)$$

$$\operatorname{rank}\left(\mathscr{A}^{\mathrm{T}}\mathscr{A}\right)=\operatorname{rank}(\mathscr{A})$$

$$\operatorname{rank}\left(\mathscr{A}+\mathscr{B}\right)\leq\operatorname{rank}(\mathscr{A})+\operatorname{rank}\left(\mathscr{B}\right)$$

$$\operatorname{rank}\left(\mathscr{A}\mathscr{B}\right)\leq\min\left\{\operatorname{rank}(\mathscr{A}),\operatorname{rank}\left(\mathscr{B}\right)\right\}$$

基本运算

- 矩阵特征的性质

$\mathscr{A}_{n\times p}$, $\mathscr{B}_{p\times q}$, $\mathscr{C}_{q\times n}$

$$\operatorname{tr}\left(\mathscr{A}\mathscr{B}\mathscr{C}\right) = \operatorname{tr}\left(\mathscr{B}\mathscr{C}\mathscr{A}\right) = \operatorname{tr}\left(\mathscr{C}\mathscr{A}\mathscr{B}\right)$$

$$\operatorname{rank}\left(\mathscr{A}\mathscr{B}\mathscr{C}\right) = \operatorname{rank}\left(\mathscr{B}\right)$$ 其中 $\mathscr{A}$, $\mathscr{C}$ 为非奇异矩阵

$\mathscr{A}_{p\times p}$

$$\left|\mathscr{A}^{-1}\right| = \left|\mathscr{A}\right|^{-1}$$

$\operatorname{rank}(\mathscr{A}) = p$ 充分必要条件为 $\mathscr{A}$ 非奇异

谱分解 (Spectral Decompositions)

- 定理 2.1 (Jordan 分解)

任一对称矩阵 $\mathscr{A}_{p\times p}$ 均可表示为

$$\mathscr{A} = \Gamma \Lambda \Gamma^{\mathrm{T}} = \sum_{j=1}^{p} \lambda_j \gamma_j \gamma_j^{\mathrm{T}}$$

其中

$$\Lambda = \mathrm{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_p\right)$$

且

$$\Gamma = \left(\gamma_1, \gamma_2, \ldots, \gamma_p\right)$$

是由 $\mathscr{A}$ 的特征向量构成的一个正交矩阵.

谱分解 (Spectral Decompositions)

对称矩阵

- 矩阵 $\mathscr{A}_{p\times p}$ 的幂

$\lambda_j > 0$, $j = 1, 2, \ldots, p$

定理 2.1

$\Longrightarrow$ $\mathscr{A} = \Gamma\Lambda\Gamma^{\mathrm{T}}$

$$\mathscr{A}^{\alpha} \triangleq \Gamma\Lambda^{\alpha}\Gamma^{\mathrm{T}}, \quad \alpha \in \mathbb{R}$$

$\Lambda^{\alpha} = \mathrm{diag}\left(\lambda_1^{\alpha}, \lambda_2^{\alpha}, \ldots, \lambda_p^{\alpha}\right)$

$\Longrightarrow$ $\mathscr{A}^{-1} = \Gamma\Lambda^{-1}\Gamma^{\mathrm{T}}$

谱分解 (Spectral Decompositions)

- 定理 2.2 (奇异值分解)

秩为 r 的任一矩阵 $\mathscr{A}_{n\times p}$ 可分解为

$$\mathscr{A} = \Gamma \, \Lambda \, \Delta^{\mathrm{T}}$$

其中 $\Gamma_{n\times r}$ 为 $n\times r$ 矩阵, $\Delta_{p\times r}$ 为 $p\times r$ 矩阵. Γ 与 Δ 均为列正交矩阵, 即

$$\Gamma^{\mathrm{T}}\Gamma = \Delta^{\mathrm{T}}\Delta = \mathscr{I}_r$$

且

$$\Lambda = \mathrm{diag}\left(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_r^{1/2}\right), \qquad \lambda_j > 0$$

这里的 $\lambda_1, \lambda_2, \ldots, \lambda_r$ 是矩阵 $\mathscr{A}\mathscr{A}^{\mathrm{T}}$ 与 $\mathscr{A}^{\mathrm{T}}\mathscr{A}$ 的非零特征值. Γ 与 Δ 是矩阵 $\mathscr{A}\mathscr{A}^{\mathrm{T}}$ 与 $\mathscr{A}^{\mathrm{T}}\mathscr{A}$ 对应的 r 个特征向量构成的矩阵.

$\Longrightarrow$ $\mathscr{A}^{-} = \Delta\,\Lambda^{-1}\,\Gamma^{\mathrm{T}}$

广义逆

$$\begin{aligned}\mathscr{A}\mathscr{A}^{-}\mathscr{A} &= \left(\Gamma\Lambda\Delta^{\mathrm{T}}\right)\left(\Delta\Lambda^{-1}\Gamma^{\mathrm{T}}\right)\left(\Gamma\Lambda\Delta^{\mathrm{T}}\right)\\ &= \Gamma\Lambda\left(\Delta^{\mathrm{T}}\Delta\right)\Lambda^{-1}\left(\Gamma^{\mathrm{T}}\Gamma\right)\Lambda\,\Delta^{\mathrm{T}}\\ &= \Gamma\Lambda\,\Lambda^{-1}\Lambda\,\Delta^{\mathrm{T}} = \Gamma\Lambda\,\Delta^{\mathrm{T}} = \mathscr{A}\end{aligned}$$

广义逆不唯一.

谱分解 (Spectral Decompositions)

```
x = c(1, 0, 0, 0)
A = matrix(x, nrow=2)
A
g_inverse_A = ginv(A)
g_inverse_A
A %*% g_inverse_A %*% A
```

$\mathscr{A}$

```
> A
     [,1] [,2]
[1,]    1    0
[2,]    0    0
```

$\mathscr{A}^-$

```
> g_inverse_A
     [,1] [,2]
[1,]    1    0
[2,]    0    0
```

$\mathscr{A}\mathscr{A}^-\mathscr{A} = \mathscr{A}$

```
> A %*% g_inverse_A %*% A
     [,1] [,2]
[1,]    1    0
[2,]    0    0
```

```
y = c(1, 0, 0, 8)
B = matrix(y, nrow=2)
B
A %*% B %*% A
```

$\mathscr{B}$ 也是 $\mathscr{A}$ 的广义逆

```
> B
     [,1] [,2]
[1,]    1    0
[2,]    0    8
```

```
> A %*% B %*% A
     [,1] [,2]
[1,]    1    0
[2,]    0    0
```

$\mathscr{A}\mathscr{B}\mathscr{A} = \mathscr{A}$

二次型 (Quadratic Forms)

- 二次型

$$Q(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}} \mathscr{A}\, \boldsymbol{x} = \sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij} x_i x_j$$

$\mathscr{A}$ 为 $(p \times p)$ 对称矩阵

$\boldsymbol{x} \in \mathbb{R}^p$

- 二次型与矩阵的正定性 (definiteness)

正定 (positive definite) 二次型： $\forall \boldsymbol{x} \neq \boldsymbol{0}, \quad \boldsymbol{x}^{\mathrm{T}} \mathscr{A}\, \boldsymbol{x} = Q(\boldsymbol{x}) > 0$

$\mathscr{A}$ 为正定矩阵 $(\mathscr{A} > 0)$

半正定 (positive semidefinite) 二次型： $\forall \boldsymbol{x} \neq \boldsymbol{0}, \quad \boldsymbol{x}^{\mathrm{T}} \mathscr{A}\, \boldsymbol{x} = Q(\boldsymbol{x}) \geq 0$

非负定二次型

$\mathscr{A}$ 为半正定矩阵或非负定矩阵 $(\mathscr{A} \geq 0)$

二次型 (Quadratic Forms)

- 二次型总可以对角化.

定理 2.3 如果 $\mathscr{A}$ 为对称矩阵，$Q(x)=\boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x}$ 是对应的二次型，则存在变换

$\boldsymbol{x}\mapsto\Gamma^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{y}$ 使得

$$\boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x}=\sum_{i=1}^{p}\lambda_i y_i^2$$

其中 λ_i 是 $\mathscr{A}$ 的特征值.

证明：

$$\mathscr{A}=\Gamma\Lambda\Gamma^{\mathrm{T}}$$

$$\boldsymbol{y}=\Gamma^{\mathrm{T}}\boldsymbol{x}$$

$$\Longrightarrow\quad \boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}\left(\Gamma\Lambda\Gamma^{\mathrm{T}}\right)\boldsymbol{x}$$

$$=\left(\Gamma^{\mathrm{T}}\boldsymbol{x}\right)^{\mathrm{T}}\Lambda\left(\Gamma^{\mathrm{T}}\boldsymbol{x}\right)$$

$$=\boldsymbol{y}^{\mathrm{T}}\Lambda\boldsymbol{y}=\sum_{i=1}^{p}\lambda_i y_i^2$$

- **定理 2.1** (Jordan 分解)

任一对称矩阵 $\mathscr{A}_{p\times p}$ 均可表示为

$$\mathscr{A}=\Gamma\Lambda\Gamma^{\mathrm{T}}=\sum_{j=1}^{p}\lambda_j\gamma_j\gamma_j^{\mathrm{T}}$$

其中

$$\Lambda=\operatorname{diag}\left(\lambda_1,\lambda_2,\ldots,\lambda_p\right)$$

且

$$\Gamma=\left(\gamma_1,\gamma_2,\ldots,\gamma_p\right)$$

是由 $\mathscr{A}$ 的特征向量构成的一个正交矩阵.

二次型 (Quadratic Forms)

- 定理 2.4 $\mathscr{A} > 0$ 的充分必要条件是所有 $\lambda_i > 0,\ i = 1, 2, \ldots, p.$

证明:

> 定理 2.3 如果 $\mathscr{A}$ 为对称矩阵，$Q(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x}$ 是对应的二次型，则存在变换 $\boldsymbol{x} \mapsto \Gamma^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ 使得
>
> $$\boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x} = \sum_{i=1}^{p} \lambda_i y_i^2$$
>
> 其中 λ_i 是 $\mathscr{A}$ 的特征值.

$$\forall \boldsymbol{x} \neq \boldsymbol{0}, \quad Q(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_p y_p^2 > 0$$

- 推论 2.1 若 $\mathscr{A} > 0$，则 $\mathscr{A}^{-1}$ 存在，且 $\left|\mathscr{A}\right| > 0.$

二次型 (Quadratic Forms)

- 在多元数据的统计分析中，我们关注有约束条件下的二次型最大化.

定理 2.5 若 $\mathscr{A}$ 与 $\mathscr{B}$ 为对称矩阵且 $\mathscr{B}>0$，则 $\dfrac{\boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\mathscr{B}\,\boldsymbol{x}}$ 的最大值等于 $\mathscr{B}^{-1}\mathscr{A}$ 的最大特征值. 更一般地，我们有

$$\max_{\boldsymbol{x}} \frac{\boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\mathscr{B}\,\boldsymbol{x}} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p = \min_{\boldsymbol{x}} \frac{\boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\mathscr{B}\,\boldsymbol{x}}$$

其中 $\lambda_1, \lambda_2, \ldots, \lambda_p$ 是 $\mathscr{B}^{-1}\mathscr{A}$ 的特征值. 使得 $\dfrac{\boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\mathscr{B}\,\boldsymbol{x}}$ 达到最大(最小)的向量是 $\mathscr{B}^{-1}\mathscr{A}$ 的最大(最小)特征值对应的 $\mathscr{B}^{-1}\mathscr{A}$ 的特征向量. 如果 $\boldsymbol{x}^{\mathrm{T}}\mathscr{B}\,\boldsymbol{x}=1$，我们则有

$$\max_{\boldsymbol{x}} \boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p = \min_{\boldsymbol{x}} \boldsymbol{x}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{x}$$

导数 (derivatives)

$$y = f(\boldsymbol{x}) = f\left(x_1, x_2, \ldots, x_p\right)$$

$$f:\ \mathbb{R}^p \to \mathbb{R}, \qquad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$$

f 的梯度 (gradient)

$$\Longrightarrow \quad \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \triangleq \begin{pmatrix} \dfrac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_p} \end{pmatrix}, \quad \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}} \triangleq \left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \ldots, \frac{\partial f(\boldsymbol{x})}{\partial x_p} \right)$$

f 的 Hessian 矩阵

$$\Longrightarrow \quad \frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x} \partial \boldsymbol{x}^{\mathrm{T}}} \triangleq \begin{pmatrix} \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_p} \\ \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_p \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_p \partial x_2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_p \partial x_p} \end{pmatrix}$$

导数 (derivatives)

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}, \quad \mathscr{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix} = \mathscr{A}^{\mathrm{T}}, \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$$

$$\Longrightarrow \quad \frac{\partial \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}^{\mathrm{T}} \boldsymbol{a}}{\partial \boldsymbol{x}} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \boldsymbol{a}$$

$$\boldsymbol{x}^{\mathrm{T}} \mathscr{A} \boldsymbol{x} = \sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij} x_i x_j$$

$$\Longrightarrow \quad \frac{\partial \boldsymbol{x}^{\mathrm{T}} \mathscr{A} \boldsymbol{x}}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial \boldsymbol{x}^{\mathrm{T}} \mathscr{A} \boldsymbol{x}}{\partial x_1} \\ \frac{\partial \boldsymbol{x}^{\mathrm{T}} \mathscr{A} \boldsymbol{x}}{\partial x_2} \\ \vdots \\ \frac{\partial \boldsymbol{x}^{\mathrm{T}} \mathscr{A} \boldsymbol{x}}{\partial x_p} \end{pmatrix} = \begin{pmatrix} 2\left(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p\right) \\ 2\left(a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p\right) \\ \vdots \\ 2\left(a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_p\right) \end{pmatrix} = 2\mathscr{A}\boldsymbol{x}$$

$\Longrightarrow$ 二次型的 Hessian 矩阵: $\dfrac{\partial^2 \boldsymbol{x}^{\mathrm{T}} \mathscr{A} \boldsymbol{x}}{\partial \boldsymbol{x} \partial \boldsymbol{x}^{\mathrm{T}}} = 2\mathscr{A}$

分块矩阵 (partitioned matrices)

$$\mathscr{A}_{n\times p} = \left(\begin{array}{ccc|ccc} a_{1,\,1} & \cdots & a_{1,\,p_1} & a_{1,\,p_1+1} & \cdots & a_{1,\,p_1+p_2} \\ \vdots & \mathscr{A}_{11} & \vdots & \vdots & \mathscr{A}_{12} & \vdots \\ a_{n_1,\,1} & \cdots & a_{n_1,\,p_1} & a_{n_1,\,p_1+1} & \cdots & a_{n_1\ p_1+p_2} \\ \hline a_{n_1+1,\,1} & \cdots & a_{n_1+1,\,p_1} & a_{n_1+1,\,p_1+1} & \cdots & a_{n_1+1\ p_1+p_2} \\ \vdots & \mathscr{A}_{21} & \vdots & \vdots & \mathscr{A}_{22} & \vdots \\ a_{n_1+n_2,\,1} & \cdots & a_{n_1+n_2,\,p_1} & a_{n_1+n_2,\,p_1+1} & \cdots & a_{n_1+n_2,\,p_1+p_2} \end{array}\right) = \begin{pmatrix} \mathscr{A}_{11} & \mathscr{A}_{12} \\ \mathscr{A}_{21} & \mathscr{A}_{22} \end{pmatrix}$$

$$\begin{cases} n_1 + n_2 = n \\ p_1 + p_2 = p \end{cases}$$

- 如果对矩阵 $\mathscr{B}_{n\times p}$ 作同样分块: $\mathscr{B} = \begin{pmatrix} \mathscr{B}_{11} & \mathscr{B}_{12} \\ \mathscr{B}_{21} & \mathscr{B}_{22} \end{pmatrix}$

$$\Longrightarrow \qquad \mathscr{A} + \mathscr{B} = \begin{pmatrix} \mathscr{A}_{11} + \mathscr{B}_{11} & \mathscr{A}_{12} + \mathscr{B}_{12} \\ \mathscr{A}_{21} + \mathscr{B}_{21} & \mathscr{A}_{22} + \mathscr{B}_{22} \end{pmatrix}$$

$$\mathscr{B}^{\mathrm{T}} = \begin{pmatrix} \mathscr{B}_{11}^{\mathrm{T}} & \mathscr{B}_{21}^{\mathrm{T}} \\ \mathscr{B}_{12}^{\mathrm{T}} & \mathscr{B}_{22}^{\mathrm{T}} \end{pmatrix}$$

$$\mathscr{A}\mathscr{B}^{\mathrm{T}} = \begin{pmatrix} \mathscr{A}_{11}\mathscr{B}_{11}^{\mathrm{T}} + \mathscr{A}_{12}\mathscr{B}_{12}^{\mathrm{T}} & \mathscr{A}_{11}\mathscr{B}_{21}^{\mathrm{T}} + \mathscr{A}_{12}\mathscr{B}_{22}^{\mathrm{T}} \\ \mathscr{A}_{21}\mathscr{B}_{11}^{\mathrm{T}} + \mathscr{A}_{22}\mathscr{B}_{12}^{\mathrm{T}} & \mathscr{A}_{21}\mathscr{B}_{21}^{\mathrm{T}} + \mathscr{A}_{22}\mathscr{B}_{22}^{\mathrm{T}} \end{pmatrix}$$

分块矩阵 (partitioned matrices)

$$\boxed{\mathscr{A}_{p\times p}} = \begin{pmatrix} \boxed{\mathscr{A}_{11}} & \mathscr{A}_{12} \\ \mathscr{A}_{21} & \boxed{\mathscr{A}_{22}} \end{pmatrix}$$

Square matrix

- 如果 $\mathscr{A}$ 非奇异 $\left(\mathscr{A}\mathscr{A}^{-1} = \mathscr{I}_p\right)$:

$$\Longrightarrow \qquad \mathscr{A}^{-1} = \begin{pmatrix} \mathscr{A}^{11} & \mathscr{A}^{12} \\ \mathscr{A}^{21} & \mathscr{A}^{22} \end{pmatrix}$$

$$\begin{cases} \mathscr{A}^{11} = \left(\mathscr{A}_{11} - \mathscr{A}_{12}\mathscr{A}_{22}^{-1}\mathscr{A}_{21}\right)^{-1} \overset{\text{def}}{=} \left(\mathscr{A}_{11\cdot 2}\right)^{-1} \\ \mathscr{A}^{12} = -\left(\mathscr{A}_{11\cdot 2}\right)^{-1}\mathscr{A}_{12}\mathscr{A}_{22}^{-1} \\ \mathscr{A}^{21} = -\mathscr{A}_{22}^{-1}\mathscr{A}_{21}\left(\mathscr{A}_{11\cdot 2}\right)^{-1} \\ \mathscr{A}^{22} = \mathscr{A}_{22}^{-1} + \mathscr{A}_{22}^{-1}\mathscr{A}_{21}\left(\mathscr{A}_{11\cdot 2}\right)^{-1}\mathscr{A}_{12}\mathscr{A}_{22}^{-1} \end{cases}$$

$$\begin{cases} \mathscr{A}_{11\cdot 2} \overset{\text{def}}{=} \mathscr{A}_{11} - \mathscr{A}_{12}\mathscr{A}_{22}^{-1}\mathscr{A}_{21} \\ \mathscr{A}_{22\cdot 1} \overset{\text{def}}{=} \mathscr{A}_{22} - \mathscr{A}_{21}\mathscr{A}_{11}^{-1}\mathscr{A}_{12} \end{cases}$$

分块矩阵 (partitioned matrices)

- 以下结论很有用

  - 如何 $\mathscr{A}_{11}$ 非奇异，则

$$\left|\mathscr{A}\right| = \left|\mathscr{A}_{11}\right| \left|\mathscr{A}_{22} - \mathscr{A}_{21}\mathscr{A}_{11}^{-1}\mathscr{A}_{12}\right| = \left|\mathscr{A}_{11}\right| \left|\mathscr{A}_{22\cdot 1}\right|$$

  - 如果 $\mathscr{A}_{22}$ 非奇异，则

$$\left|\mathscr{A}\right| = \left|\mathscr{A}_{22}\right| \left|\mathscr{A}_{11} - \mathscr{A}_{12}\mathscr{A}_{22}^{-1}\mathscr{A}_{21}\right| = \left|\mathscr{A}_{22}\right| \left|\mathscr{A}_{11\cdot 2}\right|$$

分块矩阵 (partitioned matrices)

- **定理 2.6** 设 $\mathscr{A}_{n\times p}$ 为 $n\times p$ 矩阵，$\mathscr{B}_{p\times n}$ 为 $p\times n$ 矩阵，则 $\mathscr{A}\mathscr{B}$ 与 $\mathscr{B}\mathscr{A}$ 有相同的非零特征值，且特征值的重数亦相同. 若 $\boldsymbol{x}$ 是矩阵 $\mathscr{A}\mathscr{B}$ 对应某个非零特征值 $\lambda\neq 0$ 的一个特征向量，则 $\boldsymbol{y}=\mathscr{B}\boldsymbol{x}$ 是矩阵 $\mathscr{B}\mathscr{A}$ 对应该特征值的一个特征向量.

- **推论 2.2** 设 $\mathscr{A}_{n\times p}$ 与 $\mathscr{B}_{q\times n}$ 分别是 $n\times p$、$q\times n$ 矩阵，$\boldsymbol{a}_{p\times 1}$ 与 $\boldsymbol{b}_{q\times 1}$ 分别是 $p\times 1$、$q\times 1$ 列向量，则

$$\operatorname{rank}\left(\mathscr{A}\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}\mathscr{B}\right)\leq 1.$$

如果存在非零特征值，则它等于 $\boldsymbol{b}^{\mathrm{T}}\mathscr{B}\mathscr{A}\boldsymbol{a}$，对应的特征向量是 $\mathscr{A}\boldsymbol{a}$.

几何直观 (geometrical aspects)

- 距离 (distance)

  - 定义: $\mathbb{R}^p$ 中的两点 $\boldsymbol{x}, \boldsymbol{y}$ 的**距离** $d\left(\boldsymbol{x}, \boldsymbol{y}\right)$ 定义为函数 $d: \mathbb{R}^{2p} \to \mathbb{R}_+$，它满足

$$\begin{cases} d(\boldsymbol{x}, \boldsymbol{y}) > 0, & \forall \boldsymbol{x} \neq \boldsymbol{y} \\ d(\boldsymbol{x}, \boldsymbol{y}) = 0, & \text{当且仅当 } \boldsymbol{x} = \boldsymbol{y} \\ d(\boldsymbol{x}, \boldsymbol{y}) \leq d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{z}, \boldsymbol{y}), & \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \end{cases}$$

  - 欧氏距离 (Euclidean distance):

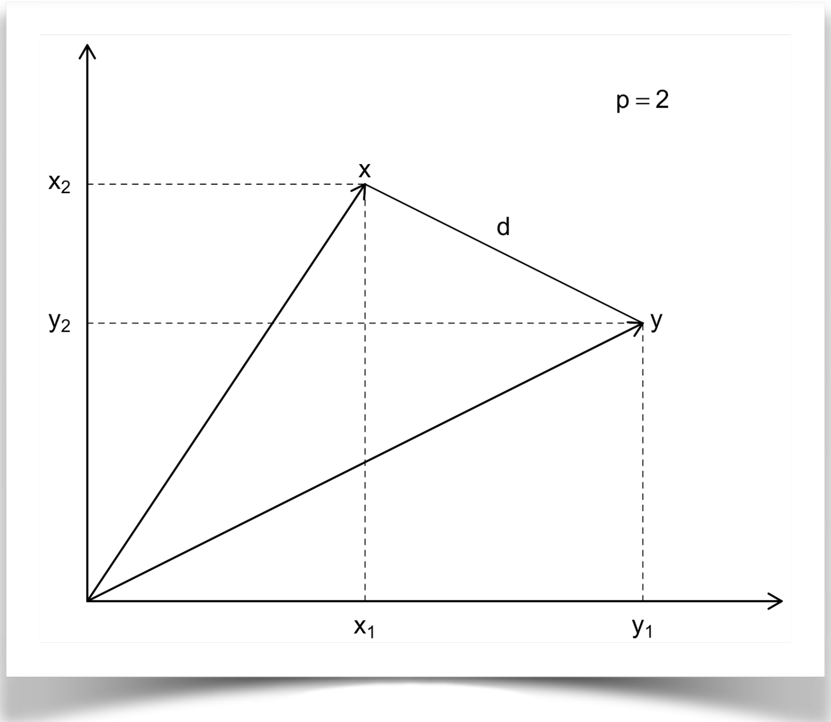
正定矩阵 ($\mathcal{A} > 0$)

$$d^2(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})^{\mathrm{T}} \boxed{\mathcal{A}} (\boldsymbol{x} - \boldsymbol{y})$$

度量 (metric)

$\mathcal{A} = \mathcal{I}_p$

$$d^2(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{p} \left(x_i - y_i\right)^2$$

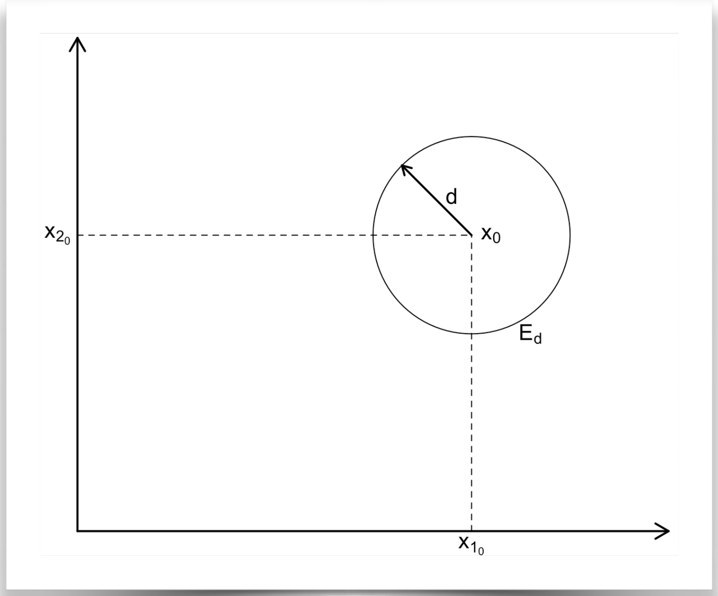


几何直观 (geometrical aspects)

- 距离 (distance)
  - 欧氏 $\mathscr{I}_p$ 距离，到点 $\boldsymbol{x}_0$ 的**等距曲线** (iso-distance curve)

中心点为 $\boldsymbol{x}_0$ 的球面，其中 d 为常数.

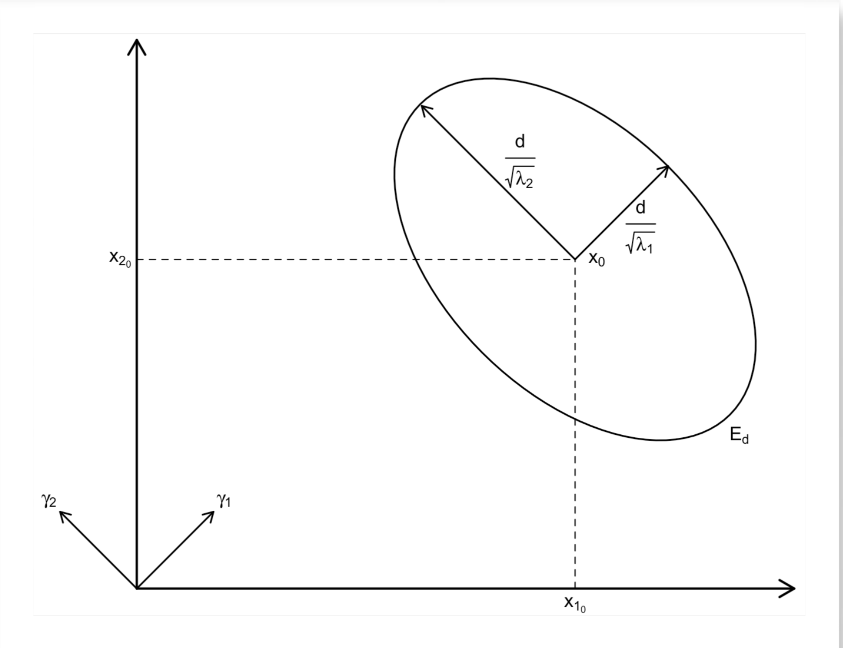
$$E_d = \left\{ \boldsymbol{x} \in \mathbb{R}^p \,\middle|\, (\boldsymbol{x} - \boldsymbol{x}_0)^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{x}_0) = d^2 \right\}$$



  - 一般的等距曲线

$$E_d = \left\{ \boldsymbol{x} \in \mathbb{R}^p \,\middle|\, (\boldsymbol{x} - \boldsymbol{x}_0)^{\mathrm{T}} \mathscr{A} (\boldsymbol{x} - \boldsymbol{x}_0) = d^2 \right\}$$

中心点为 $\boldsymbol{x}_0$ 的椭球面，其中 $\mathscr{A} > 0$ 为正定矩阵，d 为常数.



几何直观 (geometrical aspects)

- **定理 2.7** 设 $\gamma_1, \gamma_2, \ldots, \gamma_p$ 是矩阵 $\mathscr{A}$ 的特征值 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ 对应的正交特征向量.

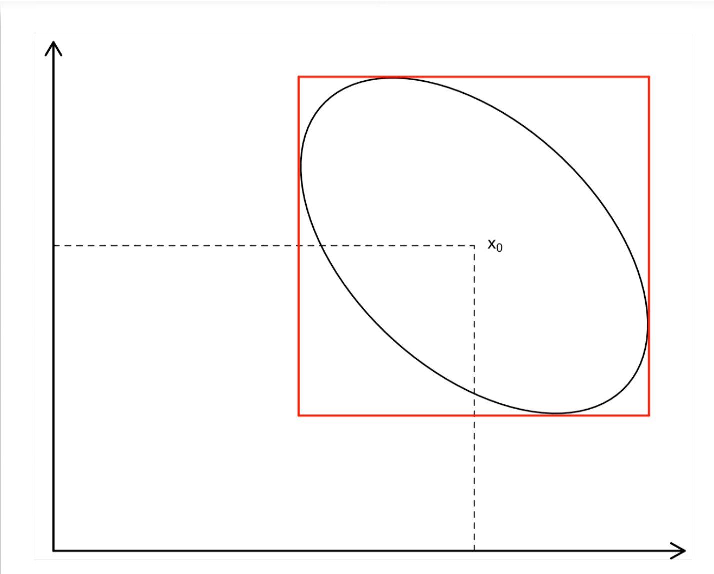
(i) 椭球面 E_d 的主轴位于 γ_i , $i = 1, 2, \ldots, p$ 的方向.

(ii) 半轴的长度为 $\sqrt{\dfrac{d^2}{\lambda_i}}$, $i = 1, 2, \ldots, p$.

(iii) 包围椭球体 E_d 的外接矩形由以下不等式确定:

$$x_{0i} - \sqrt{d^2 a^{ii}} \leq x_i \leq x_{0i} + \sqrt{d^2 a^{ii}}\ , \quad i = 1, 2, \ldots, p$$

其中 a^{ii} 是 $\mathscr{A}^{-1}$ 的 (i, i) 元素. 包围椭球体 E_d 的外接矩形我们指的是边与坐标轴平行的矩形.

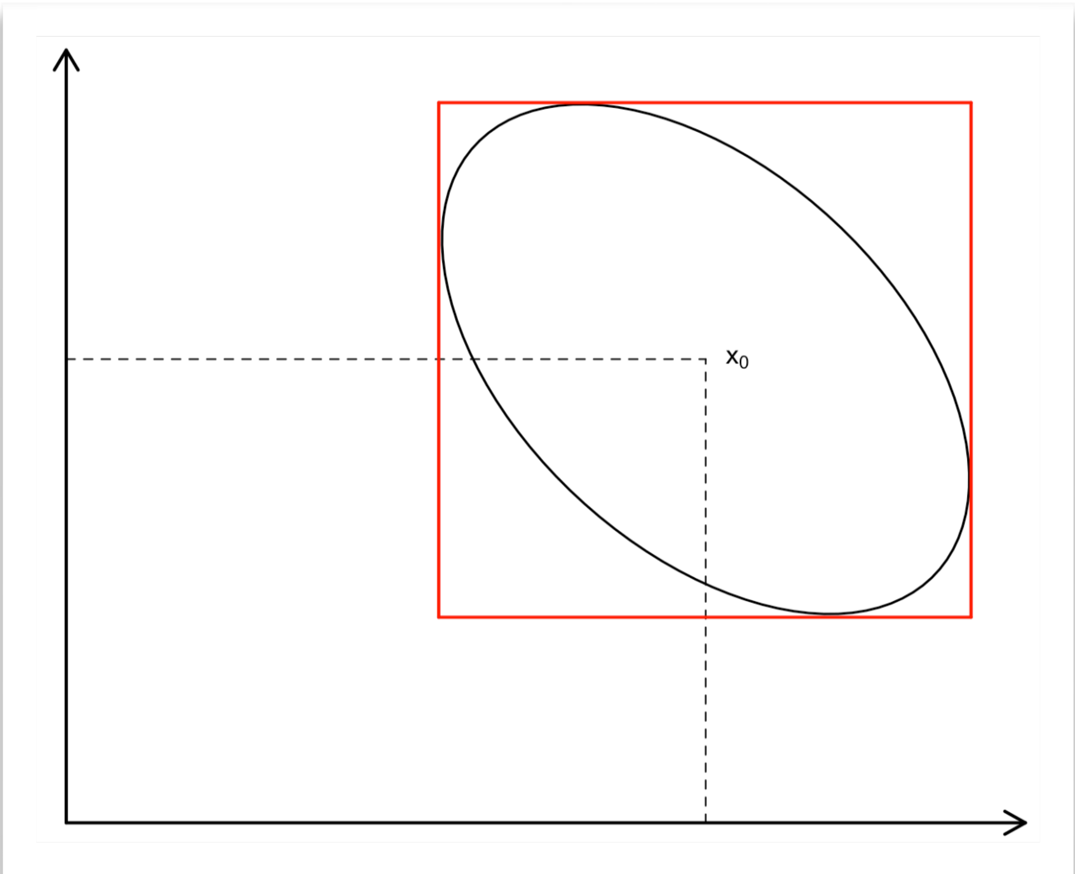


几何直观 (geometrical aspects)

- 不难确定椭球体与其平行于坐标轴的外接矩形的切点.
- 我们来确定第 j 个坐标轴方向 (正方向) 上的切点坐标.
- 为简单起见，我们假设椭球以原点 $(\boldsymbol{x}_0 = \mathbf{0})$ 为中心. 否则，只需将矩形平移 $\boldsymbol{x}_0$ 即可.
- 切点的坐标由以下问题的解给出：

$$\boldsymbol{x} = \arg \max_{\boldsymbol{x}^{\mathrm{T}} \mathscr{A} \boldsymbol{x} = d^2} \boxed{\boldsymbol{e}_j^{\mathrm{T}}} \boldsymbol{x}$$

单位矩阵 $\mathscr{I}_p$ 的第 j 列



几何直观 (geometrical aspects)

- 可通过拉格朗日乘数法来求解

$$L = \boldsymbol{e}_j^{\mathrm{T}}\boldsymbol{x} - \lambda\left(\boldsymbol{x}^{\mathrm{T}}\mathscr{A}\boldsymbol{x} - d^2\right)$$

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}, \quad \mathscr{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix} = \mathscr{A}^{\mathrm{T}}, \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$$

$$\Longrightarrow \quad \frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{a}}{\partial \boldsymbol{x}} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \boldsymbol{a}, \quad \frac{\partial \boldsymbol{x}^{\mathrm{T}}\mathscr{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = 2\mathscr{A}\boldsymbol{x}$$

$$\begin{cases} \dfrac{\partial L}{\partial \boldsymbol{x}} = \boldsymbol{e}_j - 2\lambda\mathscr{A}\boldsymbol{x} = \boldsymbol{0} \\[2ex] \dfrac{\partial L}{\partial \lambda} = \boldsymbol{x}^{\mathrm{T}}\mathscr{A}\boldsymbol{x} - d^2 = 0 \end{cases}$$

$\mathscr{A}^{-1}$ 的 (i, j) 元素

$$\Longrightarrow \quad \boldsymbol{x} = \frac{1}{2\lambda}\mathscr{A}^{-1}\boldsymbol{e}_j \quad \Longrightarrow \quad x_i = \frac{1}{2\lambda}a^{ij}, \; i = 1, 2, \ldots, p$$

几何直观 (geometrical aspects)

$$\begin{cases} \dfrac{\partial L}{\partial \boldsymbol{x}} = \boldsymbol{e}_j - 2\lambda \mathscr{A}\boldsymbol{x} = \boldsymbol{0} \\ \dfrac{\partial L}{\partial \lambda} = \boldsymbol{x}^{\mathrm{T}}\mathscr{A}\boldsymbol{x} - d^2 = 0 \end{cases}$$

左乘 $\boldsymbol{x}^{\mathrm{T}}$

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{e}_j - 2\lambda \boldsymbol{x}^{\mathrm{T}}\mathscr{A}\boldsymbol{x} = \boldsymbol{0}$$

$$\Longrightarrow \quad \boldsymbol{x}^{\mathrm{T}}\boldsymbol{e}_j = 2\lambda d^2 \quad \Longrightarrow \quad x_j = 2\lambda d^2$$

$\mathscr{A}^{-1}$ 的 (i, j) 元素

$$x_i = \frac{1}{2\lambda} a^{ij}, \; i = 1, 2, \ldots, p$$

$i = j$

$$\frac{1}{2\lambda} a^{jj} = 2\lambda d^2 \quad \Longrightarrow \quad 2\lambda = \sqrt{\frac{a^{jj}}{d^2}}$$

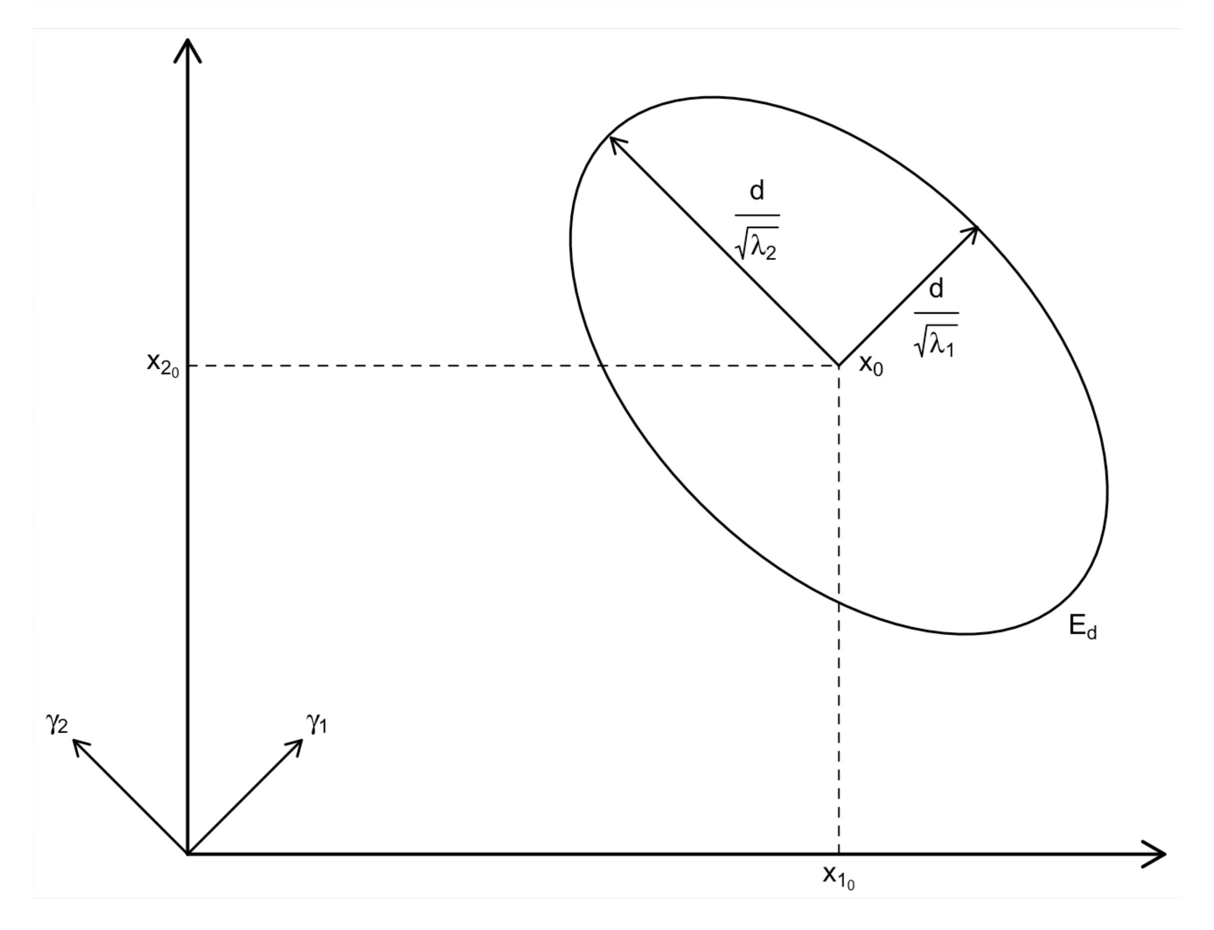
- 椭球面与包围它的矩形在第 j 个轴的正方向上的切点坐标为

$$x_i = \sqrt{\frac{d^2}{a^{jj}}}\, a^{ij}, \quad i = 1, 2, \ldots, p$$

$i = j$ 定理 2.7 (*iii*)

几何直观 (geometrical aspects)

- 定理 2.7 的作用
  - 首先，它为绘制二维椭圆提供了一个有用的工具.
  - 可以证明，多元正态总体均值向量 μ 的置信域可由一个特定的椭球体给出，椭球的参数取决于样本的特征.
  - 还将证明，多元正态分布密度的等值面亦由椭球体给出，椭球体的参数取决于多元正态分布的均值向量和协方差矩阵.



几何直观 (geometrical aspects)

- 向量的范数 (norm)

  - 给定 $\mathbb{R}^p$ 中的一个向量 $\boldsymbol{x}$，定义 $\boldsymbol{x}$ (关于度量 $\mathscr{I}_p$) 的范数或长度 (length) 为

$$\|\boldsymbol{x}\| = d\left(\boldsymbol{0}_p, \boldsymbol{x}\right) = \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$$

  - 当 $\|\boldsymbol{x}\| = 1$ 时，称 $\boldsymbol{x}$ 为单位向量 (unit vector).

  - 更一般地，$\boldsymbol{x}$ (关于度量 $\mathscr{A}$) 的范数定义为

$$\|\boldsymbol{x}\|_{\mathscr{A}} = \sqrt{\boldsymbol{x}^{\mathrm{T}}\mathscr{A}\boldsymbol{x}}$$

几何直观 (geometrical aspects)

- 两个向量的夹角
  - 考虑两个向量 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^p$，$\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的夹角 θ 可由其余弦如下定义：

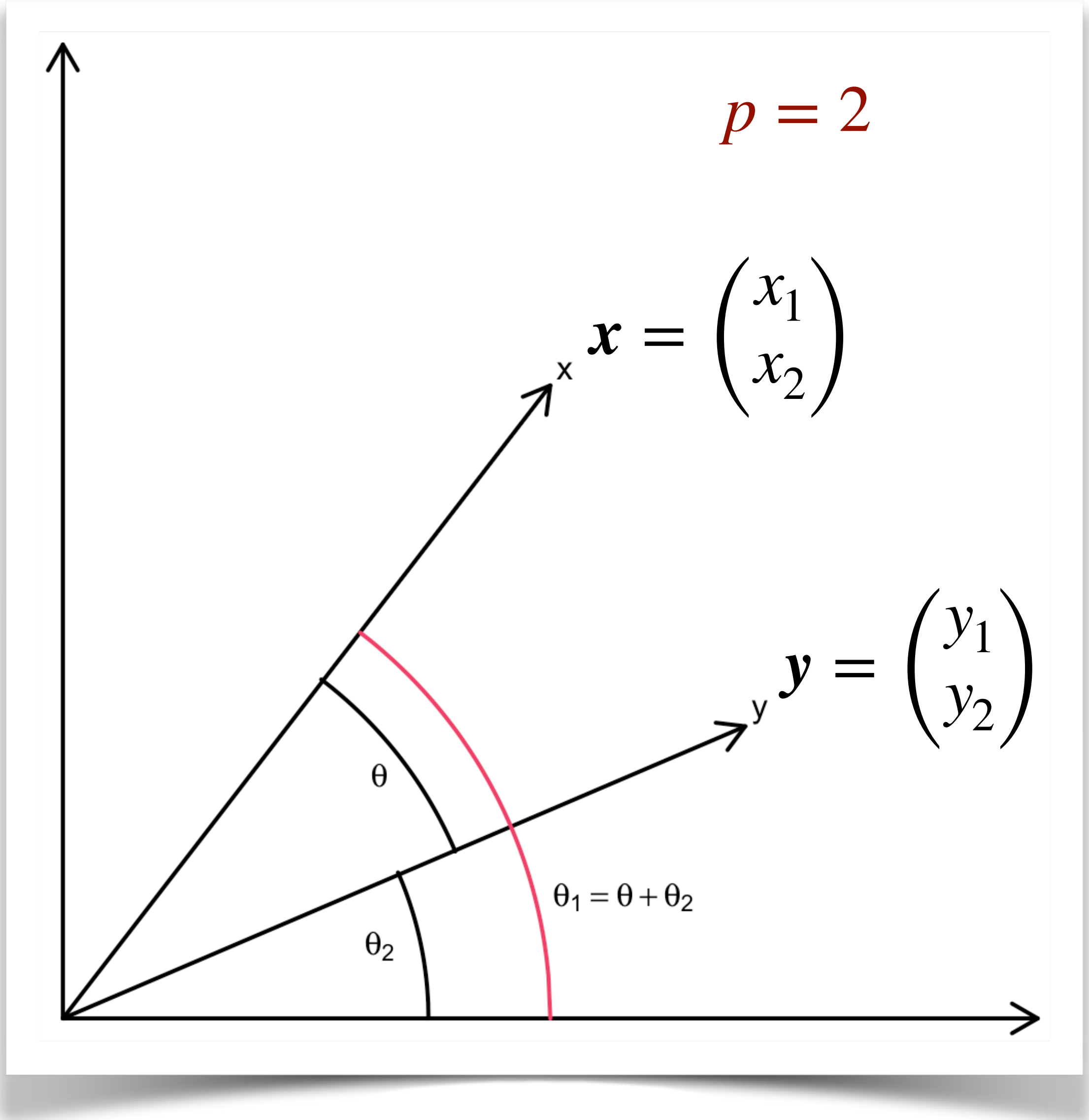
$$\cos\theta = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}{\|\boldsymbol{x}\|\ \|\boldsymbol{y}\|}$$

$$\begin{cases} \|\boldsymbol{x}\|\cos\theta_1 = x_1 \\ \|\boldsymbol{x}\|\sin\theta_1 = x_2 \end{cases} \qquad \begin{cases} \|\boldsymbol{y}\|\cos\theta_2 = y_1 \\ \|\boldsymbol{y}\|\sin\theta_2 = y_2 \end{cases}$$

$$\Longrightarrow \quad \cos\theta = \cos(\theta_1 - \theta_2)$$

$$= \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2$$

$$= \frac{x_1}{\|\boldsymbol{x}\|}\frac{y_1}{\|\boldsymbol{y}\|} + \frac{x_2}{\|\boldsymbol{x}\|}\frac{y_2}{\|\boldsymbol{y}\|} = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}{\|\boldsymbol{x}\|\ \|\boldsymbol{y}\|}$$



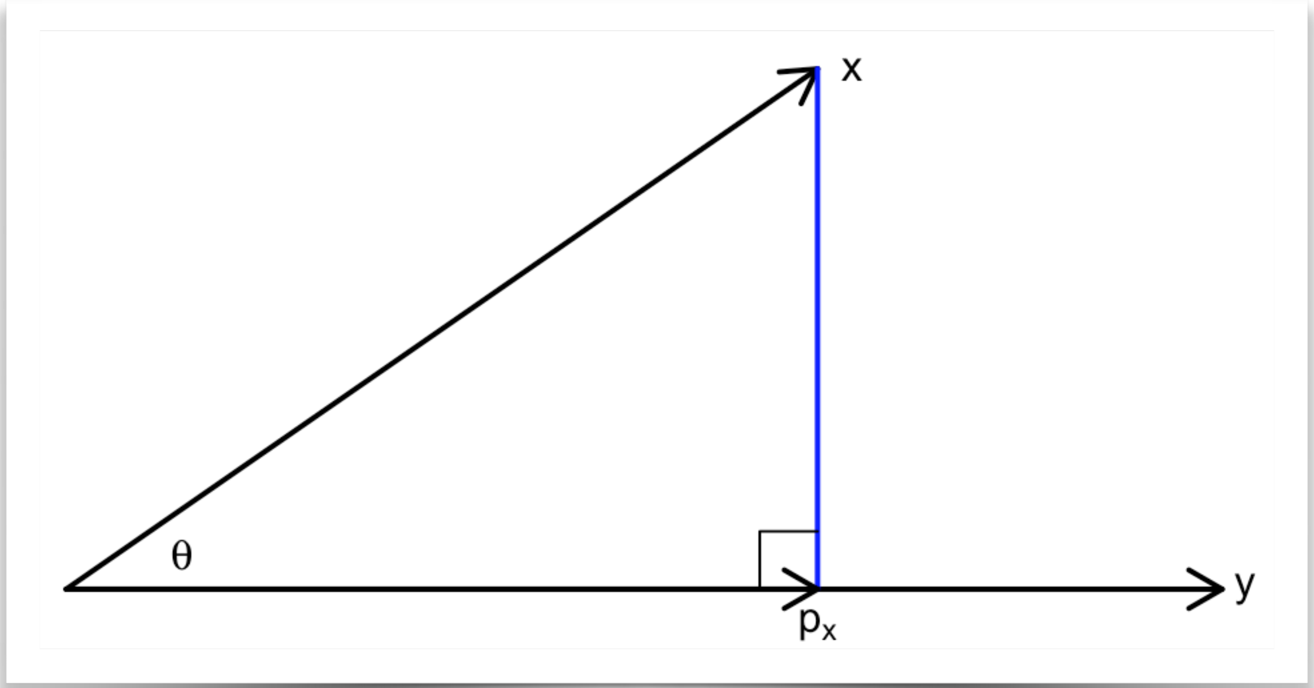
几何直观 (geometrical aspects)

- 注 2.1 如果 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=0$，则夹角 $\theta=\dfrac{\pi}{2}$.

$$\cos\theta=\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|}$$

$$\|\boldsymbol{p}_x\| = \|\boldsymbol{x}\|\cdot|\cos\theta| = \frac{\left|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}\right|}{\|\boldsymbol{y}\|}$$

$\boldsymbol{x}$ 在 $\boldsymbol{y}$ 上的投影：$\boldsymbol{p}_x=\dfrac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}\boldsymbol{y}$



- 关于一般度量 $\mathscr{A}$ 也可以定义向量的夹角

$$\cos\theta=\frac{\boldsymbol{x}^{\mathrm{T}}\mathscr{A}\boldsymbol{y}}{\|\boldsymbol{x}\|_{\mathscr{A}}\cdot\|\boldsymbol{y}\|_{\mathscr{A}}}$$

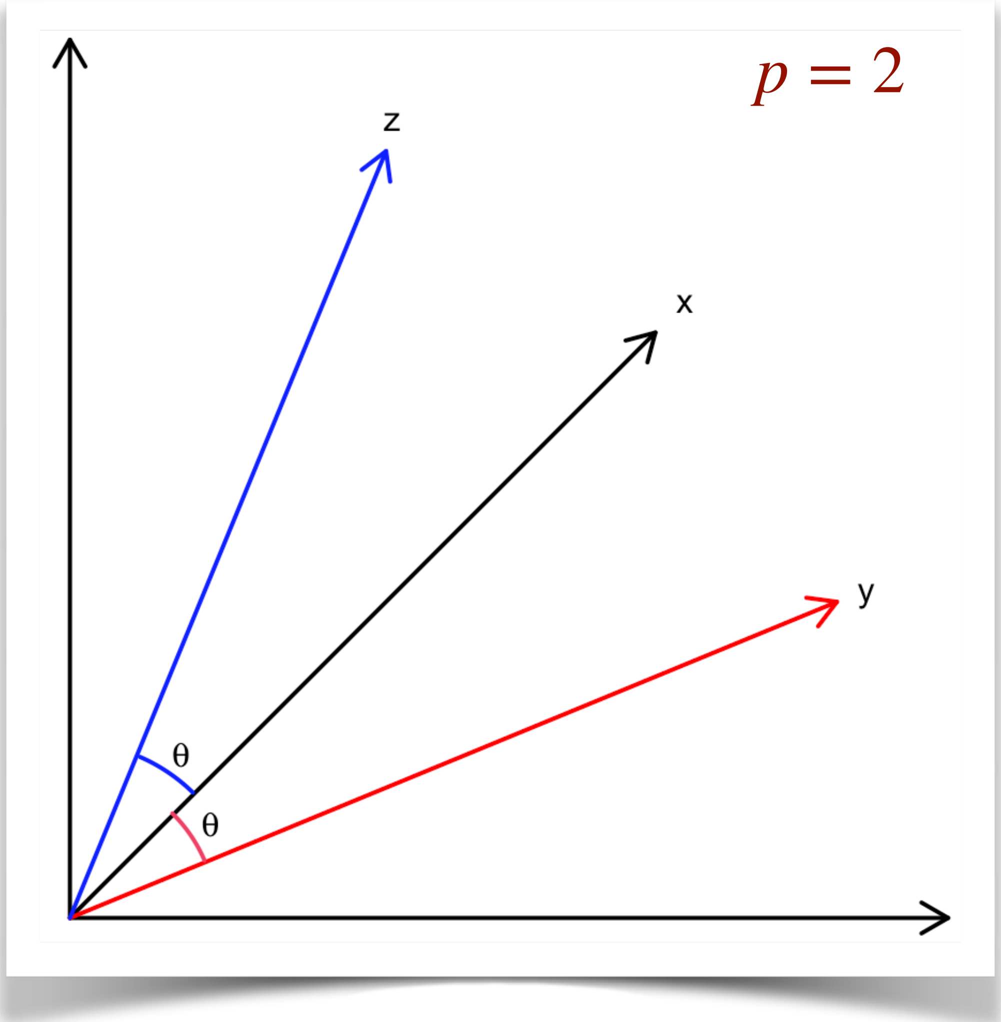
- 如果 $\cos\theta=0$，则称 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 关于度量 $\mathscr{A}$ 正交.

几何直观 (geometrical aspects)

- 旋转 (rotation)

正交矩阵：$\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}} = \boldsymbol{I}$

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \boldsymbol{\Gamma} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$



- 关于原点顺时针旋转一个角度 θ

$$\boldsymbol{y} = \boldsymbol{\Gamma}\,\boldsymbol{x}$$

- 关于原点逆时针旋转一个角度 θ

$$\boldsymbol{z} = \boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{x}$$

- 一般地，向量 $\boldsymbol{x}$ 左乘一个正交矩阵 $\boldsymbol{\Gamma}$，在几何上对应于坐标轴的旋转，从而第一个新轴由 $\boldsymbol{\Gamma}$ 的第一行确定.

几何直观 (geometrical aspects)

- 矩阵的列空间 (Column Space) 与零空间 (Null Space)

  - 对矩阵 $\mathscr{X}_{n\times p}$，定义 $\mathscr{X}$ 的列空间，或 $\mathscr{X}$ 的列向量生成的空间为

$$C\left(\mathscr{X}\right) \triangleq \left\{\boldsymbol{x} \in \mathbb{R}^n \,\middle|\, \text{存在向量}\, \boldsymbol{a} \in \mathbb{R}^p \,\text{使得}\, \mathscr{X}\boldsymbol{a} = \boldsymbol{x}\right\}$$

$$\Longrightarrow \quad C\left(\mathscr{X}\right) \subseteq \mathbb{R}^n, \quad \dim\left\{C\left(\mathscr{X}\right)\right\} = \text{rank}\left(\mathscr{X}\right) = r \leq \min\left\{n, p\right\}$$

  - $\mathscr{X}$ 的零空间定义为

$$N\left(\mathscr{X}\right) \triangleq \left\{\boldsymbol{y} \in \mathbb{R}^p \,\middle|\, \mathscr{X}\boldsymbol{y} = \boldsymbol{0}\right\}$$

$$\Longrightarrow \quad N\left(\mathscr{X}\right) \subseteq \mathbb{R}^p, \quad \dim\left\{N\left(\mathscr{X}\right)\right\} = p - r$$

几何直观 (geometrical aspects)

- 注 2.2 $N\left(\mathscr{X}^{\mathrm{T}}\right)$ 是 $\mathbb{R}^n$ 中 $C\left(\mathscr{X}\right)$ 的正交补 (orthogonal complement)，即，给定向量 $\boldsymbol{b} \in \mathbb{R}^n$，则 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b} = 0$ 对所有 $\boldsymbol{x} \in C\left(\mathscr{X}\right)$ 都成立的充分必要条件是 $\boldsymbol{b} \in N\left(\mathscr{X}^{\mathrm{T}}\right)$.

- 设 $\mathscr{X} = \begin{pmatrix} 2 & 3 & 5 \\ 4 & 6 & 7 \\ 6 & 8 & 6 \\ 8 & 2 & 4 \end{pmatrix}$. $\Longrightarrow$ $\operatorname{rank}\left(\mathscr{X}\right) = 3.$ $\Longrightarrow$ $\dim\left\{C\left(\mathscr{X}\right)\right\} = 3.$

  $\Longrightarrow$ $N\left(\mathscr{X}\right) = \left\{\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\right\}.$ $\Longrightarrow$ $\dim\left\{N\left(\mathscr{X}\right)\right\} = \operatorname{rank}\left(\mathscr{X}\right) - 3 = 0$

- 设 $\mathscr{X} = \begin{pmatrix} 2 & 3 & 1 \\ 4 & 6 & 2 \\ 6 & 8 & 3 \\ 8 & 2 & 4 \end{pmatrix}$. $\Longrightarrow$ $\operatorname{rank}\left(\mathscr{X}\right) = 2.$ $\Longrightarrow$ $\dim\left\{C\left(\mathscr{X}\right)\right\} = 2.$

  $\Longrightarrow$ $\dim\left\{N\left(\mathscr{X}\right)\right\} = 3 - \operatorname{rank}\left(\mathscr{X}\right) = 1$

几何直观 (geometrical aspects)

- **投影矩阵** (projection matrix)
  - $\mathbb{R}^n$ 中的一个矩阵 $\mathscr{P}_{n\times n}$ 称为**投影矩阵**，当且仅当
    $$\mathscr{P} = \mathscr{P}^{\mathrm{T}} = \mathscr{P}^2 \quad (\mathscr{P}\text{ 是对称幂等矩阵})$$
  - 设 $\boldsymbol{b} \in \mathbb{R}^n$，则 $\boldsymbol{b}$ 在 $C(\mathscr{P})$ 上的投影为 $\boldsymbol{a} = \mathscr{P}\boldsymbol{b}$.

几何直观 (geometrical aspects)

- 在 $C(\mathscr{X})$ 上的投影
  - 考虑矩阵 $\mathscr{X}_{n\times p}$，记

$\mathscr{P}$ 是对称矩阵：$\mathscr{P}^{\mathrm{T}} = \mathscr{P}$

$\mathscr{Q}$ 也是对称矩阵：$\mathscr{Q}^{\mathrm{T}} = \mathscr{Q}$

$$\mathscr{P} = \mathscr{X}\left(\mathscr{X}^{\mathrm{T}}\mathscr{X}\right)^{-1}\mathscr{X}^{\mathrm{T}}, \quad \mathscr{Q} = \mathscr{I}_n - \mathscr{P}$$

$$\Longrightarrow \quad \mathscr{P}^2 = \left[\mathscr{X}\left(\mathscr{X}^{\mathrm{T}}\mathscr{X}\right)^{-1}\mathscr{X}^{\mathrm{T}}\right]\mathscr{X}\left(\mathscr{X}^{\mathrm{T}}\mathscr{X}\right)^{-1}\mathscr{X}^{\mathrm{T}}$$

$$= \mathscr{X}\left(\mathscr{X}^{\mathrm{T}}\mathscr{X}\right)^{-1}\mathscr{X}^{\mathrm{T}} = \mathscr{P} \quad \Longrightarrow \mathscr{P} \text{ 是幂等阵}$$

$$\Longrightarrow \quad \mathscr{Q}^2 = \left(\mathscr{I}_n - \mathscr{P}\right)\left(\mathscr{I}_n - \mathscr{P}\right)$$

$$= \mathscr{I}_n - \mathscr{P} - \mathscr{P} + \mathscr{P}^2 = \mathscr{I}_n - \mathscr{P} = \mathscr{Q} \quad \Longrightarrow \mathscr{Q} \text{ 也是幂等阵}$$

几何直观 (geometrical aspects)

- 在 $C(\mathscr{X})$ 上的投影
  - 考虑矩阵 $\mathscr{X}_{n\times p}$，记

$$\mathscr{P} = \mathscr{X}\left(\mathscr{X}^{\mathrm{T}}\mathscr{X}\right)^{-1}\mathscr{X}^{\mathrm{T}}, \quad \mathscr{Q} = \mathscr{I}_n - \mathscr{P}$$

$$\Longrightarrow \quad \mathscr{P}\mathscr{X} = \mathscr{X}\left(\mathscr{X}^{\mathrm{T}}\mathscr{X}\right)^{-1}\mathscr{X}^{\mathrm{T}}\cdot\mathscr{X} = \mathscr{X}$$

  - 由于 $\mathscr{X}$ 的列向量到自身的投影就是它自己，所以对任何向量 $\boldsymbol{b}\in\mathbb{R}^n$，投影矩阵 $\mathscr{P}$ 将它投影到了 $C(\mathscr{X})$ 上.

$$\Longrightarrow \quad \mathscr{Q}\mathscr{X} = \left(\mathscr{I}_n - \mathscr{P}\right)\mathscr{X} = \mathscr{X} - \mathscr{P}\mathscr{X} = 0$$

  - 类似地，投影矩阵 $\mathscr{Q}$ 将 $\mathbb{R}^n$ 中的任一向量 $\boldsymbol{b}\in\mathbb{R}^n$ 投影到了 $C(\mathscr{X})$ 的正交补上.

几何直观 (geometrical aspects)

- 定理 2.8 设 $\mathscr{P} = \mathscr{X}\left(\mathscr{X}^{\mathrm{T}}\mathscr{X}\right)^{-1}\mathscr{X}^{\mathrm{T}}$ 为投影矩阵，$\mathscr{Q} = \mathscr{I}_n - \mathscr{P}$ 是其正交补，则

  (i) $\boldsymbol{x} = \mathscr{P}\boldsymbol{b} \implies \boldsymbol{x} \in C\left(\mathscr{X}\right).$

  (ii) $\boldsymbol{y} = \mathscr{Q}\boldsymbol{b} \implies \forall \boldsymbol{x} \in C\left(\mathscr{X}\right), \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} = 0.$

证明 (i)

$$\boldsymbol{x} = \mathscr{P}\boldsymbol{b} = \mathscr{X}\boxed{\left(\mathscr{X}^{\mathrm{T}}\mathscr{X}\right)^{-1}\mathscr{X}^{\mathrm{T}}\boldsymbol{b}}$$

$$\boldsymbol{a} = \left(\mathscr{X}^{\mathrm{T}}\mathscr{X}\right)^{-1}\mathscr{X}^{\mathrm{T}}\boldsymbol{b} \in \mathbb{R}^p$$

$$= \mathscr{X}\boldsymbol{a}$$

$$= \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_p \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}$$

$$= a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_p\boldsymbol{x}_p \in C\left(\mathscr{X}\right)$$

几何直观 (geometrical aspects)

- **定理 2.8** 设 $\mathscr{P} = \mathscr{X}\left(\mathscr{X}^{\mathrm{T}}\mathscr{X}\right)^{-1}\mathscr{X}^{\mathrm{T}}$ 为投影矩阵，$\mathscr{Q} = \mathscr{I}_n - \mathscr{P}$ 是其正交补，则

  (i) $\boldsymbol{x} = \mathscr{P}\boldsymbol{b} \quad \Longrightarrow \quad \boldsymbol{x} \in C\left(\mathscr{X}\right).$

  (ii) $\boldsymbol{y} = \mathscr{Q}\boldsymbol{b} \quad \Longrightarrow \quad \forall \boldsymbol{x} \in C\left(\mathscr{X}\right),\ \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} = 0.$

证明 *(ii)*

$$\begin{aligned}
\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} &= \left(\mathscr{Q}\boldsymbol{b}\right)^{\mathrm{T}}\mathscr{X}\boldsymbol{a} = \boldsymbol{b}^{\mathrm{T}}\left(\mathscr{I}_n - \mathscr{P}\right)^{\mathrm{T}}\mathscr{X}\boldsymbol{a} \\
&= \boldsymbol{b}^{\mathrm{T}}\left(\mathscr{I}_n - \mathscr{P}\right)\mathscr{X}\boldsymbol{a} \\
&= \boldsymbol{b}^{\mathrm{T}}\mathscr{X}\boldsymbol{a} - \boldsymbol{b}^{\mathrm{T}}\mathscr{X}\underbrace{\left(\mathscr{X}^{\mathrm{T}}\mathscr{X}\right)^{-1}\mathscr{X}^{\mathrm{T}}\mathscr{X}}_{\mathscr{I}_p}\boldsymbol{a} \\
&= \boldsymbol{b}^{\mathrm{T}}\mathscr{X}\boldsymbol{a} - \boldsymbol{b}^{\mathrm{T}}\mathscr{X}\boldsymbol{a} = 0
\end{aligned}$$

几何直观 (geometrical aspects)

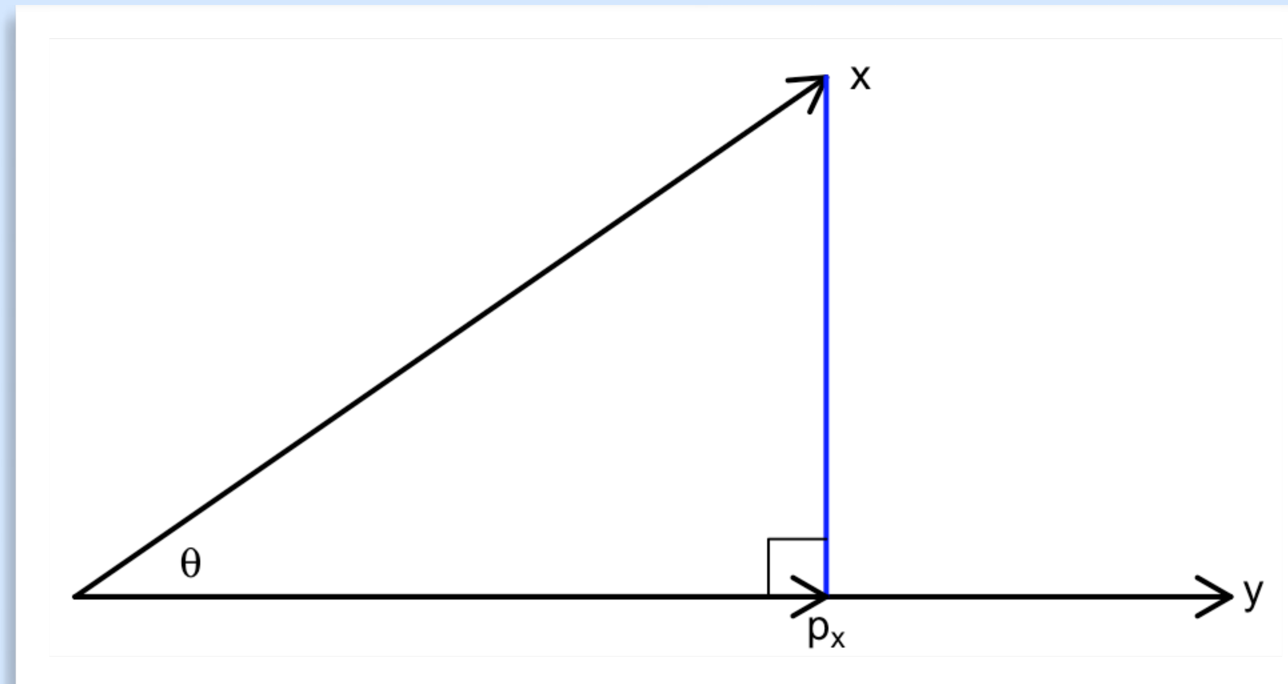
- 注 2.3 设 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$，$\boldsymbol{p}_x \in \mathbb{R}^n$ 表示 $\boldsymbol{x}$ 在 $\boldsymbol{y}$ 上的投影向量.

取 $\mathscr{X} = \boldsymbol{y}$ 时，我们有

$$\boldsymbol{p}_x = \overbrace{\boxed{\boldsymbol{y}\left(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}\right)^{-1}\boldsymbol{y}^{\mathrm{T}}}}^{\text{投影矩阵}}\boldsymbol{x}$$

$$= \frac{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}}{\|\boldsymbol{y}\|^2}\boldsymbol{y}$$

$$\Longrightarrow \quad \|\boldsymbol{p}_x\| = \sqrt{\boldsymbol{p}_x^{\mathrm{T}}\boldsymbol{p}_x} = \frac{\left|\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}\right|}{\|\boldsymbol{y}\|}$$



向量 $\boldsymbol{x}$ 在 $\boldsymbol{y}$ 上的投影：$\boldsymbol{p}_x = \dfrac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}\boldsymbol{y}$