《矩阵论》习题解答

民间版

作者：杨志豪、徐若彭、张卓雨
组织：浙江大学信息与电子工程学院
时间：November, 2023
版本：1.0

“考前勤奋，考后乐观。”

目录

第 1 次　作业 1

第 2 次　作业 8

第 3 次　作业 13

第 4 次　作业 16

第 5 次　作业 22

第一次 作业

1.8 利用初等行变换求解线性方程组

$$
\begin{aligned}
2x_1-4x_2+3x_3-4x_4-11x_5&=28\\
-x_1+2x_2-x_3+2x_4+5x_5&=-13\\
-3x_3+2x_4+5x_5&=-10\\
3x_1-5x_2+10x_3-7x_4+12x_5&=31
\end{aligned}
$$

解 根据原方程得增广矩阵

$$
\begin{bmatrix}
2 & -4 & 3 & -4 & -11 & 28\\
-1 & 2 & -1 & 2 & 5 & -13\\
0 & 0 & -3 & 2 & 5 & -10\\
3 & -5 & 10 & -7 & 12 & 31
\end{bmatrix}
$$

利用初等行变换进行化简

(a) $\boldsymbol{r}_1+2\boldsymbol{r}_2\to\boldsymbol{r}_1,\boldsymbol{r}_4+3\boldsymbol{r}_2\to\boldsymbol{r}_4$

$$
\begin{bmatrix}
0 & 0 & 1 & 0 & -1 & 2\\
-1 & 2 & -1 & 2 & 5 & -13\\
0 & 0 & -3 & 2 & 5 & -10\\
0 & 1 & 7 & -1 & 27 & -8
\end{bmatrix}
$$

(b) $\boldsymbol{r}_2-2\boldsymbol{r}_4\to\boldsymbol{r}_2$

$$
\begin{bmatrix}
0 & 0 & 1 & 0 & -1 & 2\\
-1 & 0 & -15 & 4 & -49 & 3\\
0 & 0 & -3 & 2 & 5 & -10\\
0 & 1 & 7 & -1 & 27 & -8
\end{bmatrix}
$$

(c) $\boldsymbol{r}_2+15\boldsymbol{r}_1\to\boldsymbol{r}_2,\boldsymbol{r}_3+3\boldsymbol{r}_1\to\boldsymbol{r}_3,\boldsymbol{r}_4-7\boldsymbol{r}_1\to\boldsymbol{r}_4$

$$
\begin{bmatrix}
0 & 0 & 1 & 0 & -1 & 2\\
-1 & 0 & 0 & 4 & -64 & 33\\
0 & 0 & 0 & 2 & 2 & -4\\
0 & 1 & 0 & -1 & 34 & -22
\end{bmatrix}
$$

(d) $\dfrac{1}{2}\times\boldsymbol{r}_3\to\boldsymbol{r}_3$

$$
\begin{bmatrix}
0 & 0 & 1 & 0 & -1 & 2\\
-1 & 0 & 0 & 4 & -64 & 33\\
0 & 0 & 0 & 1 & 1 & -2\\
0 & 1 & 0 & -1 & 34 & -22
\end{bmatrix}
$$

(e) $-\boldsymbol{r}_2\to\boldsymbol{r}_2,\boldsymbol{r}_2+4\boldsymbol{r}_3\to\boldsymbol{r}_2,\boldsymbol{r}_4+\boldsymbol{r}_3\to\boldsymbol{r}_4$

$$
\begin{bmatrix}
0 & 0 & 1 & 0 & -1 & 2\\
1 & 0 & 0 & 0 & 68 & -41\\
0 & 0 & 0 & 1 & 1 & -2\\
0 & 1 & 0 & 0 & 35 & -24
\end{bmatrix}
$$

(f) $\boldsymbol{r}_1 \leftrightarrow \boldsymbol{r}_2, \boldsymbol{r}_2 \leftrightarrow \boldsymbol{r}_4, \boldsymbol{r}_3 \leftrightarrow \boldsymbol{r}_4$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 68 & -41 \\ 0 & 1 & 0 & 0 & 35 & -24 \\ 0 & 0 & 1 & 0 & -1 & 2 \\ 0 & 0 & 0 & 1 & 1 & -2 \end{bmatrix}$$

方程组有特解 $[27, 11, 1, -1, -1]^T$，对应齐次方程组通解为 $[68, 35, -1, 1, -1]^T$。方程组的解为

$$\boldsymbol{x} = k \begin{bmatrix} 68 \\ 35 \\ -1 \\ 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 27 \\ 11 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \quad \forall k \in \mathbb{R}$$

1.12 证明自协方差矩阵和互协方差的下列性质：

(1) $\mathbf{Var}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}) = \boldsymbol{A}\mathbf{Var}(\boldsymbol{x})\boldsymbol{A}^{\mathbf{H}}$，其中 $\boldsymbol{A}$ 和 $\boldsymbol{b}$ 分别为常数矩阵和常数向量。

(2) $\mathbf{Cov}(\boldsymbol{x}, \boldsymbol{y}) = [\mathbf{Cov}(\boldsymbol{y}, \boldsymbol{x})]^{\mathbf{H}}$。

(3) $\mathbf{Cov}(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{B}\boldsymbol{y}) = \boldsymbol{A}\mathbf{Cov}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{B}^{\mathbf{H}}$。

证明

(1)

$$\begin{aligned} \mathbf{E}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}) &= \boldsymbol{A}\mathbf{E}(\boldsymbol{x}) + \boldsymbol{b} \triangleq \boldsymbol{\mu}_x \\ \mathbf{Var}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}) &= \mathbf{E}\left\{[\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} - \boldsymbol{\mu}_x][\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} - \boldsymbol{\mu}_x]^{\mathbf{H}}\right\} \\ &= \mathbf{E}\left\{\left[\boldsymbol{A}[\boldsymbol{x} - \mathbf{E}(\boldsymbol{x})]\right]\left[\boldsymbol{A}[\boldsymbol{x} - \mathbf{E}(\boldsymbol{x})]\right]^{\mathbf{H}}\right\} \\ &= \boldsymbol{A}\mathbf{E}\left\{[\boldsymbol{x} - \mathbf{E}(\boldsymbol{x})][\boldsymbol{x} - \mathbf{E}(\boldsymbol{x})]^{\mathbf{H}}\right\}\boldsymbol{A}^{\mathbf{H}} \\ &= \boldsymbol{A}\mathbf{Var}(\boldsymbol{x})\boldsymbol{A}^{\mathbf{H}} \end{aligned}$$

(2)

$$\begin{aligned} \mathbf{Cov}(\boldsymbol{x}, \boldsymbol{y}) &= \mathbf{E}\left\{[\boldsymbol{x} - \mathbf{E}(\boldsymbol{x})][\boldsymbol{y} - \mathbf{E}(\boldsymbol{y})]^{\mathbf{H}}\right\} \\ &= \left\{\mathbf{E}\{[\boldsymbol{x} - \mathbf{E}(\boldsymbol{x})][\boldsymbol{y} - \mathbf{E}(\boldsymbol{y})]^{\mathbf{H}}\}^{\mathbf{H}}\right\}^{\mathbf{H}} \\ &= \left\{\mathbf{E}\{[\boldsymbol{y} - \mathbf{E}(\boldsymbol{y})][\boldsymbol{x} - \mathbf{E}(\boldsymbol{x})]^{\mathbf{H}}\}\right\}^{\mathbf{H}} \\ &= [\mathbf{Cov}(\boldsymbol{y}, \boldsymbol{x})]^{\mathbf{H}} \end{aligned}$$

(3)

$$\begin{aligned} \mathbf{Cov}(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{B}\boldsymbol{y}) &= \mathbf{E}\left\{[\boldsymbol{A}\boldsymbol{x} - \mathbf{E}(\boldsymbol{A}\boldsymbol{x})][\boldsymbol{B}\boldsymbol{y} - \mathbf{E}(\boldsymbol{B}\boldsymbol{y})]^{\mathbf{H}}\right\} \\ &= \mathbf{E}\left\{\boldsymbol{A}[\boldsymbol{x} - \mathbf{E}(\boldsymbol{x})][\boldsymbol{y} - \mathbf{E}(\boldsymbol{y})]^{\mathbf{H}}\boldsymbol{B}^{\mathbf{H}}\right\} \\ &= \boldsymbol{A}\mathbf{E}\left\{[(\boldsymbol{x} - \mathbf{E}(\boldsymbol{x})][\boldsymbol{y} - \mathbf{E}(\boldsymbol{y})]^{\mathbf{H}}\right\}\boldsymbol{B}^{\mathbf{H}} \\ &= \boldsymbol{A}\mathbf{Cov}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{B}^{\mathbf{H}} \end{aligned}$$

1.17 矩阵的秩在工程控制系统的设计中起着重要的作用。一个离散时间的控制系统的状态空间模型包括了差分方程

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k, \qquad k = 0, 1, \cdots$$

式中，$\boldsymbol{A} \in \mathbb{R}^{n\times n}, \boldsymbol{B} \in \mathbb{R}^{n\times m}$，并且 $\boldsymbol{x}_k \in \mathbb{R}^n$ 为描述系统 k 时刻状态的变量，简称状态向量；而 $\boldsymbol{u}_k$ 为系统在 k 时刻的输入或控制向量。矩阵对 $(\boldsymbol{A}, \boldsymbol{B})$ 称为可控的，若

$$\mathrm{rank}([\boldsymbol{B}, \boldsymbol{A}\boldsymbol{B}, \boldsymbol{A}^2\boldsymbol{B}, \cdots, \boldsymbol{A}^{n-1}\boldsymbol{B}]) = n$$

若 $(\boldsymbol{A}, \boldsymbol{B})$ 是可控的，则最多用 n 步即可将系统控制到任意一个指定的状态 $\boldsymbol{x}$，试确定以下矩阵对是否可控：

(1) $\boldsymbol{A} = \begin{bmatrix} 0.9 & 1 & 0 \\ 0 & -0.9 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

(2) $\boldsymbol{A} = \begin{bmatrix} 0.8 & -0.3 & 0 \\ 0.2 & 0.5 & 1 \\ 0 & 0 & -0.5 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$

解 矩阵为三阶矩阵，所以 $n = 3$

(1)

$$[\boldsymbol{B}, \boldsymbol{A}\boldsymbol{B}, \boldsymbol{A}^2\boldsymbol{B}] = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -0.9 & 0.81 \\ 1 & 0.5 & 0.25 \end{bmatrix}$$

其秩为 3，所以矩阵对 $(\boldsymbol{A}, \boldsymbol{B})$ 是可控的。

(2)

$$[\boldsymbol{B}, \boldsymbol{A}\boldsymbol{B}, \boldsymbol{A}^2\boldsymbol{B}] = \begin{bmatrix} 1 & 0.5 & 0.19 \\ 1 & 0.7 & 0.45 \\ 0 & 0 & 0 \end{bmatrix}$$

其秩为 2，所以矩阵对 $(\boldsymbol{A}, \boldsymbol{B})$ 是不可控的。

1.19 令正方矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 有同样的维数，证明 $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{B}\boldsymbol{A})$

证明 设矩阵维数为 n，有

$$(\boldsymbol{A}\boldsymbol{B})_{ii} = \sum_{j=1}^{n} a_{ij} b_{ji}$$

$$\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \sum_{i=1}^{n} (\boldsymbol{A}\boldsymbol{B})_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} b_{ji} = \sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} a_{ji}$$

$$\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{B}\boldsymbol{A})$$

1.37 当 α 取何值时，线性方程组

$$(\alpha+3)x_1 + x_2 + 2x_3 = \alpha$$
$$3(\alpha+1)x_1 + \alpha x_2 + (\alpha+3)x_3 = 3$$
$$\alpha x_1 + (\alpha-1)x_2 + x_3 = \alpha$$

有唯一解、无解和无穷多解。当方程组有无穷多解时，求出它的通解。

解 增广矩阵

$$\begin{bmatrix} \alpha+3 & 1 & 2 & \alpha \\ 3(\alpha+1) & \alpha & \alpha+3 & 3 \\ \alpha & \alpha-1 & 1 & \alpha \end{bmatrix} \xRightarrow{\text{初等行变换}} \begin{bmatrix} 1 & \dfrac{2-\alpha}{3} & \dfrac{1}{3} & 0 \\ 0 & \alpha-(2-\alpha)(\alpha+1) & 2 & 3 \\ 0 & \alpha-1-\dfrac{\alpha(2-\alpha)}{3} & 1-\dfrac{\alpha}{3} & \alpha \end{bmatrix}$$

(1) 方程组有唯一解，要求系数矩阵可逆，行列式 $\neq 0$

$$[\alpha-(2-\alpha)(\alpha+1)](1-\frac{\alpha}{3})-2[\alpha-1-\frac{\alpha(2-\alpha)}{3}]=\frac{1}{3}(\alpha^2-\alpha^3)\neq 0$$

则 $\alpha\neq 0$ 且 $\alpha\neq 1$ 时，方程组有唯一解。

(2) 方程组无解，要求系数矩阵不可逆，行列式 $=0$

$$[\alpha-(2-\alpha)(\alpha+1)](1-\frac{\alpha}{3})-2[\alpha-1-\frac{\alpha(2-\alpha)}{3}]=\frac{1}{3}(\alpha^2-\alpha^3)=0$$

要求方程为非一致性方程，系数矩阵后两行比例与向量 $\begin{bmatrix}0 & 3 & \alpha\end{bmatrix}^T$ 后两个元素所成比例不同，则 $\alpha=0$

(3) 方程组有无穷多解，要求系数矩阵不可逆，行列式 $=0$

$$[\alpha-(2-\alpha)(\alpha+1)](1-\frac{\alpha}{3})-2[\alpha-1-\frac{\alpha(2-\alpha)}{3}]=\frac{1}{3}(\alpha^2-\alpha^3)=0$$

且方程为一致性方程，则 $\alpha=1$。此时方程的通解为 $x_2=2x_3-3$, $x_1=1-x_3$。

1.42 已知向量组

$$\boldsymbol{a}_1=\begin{bmatrix}a\\2\\10\end{bmatrix},\quad \boldsymbol{a}_2=\begin{bmatrix}-2\\1\\5\end{bmatrix},\quad \boldsymbol{a}_3=\begin{bmatrix}-1\\1\\4\end{bmatrix},\quad \boldsymbol{b}=\begin{bmatrix}1\\b\\c\end{bmatrix}$$

试分别求出满足以下条件的 a,b,c 值:

(1) $\boldsymbol{b}$ 可由 $\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3$ 线性表示，且唯一。

(2) $\boldsymbol{b}$ 不能由 $\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3$ 线性表示。

(3) $\boldsymbol{b}$ 可由 $\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3$ 线性表示，但表示不唯一。并求出一般表达式。

解 令 $\boldsymbol{A}=[\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3]$

(1) 该问题等价于求 a,b,c 为何值时，$\boldsymbol{Ax}=\boldsymbol{b}$ 有唯一解

$$\text{增广矩阵:}\begin{bmatrix} a & -2 & -1 & 1 \\ 2 & 1 & 1 & b \\ 10 & 5 & 4 & c \end{bmatrix} \xRightarrow{\text{初等行变换}} \begin{bmatrix} a & -2 & -1 & 1 \\ 2+a & -1 & 0 & b+1 \\ 4+a & 0 & 0 & c-3b+1 \end{bmatrix}$$

$\boldsymbol{Ax}=\boldsymbol{b}$ 有唯一解，系数矩阵 $\boldsymbol{A}$ 与增广矩阵秩相同，且等于未知数个数。则 $a\neq -4$ 时，方程组有唯一解

$$\boldsymbol{x}=\left[\frac{c-3b+1}{4+a},\frac{2c+ac-10b-4ab-2}{4+a},\frac{-ac+5ab-4c+20b}{4+a}\right]^T$$

(2) 当 $a=-4$，且 $c\neq 3b-1$ 时，方程组无解

(3) 当 $a=-4$，且 $c=3b-1$ 时，方程组无穷多解, 设 x_1 为自由量，有

$$
\begin{aligned}
x_1 &= x_1 \\
x_2 &= -2x_1-(b+1) \\
x_3 &= 2b+1
\end{aligned}
$$

基础解系 $\boldsymbol{\xi}=[1,-2,0]^T$，特解 $\boldsymbol{\xi}_0=[0,-b-1,2b+1]^T$，方程组有无穷多解 $\boldsymbol{x}=k\boldsymbol{\xi}+\boldsymbol{\xi}_0, \forall k\in\mathbb{R}$

1.73 已知 $\boldsymbol{AB}=\boldsymbol{BA}=\boldsymbol{O}$（零矩阵）和 $\operatorname{rank}(\boldsymbol{A}^2)=\operatorname{rank}(\boldsymbol{A})$，证明

(1) $\operatorname{rank}(\boldsymbol{A}+\boldsymbol{B})=\operatorname{rank}(\boldsymbol{A})+\operatorname{rank}(\boldsymbol{B})$。

(2) $\operatorname{rank}(\boldsymbol{A}^k+\boldsymbol{B}^k)=\operatorname{rank}(\boldsymbol{A}^k)+\operatorname{rank}(\boldsymbol{B}^k)$，其中 k 是某个整数。

证明

(1) 首先对矩阵 $\boldsymbol{A}$ 进行 Jordan 标准型分解

$$
\boldsymbol{A}=\boldsymbol{S}\begin{bmatrix} \boldsymbol{J}_{n_1}(\lambda_1) & & & 0 \\ & \boldsymbol{J}_{n_2}(\lambda_2) & & \\ & & \ddots & \\ 0 & & & \boldsymbol{J}_{n_k}(\lambda_k) \end{bmatrix}\boldsymbol{S}^{-1}=\boldsymbol{S\Sigma S}^{-1}
$$

其中 $\boldsymbol{S}$ 为可逆矩阵

$$
\boldsymbol{J}_{n_k}(\lambda_k)=\begin{bmatrix} \lambda_k & 1 & & & 0 \\ & \lambda_k & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & \lambda_k \end{bmatrix}
$$

据此可以得到

$$
\boldsymbol{A}^2=\boldsymbol{S}\begin{bmatrix} \boldsymbol{J}^2_{n_1}(\lambda_1) & & & 0 \\ & \boldsymbol{J}^2_{n_2}(\lambda_2) & & \\ & & \ddots & \\ 0 & & & \boldsymbol{J}^2_{n_k}(\lambda_k) \end{bmatrix}\boldsymbol{S}^{-1}=\boldsymbol{S\Sigma}^2\boldsymbol{S}^{-1}
$$

从上面可以看出，经过 Jordan 标准型分解以后，$\boldsymbol{A}$ 和 $\boldsymbol{A}^2$ 的秩实际上由 $\boldsymbol{\Sigma}$ 和 $\boldsymbol{\Sigma}^2$ 决定。因为已知 $\operatorname{rank}(\boldsymbol{A}^2)=\operatorname{rank}(\boldsymbol{A})$，所以可以断定：矩阵 $\boldsymbol{A}$ 中一定不存在下面形式的 Jordan 块

$$
\boldsymbol{J}=\begin{bmatrix} 0 & 1 & & & 0 \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & 0 \end{bmatrix}
$$

否则，$\boldsymbol{\Sigma}^2$ 的秩将小于 $\boldsymbol{\Sigma}$ 的秩，$\boldsymbol{A}^2$ 的秩将小于 $\boldsymbol{A}$ 的秩。也就是说，矩阵 $\boldsymbol{A}$ 的 0 特征值对应的 Jordan 块一定是一阶的。

若定义 $\boldsymbol{C}=\boldsymbol{S}^{-1}\boldsymbol{BS}$，则易知

$$
\boldsymbol{AB}=\boldsymbol{O}\Leftrightarrow\boldsymbol{S\Sigma C}=\boldsymbol{O}\Leftrightarrow\boldsymbol{\Sigma C}=\boldsymbol{O}
$$

若将 $\boldsymbol{C}$ 按行分块，则与 $\boldsymbol{\Sigma}$ 的各个 Jordan 块的维数相对应，即

$$\boldsymbol{C}=\begin{bmatrix}\boldsymbol{C}_1 & \boldsymbol{C}_2 & \cdots & \boldsymbol{C}_{n_k}\end{bmatrix}^T$$

由 $\boldsymbol{\Sigma C}=\boldsymbol{O}$ 可知，$\boldsymbol{C}$ 矩阵与 $\boldsymbol{\Sigma}$ 非零 Jordan 块对应的行一定为零。

同理，由于 $\boldsymbol{BA}=\boldsymbol{O}$，则 $\boldsymbol{C}$ 矩阵与 $\boldsymbol{\Sigma}$ 非零 Jordan 块对应的列也一定为零。有

$$\operatorname{rank}(\boldsymbol{\Sigma}+\boldsymbol{C})=\operatorname{rank}(\boldsymbol{\Sigma})+\operatorname{rank}(\boldsymbol{C})$$

矩阵左乘 $\boldsymbol{S}$ 右乘 $\boldsymbol{S}^{-1}$ 不影响矩阵的秩，有

$$\operatorname{rank}(\boldsymbol{A}+\boldsymbol{B})=\operatorname{rank}(\boldsymbol{\Sigma}+\boldsymbol{C})=\operatorname{rank}(\boldsymbol{\Sigma})+\operatorname{rank}(\boldsymbol{C})=\operatorname{rank}(\boldsymbol{A})+\operatorname{rank}(\boldsymbol{B})$$

(2) 因为已知 $\operatorname{rank}(\boldsymbol{A}^2)=\operatorname{rank}(\boldsymbol{A})$，所以有

$$\operatorname{rank}(\boldsymbol{A}^{2k})=\operatorname{rank}(\boldsymbol{A}^k)$$

因为 $\boldsymbol{AB}=\boldsymbol{BA}=\boldsymbol{O}$，所以

$$\boldsymbol{A}^k\boldsymbol{B}^k=\boldsymbol{B}^k\boldsymbol{A}^k=\boldsymbol{O}$$

由 (1) 结论可得，$\operatorname{rank}(\boldsymbol{A}^k+\boldsymbol{B}^k)=\operatorname{rank}(\boldsymbol{A}^k)+\operatorname{rank}(\boldsymbol{B}^k)$

1.81 证明逆矩阵的以下性质：

(1) 逆矩阵的行列式等于原矩阵行列式的倒数，即 $|\boldsymbol{A}^{-1}|=\dfrac{1}{|\boldsymbol{A}|}$

(2) 逆矩阵是非奇异的。

(3) $(\boldsymbol{A}^{-1})^{-1}=\boldsymbol{A}$。

(4) 复共轭转置矩阵的逆矩阵 $(\boldsymbol{A}^{\mathrm{H}})^{-1}=(\boldsymbol{A}^{-1})^{\mathrm{H}}=\boldsymbol{A}^{-\mathrm{H}}$。

(5) 若 $\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$，则 $(\boldsymbol{A}^{-1})^{\mathrm{H}}=\boldsymbol{A}^{-1}$。

(6) $(\boldsymbol{A}^{*})^{-1}=(\boldsymbol{A}^{-1})^{*}$。

证明

(1)

$$\begin{aligned}&\boldsymbol{AA}^{-1}=\boldsymbol{I},\quad |\boldsymbol{A}||\boldsymbol{A}^{-1}|=1\\&|\boldsymbol{A}^{-1}|=\frac{1}{|\boldsymbol{A}|}\end{aligned}$$

(2)

$$\because |\boldsymbol{A}||\boldsymbol{A}^{-1}|=1,\quad \therefore |\boldsymbol{A}^{-1}|\neq 0$$

即逆矩阵非奇异

(3)

$$\begin{aligned}&(\boldsymbol{A}^{-1})^{-1}\boldsymbol{A}^{-1}=(\boldsymbol{AA}^{-1})^{-1}=\boldsymbol{I}\\&\boldsymbol{A}^{-1}(\boldsymbol{A}^{-1})^{-1}=(\boldsymbol{A}^{-1}\boldsymbol{A})^{-1}=\boldsymbol{I}\\&\therefore\ (\boldsymbol{A}^{-1})^{-1}=\boldsymbol{A}\end{aligned}$$

(4)

$$\begin{aligned}&(\boldsymbol{A}^{\mathrm{H}})^{-1}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{I},\quad \boldsymbol{AA}^{-1}=\boldsymbol{I}\\&(\boldsymbol{AA}^{-1})^{\mathrm{H}}=(\boldsymbol{A}^{-1})^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{I}^{\mathrm{H}}=\boldsymbol{I}\\&\text{同理 }\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}^{-1})^{\mathrm{H}}=\boldsymbol{I}\\&\therefore\ (\boldsymbol{A}^{\mathrm{H}})^{-1}=(\boldsymbol{A}^{-1})^{\mathrm{H}}=\boldsymbol{A}^{-\mathrm{H}}\end{aligned}$$

(5)

$$
\begin{aligned}
&\text{由（4）可知}(\boldsymbol{A}^{-1})^{\mathrm{H}} = (\boldsymbol{A}^{\mathrm{H}})^{-1} \\
&\because \boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A} \\
&\therefore (\boldsymbol{A}^{-1})^{\mathrm{H}} = (\boldsymbol{A}^{\mathrm{H}})^{-1} = \boldsymbol{A}^{-1}
\end{aligned}
$$

(6)

$$
\begin{aligned}
&|A|A^{-1} = A^{*}, \quad A^{-1} = \frac{A^{*}}{|A|} \\
&(A^{-1})^{*} = (\frac{A^{*}}{|A|})^{*} = \frac{(A^{*})^{*}}{|A|^{n-1}} = \frac{A}{|A|} \\
&(A^{*})^{-1} = \frac{A}{|A|} \\
&\therefore (A^{*})^{-1} = (A^{-1})^{*}
\end{aligned}
$$

1.103 设 $\boldsymbol{A}$ 是 Hermitian 矩阵，并且 $\boldsymbol{M}$ 是 $\boldsymbol{A}$ 的 Moore-Penrose 逆矩阵。证明：矩阵 $\boldsymbol{M}^2$ 是 $\boldsymbol{A}^2$ 的 Moore-Penrose 逆矩阵

解 由 Moore-Penrose 逆矩阵定义已知

$$
\boldsymbol{AMA} = \boldsymbol{A}, \boldsymbol{MAM} = \boldsymbol{M}, (\boldsymbol{AM})^{\mathrm{H}} = \boldsymbol{AM}, (\boldsymbol{MA})^{\mathrm{H}} = \boldsymbol{MA}
$$

(1)

$$
\begin{aligned}
\boldsymbol{A}^2\boldsymbol{M}^2\boldsymbol{A}^2 &= \boldsymbol{AAMMAA} = \boldsymbol{A}(\boldsymbol{AM})^{\mathrm{H}}(\boldsymbol{MA})^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{AM}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{A} \\
&= \boldsymbol{AM}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{AM}^{\mathrm{H}}\boldsymbol{AAM}^{\mathrm{H}}\boldsymbol{A} \\
&= (\boldsymbol{AM}^{\mathrm{H}}\boldsymbol{A})^2 = [(\boldsymbol{AMA})^{\mathrm{H}}]^2 = (\boldsymbol{A}^{\mathrm{H}})^2 \\
&= \boldsymbol{A}^2
\end{aligned}
$$

(2)

$$
\begin{aligned}
\boldsymbol{M}^2\boldsymbol{A}^2\boldsymbol{M}^2 &= \boldsymbol{MMAAMM} = \boldsymbol{M}(\boldsymbol{MA})^{\mathrm{H}}\boldsymbol{AMM} = \boldsymbol{MA}^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{AMM} \\
&= \boldsymbol{MA}^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{MM} = \boldsymbol{MA}^{\mathrm{H}}\boldsymbol{MM} = \boldsymbol{MAMM} \\
&= \boldsymbol{M}^2
\end{aligned}
$$

(3)

$$
\begin{aligned}
(\boldsymbol{M}^2\boldsymbol{A}^2)^{\mathrm{H}} &= \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}}(\boldsymbol{MA})^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}}\boldsymbol{MAM}^{\mathrm{H}} \\
&= \boldsymbol{AMAM}^{\mathrm{H}} = \boldsymbol{AM}^{\mathrm{H}} = (\boldsymbol{MA})^{\mathrm{H}} = \boldsymbol{MA} \\
\boldsymbol{M}^2\boldsymbol{A}^2 &= \boldsymbol{MMAA} = \boldsymbol{M}(\boldsymbol{MA})^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{MA}^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{MA}^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}} \\
&= \boldsymbol{M}(\boldsymbol{AMA})^{\mathrm{H}} = \boldsymbol{MA}^{\mathrm{H}} = \boldsymbol{MA} \\
\boldsymbol{M}^2\boldsymbol{A}^2 &= (\boldsymbol{M}^2\boldsymbol{A}^2)^{\mathrm{H}}
\end{aligned}
$$

(4)

$$
\begin{aligned}
(\boldsymbol{A}^2\boldsymbol{M}^2)^{\mathrm{H}} &= \boldsymbol{M}^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{M}^{\mathrm{H}}(\boldsymbol{AM})^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}} \\
&= \boldsymbol{M}^{\mathrm{H}}\boldsymbol{AMA} = \boldsymbol{M}^{\mathrm{H}}\boldsymbol{A} = (\boldsymbol{AM})^{\mathrm{H}} = \boldsymbol{AM} \\
\boldsymbol{A}^2\boldsymbol{M}^2 &= \boldsymbol{AAMM} = \boldsymbol{A}(\boldsymbol{AM})^{\mathrm{H}}\boldsymbol{M} = \boldsymbol{AM}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{M} \\
&= (\boldsymbol{AMA})^{\mathrm{H}}\boldsymbol{M} = \boldsymbol{AM} \\
(\boldsymbol{A}^2\boldsymbol{M}^2)^{\mathrm{H}} &= \boldsymbol{A}^2\boldsymbol{M}^2
\end{aligned}
$$

第二次 作业

✍ 课外习题：设 $\boldsymbol{X}$ 为 $n\times n$ 维矩阵变量，分别求 $\mathrm{tr}(\boldsymbol{X})$ 和 $\det(\boldsymbol{X})$ 的梯度矩阵

解

(1).1

$$\begin{aligned}\mathrm{dtr}(\boldsymbol{X}) &= \mathrm{tr}(\mathrm{d}\boldsymbol{X})\\ \mathrm{D}_{\boldsymbol{X}}\mathrm{tr}(\boldsymbol{X}) &= \boldsymbol{I}_n\\ \boldsymbol{\nabla}_{\boldsymbol{X}}\mathrm{tr}(\boldsymbol{X}) &= \boldsymbol{I}_n^{\mathrm{T}} = \boldsymbol{I}_n\end{aligned}$$

(1).2

$$\begin{aligned}\boldsymbol{\nabla}_{\boldsymbol{X}}\mathrm{tr}(\boldsymbol{X}) = \boldsymbol{\nabla}_{\boldsymbol{X}}\left(\sum_{i=1}^{n}x_{ii}\right) &= \begin{bmatrix}\frac{\partial\sum_{i=1}^{n}x_{ii}}{\partial x_{11}} & \frac{\partial\sum_{i=1}^{n}x_{ii}}{\partial x_{12}} & \cdots & \frac{\partial\sum_{i=1}^{n}x_{ii}}{\partial x_{1n}}\\ \frac{\partial\sum_{i=1}^{n}x_{ii}}{\partial x_{21}} & \frac{\partial\sum_{i=1}^{n}x_{ii}}{\partial x_{22}} & \cdots & \frac{\partial\sum_{i=1}^{n}x_{ii}}{\partial x_{2n}}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial\sum_{i=1}^{n}x_{ii}}{\partial x_{n1}} & \frac{\partial\sum_{i=1}^{n}x_{ii}}{\partial x_{n2}} & \cdots & \frac{\partial\sum_{i=1}^{n}x_{ii}}{\partial x_{nn}}\end{bmatrix}\\ &= \begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}_{n\times n}\\ &= \boldsymbol{I}_n\end{aligned}$$

(2).1 若 $\boldsymbol{X}$ 可逆

$$\begin{aligned}\mathrm{d}|\boldsymbol{X}| &= |\boldsymbol{X}|\mathrm{tr}(\boldsymbol{X}^{-1}\mathrm{d}\boldsymbol{X})\\ \mathrm{D}_{\boldsymbol{X}}|\boldsymbol{X}| &= |\boldsymbol{X}|\boldsymbol{X}^{-1}\\ \boldsymbol{\nabla}_{\boldsymbol{X}}|\boldsymbol{X}| &= |\boldsymbol{X}|\boldsymbol{X}^{-\mathrm{T}}\end{aligned}$$

(2).2 若 $\boldsymbol{X}$ 可逆

由 Laplace 定理得矩阵 $\boldsymbol{X}$ 的行列式可通过如下方式计算：

$$\det(\boldsymbol{X}) = \sum_{i=1}^{n}(-1)^{i+j}x_{ij}M_{ij}$$

其中 M_{ij} 是 $\boldsymbol{X}$ 在 (i,j) 处的余子式。因此有

$$\begin{aligned}\boldsymbol{\nabla}_{\boldsymbol{X}}\det(X) &= \boldsymbol{\nabla}_{\boldsymbol{X}}\left[\sum_{i=1}^{n}(-1)^{i+j}x_{ij}M_{ij}\right]\\ &= \begin{bmatrix}\frac{\partial\sum_{i=1}^{n}(-1)^{i+j}x_{ij}M_{ij}}{\partial x_{11}} & \frac{\partial\sum_{i=1}^{n}(-1)^{i+j}x_{ij}M_{ij}}{\partial x_{12}} & \cdots & \frac{\partial\sum_{i=1}^{n}(-1)^{i+j}x_{ij}M_{ij}}{\partial x_{1n}}\\ \frac{\partial\sum_{i=1}^{n}(-1)^{i+j}x_{ij}M_{ij}}{\partial x_{21}} & \frac{\partial\sum_{i=1}^{n}(-1)^{i+j}x_{ij}M_{ij}}{\partial x_{22}} & \cdots & \frac{\partial\sum_{i=1}^{n}(-1)^{i+j}x_{ij}M_{ij}}{\partial x_{2n}}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial\sum_{i=1}^{n}(-1)^{i+j}x_{ij}M_{ij}}{\partial x_{n1}} & \frac{\partial\sum_{i=1}^{n}(-1)^{i+j}x_{ij}M_{ij}}{\partial x_{n2}} & \cdots & \frac{\partial\sum_{i=1}^{n}(-1)^{i+j}x_{ij}M_{ij}}{\partial x_{nn}}\end{bmatrix}\\ &= \left[(-1)^{i+j}M_{ij}\right]_{n\times n} = (X_{ij})_{n\times n}\\ &= (\boldsymbol{X}^*)^{\mathrm{T}} = \det(\boldsymbol{X})(\boldsymbol{X}^{-1})^{\mathrm{T}}\end{aligned}$$

其中 $\boldsymbol{X}^*$ 为 $\boldsymbol{X}$ 的伴随矩阵。

若 X 不可逆，$\boldsymbol{\nabla}_X|\boldsymbol{X}| = \boldsymbol{O}$

✍ **3.2** 证明：若 $\phi(\boldsymbol{x}) = (\boldsymbol{f}(\boldsymbol{x}))^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}(\boldsymbol{x})$，则

$$\mathrm{D}_{\boldsymbol{x}}\phi(\boldsymbol{x}) = (\boldsymbol{g}(\boldsymbol{x}))^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\mathrm{D}_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x}) + (\boldsymbol{f}(\boldsymbol{x}))^{\mathrm{T}}\boldsymbol{A}\mathrm{D}_{\boldsymbol{x}}\boldsymbol{g}(\boldsymbol{x})$$

证明

1.

$$\begin{aligned}
\mathrm{d}\phi(\boldsymbol{x}) &= \mathrm{d}(\boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}(\boldsymbol{x})) \\
&= \mathrm{d}(\boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}})\boldsymbol{A}\boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{g}(\boldsymbol{x}) \\
&= \mathrm{d}\boldsymbol{x}^{\mathrm{T}}\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}\boldsymbol{A}\boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{A}\frac{\partial \boldsymbol{g}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}\mathrm{d}\boldsymbol{x} \\
&= \mathrm{tr}(\mathrm{d}\boldsymbol{x}^{\mathrm{T}}\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}\boldsymbol{A}\boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{A}\frac{\partial \boldsymbol{g}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}\mathrm{d}\boldsymbol{x}) \\
&= \mathrm{tr}(\boldsymbol{g}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}\mathrm{d}\boldsymbol{x} + \boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{A}\frac{\partial \boldsymbol{g}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}\mathrm{d}\boldsymbol{x}) \\
&= \mathrm{tr}([\boldsymbol{g}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\mathrm{D}_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{A}\mathrm{D}_{\boldsymbol{x}}\boldsymbol{g}(\boldsymbol{x})]\mathrm{d}\boldsymbol{x}) \\
\mathrm{D}_{\boldsymbol{x}}\phi(\boldsymbol{x}) &= (\boldsymbol{g}(\boldsymbol{x}))^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\mathrm{D}_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x}) + (\boldsymbol{f}(\boldsymbol{x}))^{\mathrm{T}}\boldsymbol{A}\mathrm{D}_{\boldsymbol{x}}\boldsymbol{g}(\boldsymbol{x})
\end{aligned}$$

2.

记

$$\begin{aligned}
\boldsymbol{f}(\boldsymbol{x}) &= [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_m(\boldsymbol{x})]^{\mathrm{T}} \\
\boldsymbol{g}(\boldsymbol{x}) &= [g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), \cdots, g_n(\boldsymbol{x})]^{\mathrm{T}} \\
\boldsymbol{A} &= (a_{ij})_{m\times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}
\end{aligned}$$

则可以将 $\phi(\boldsymbol{x})$ 写成标量的形式：

$$\phi(\boldsymbol{x}) = (\boldsymbol{f}(\boldsymbol{x}))^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}(\boldsymbol{x}) = \sum_{i=1}^{m}\sum_{j=1}^{n} f_i(\boldsymbol{x})a_{ij}g_j(\boldsymbol{x})$$

因此有

$$\begin{aligned}
\mathrm{D}_{\boldsymbol{x}}\phi(\boldsymbol{x}) &= \mathrm{D}_{\boldsymbol{x}}\sum_{i=1}^{m}\sum_{j=1}^{n} f_i(\boldsymbol{x})a_{ij}g_j(\boldsymbol{x}) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{D}_{\boldsymbol{x}} f_i(\boldsymbol{x})a_{ij}g_j(\boldsymbol{x}) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}[(\mathrm{D}_{\boldsymbol{x}} f_i(\boldsymbol{x}))g_j(\boldsymbol{x}) + f_i(\boldsymbol{x})(\mathrm{D}_{\boldsymbol{x}} g_j(\boldsymbol{x}))] \\
&= \sum_{i=1}^{m}\mathrm{D}_{\boldsymbol{x}} f_i(\boldsymbol{x})\sum_{j=1}^{n} a_{ij}g_j(\boldsymbol{x}) + \sum_{i=1}^{m} f_i(\boldsymbol{x})\sum_{j=1}^{n} a_{ij}\mathrm{D}_{\boldsymbol{x}} g_j(\boldsymbol{x}) \\
&= (\boldsymbol{g}(\boldsymbol{x}))^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\mathrm{D}_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x}) + (\boldsymbol{f}(\boldsymbol{x}))^{\mathrm{T}}\boldsymbol{A}\mathrm{D}_{\boldsymbol{x}}\boldsymbol{g}(\boldsymbol{x})
\end{aligned}$$

3.7 证明

$$\mathrm{d}[\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})] = 2\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X})$$

证明

$$\begin{aligned}\mathrm{d}[\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})] &= \mathrm{tr}[\mathrm{d}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})] \\ &= \mathrm{tr}[\mathrm{d}(\boldsymbol{X}^{\mathrm{T}})\boldsymbol{X} + \boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}] \\ &= \mathrm{tr}[\mathrm{d}(\boldsymbol{X}^{\mathrm{T}})\boldsymbol{X}] + \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}) \\ &= \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}) + \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}) \\ &= 2\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X})\end{aligned}$$

3.11 求实标量函数 $f(\boldsymbol{x}) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$ 和 $f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ 的 Hessian 矩阵。

解

(1)

$$\boldsymbol{H}[f(\boldsymbol{x})] = \frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}(\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}) = \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{x}^{\mathrm{T}}} = \boldsymbol{O}$$

(2)

$$\boldsymbol{H}[f(\boldsymbol{x})] = \frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}(\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}) = \frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}) = \boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}$$

3.14 求矩阵函数 $\boldsymbol{AXB}$ 和 $\boldsymbol{AX}^{-1}\boldsymbol{B}$ 的 Jacobian 矩阵。

解

(1)

$$\mathrm{d}(\boldsymbol{AXB}) = \boldsymbol{A}\mathrm{d}(\boldsymbol{X})\boldsymbol{B}$$

已知

$$\mathrm{d}\boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{A}(\mathrm{d}\boldsymbol{X})\boldsymbol{B} \Leftrightarrow \mathrm{d}(\mathrm{vec}\boldsymbol{F}(\boldsymbol{X})) = (\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{A})\mathrm{d}(\mathrm{vec}\boldsymbol{X})$$

则

$$\mathrm{D}(\boldsymbol{AXB}) = \boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{A}$$

(2)

$$\mathrm{d}(\boldsymbol{AX}^{-1}\boldsymbol{B}) = \boldsymbol{A}\mathrm{d}(\boldsymbol{X}^{-1})\boldsymbol{B} = -\boldsymbol{AX}^{-1}\mathrm{d}(\boldsymbol{X})\boldsymbol{X}^{-1}\boldsymbol{B}$$

已知

$$\mathrm{d}\boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{A}(\mathrm{d}\boldsymbol{X})\boldsymbol{B} \Leftrightarrow \mathrm{d}(\mathrm{vec}\boldsymbol{F}(\boldsymbol{X}) = (\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{A})\mathrm{d}(\mathrm{vec}\boldsymbol{X})$$

则

$$\mathrm{D}(\boldsymbol{AX}^{-1}\boldsymbol{B}) = -(\boldsymbol{X}^{-1}\boldsymbol{B})^{\mathrm{T}}\otimes\boldsymbol{AX}^{-1}$$

3.23 求下列实值函数的 Hessian 矩阵：

(1)$(\boldsymbol{Ax}+\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{Dx}+\boldsymbol{e})$。

(2)$(\boldsymbol{Ax}+\boldsymbol{b})^{\mathrm{T}}\boldsymbol{C}(\boldsymbol{Dx}+\boldsymbol{e})$。

(3)$(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}\boldsymbol{C}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})$。

解

1.

(1) 设 $f(\boldsymbol{x})=(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}+\boldsymbol{e})$

$$
\begin{aligned}
f(\boldsymbol{x}) &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{x}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{x}+\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{e}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{e}\\
\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}} &= \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D})+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{D}+\boldsymbol{e}^{\mathrm{T}}\boldsymbol{A}\\
\boldsymbol{H}[f(\boldsymbol{x})] &= \frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{D}+\boldsymbol{D}^{\mathrm{T}}\boldsymbol{A}
\end{aligned}
$$

(2) 设 $f(\boldsymbol{x})=(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}\boldsymbol{C}(\boldsymbol{D}\boldsymbol{x}+\boldsymbol{e})$

$$
\begin{aligned}
f(\boldsymbol{x}) &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{D}\boldsymbol{x}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{D}\boldsymbol{x}+\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{e}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{e}\\
\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}} &= \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{D})+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{D}+\boldsymbol{e}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\\
\boldsymbol{H}[f(\boldsymbol{x})] &= \frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{D}+\boldsymbol{D}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}
\end{aligned}
$$

(3) 设 $f(\boldsymbol{x})=(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}\boldsymbol{C}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})$

由 (2) 可得

$$\boldsymbol{H}[f(\boldsymbol{x})]=\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{C}+\boldsymbol{C}^{\mathrm{T}})\boldsymbol{A}$$

2.

(1) 先求一阶微分:

$$
\begin{aligned}
\mathrm{d}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}+\boldsymbol{e}) &= (\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}\mathrm{d}(\boldsymbol{D}\boldsymbol{x}+\boldsymbol{e})+\mathrm{d}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}+\boldsymbol{e})\\
&= (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{b}^{\mathrm{T}})\boldsymbol{D}\mathrm{d}\boldsymbol{x}+(\mathrm{d}\boldsymbol{x}^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}+\boldsymbol{e})\\
&= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D}\mathrm{d}\boldsymbol{x}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{D}\mathrm{d}\boldsymbol{x}+(\mathrm{d}\boldsymbol{x}^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{x}+(\mathrm{d}\boldsymbol{x}^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}\boldsymbol{e}
\end{aligned}
$$

再求二阶微分:

$$\mathrm{d}^2(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}+\boldsymbol{e})=\mathrm{d}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D}\mathrm{d}\boldsymbol{x})+\mathrm{d}[(\mathrm{d}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{x}]=2(\mathrm{d}\boldsymbol{x}^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D}\mathrm{d}\boldsymbol{x}$$

因此其 Hessian 矩阵为:

$$\boldsymbol{H}[(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}+\boldsymbol{e})]=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D}+(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D})^{\mathrm{T}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D}+\boldsymbol{D}^{\mathrm{T}}\boldsymbol{A}$$

(2) 先求一阶微分:

$$
\begin{aligned}
\mathrm{d}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}\boldsymbol{C}(\boldsymbol{D}\boldsymbol{x}+\boldsymbol{e}) &= (\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}\boldsymbol{C}\mathrm{d}(\boldsymbol{D}\boldsymbol{x}+\boldsymbol{e})+\mathrm{d}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}\mathrm{C}(\boldsymbol{D}\boldsymbol{x}+\boldsymbol{e})\\
&= (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{b}^{\mathrm{T}})\boldsymbol{C}\boldsymbol{D}\mathrm{d}\boldsymbol{x}+(\mathrm{d}\boldsymbol{x}^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}(\boldsymbol{D}\boldsymbol{x}+\boldsymbol{e})\\
&= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{D}\mathrm{d}\boldsymbol{x}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{D}\mathrm{d}\boldsymbol{x}+(\mathrm{d}\boldsymbol{x}^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{D}\boldsymbol{x}+(\mathrm{d}\boldsymbol{x}^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{e}
\end{aligned}
$$

再求二阶微分:

$$\mathrm{d}^2(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}C(\boldsymbol{D}\boldsymbol{x}+\boldsymbol{e})=\mathrm{d}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{D}\mathrm{d}\boldsymbol{x})+\mathrm{d}[(\mathrm{d}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{D}\boldsymbol{x}]=2(\mathrm{d}\boldsymbol{x}^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{D}\mathrm{d}\boldsymbol{x}$$

因此其 Hessian 矩阵为:

$$\boldsymbol{H}[(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}\boldsymbol{C}(\boldsymbol{D}\boldsymbol{x}+\boldsymbol{e})]=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{D}+(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{D})^{\mathrm{T}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{D}+\boldsymbol{D}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}$$

(3) 令 (2) 中的 $\boldsymbol{D}=\boldsymbol{A},\boldsymbol{e}=\boldsymbol{b}$，便有

$$\boldsymbol{H}[(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^{\mathrm{T}}\boldsymbol{C}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})]=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{A}+(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{A})^{\mathrm{T}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{C}^{\mathrm{T}}+\boldsymbol{C})\boldsymbol{A}$$

3.28 令 $\boldsymbol{x}$ 为复向量，求下列实值函数的复 Hessian 矩阵：
(1)$(\boldsymbol{Ax}+\boldsymbol{b})^{\mathrm{H}}(\boldsymbol{Dx}+\boldsymbol{e})$。
(2)$(\boldsymbol{Ax}+\boldsymbol{b})^{\mathrm{H}}\boldsymbol{C}(\boldsymbol{Dx}+\boldsymbol{e})$。
(3)$(\boldsymbol{Ax}+\boldsymbol{b})^{\mathrm{H}}\boldsymbol{C}(\boldsymbol{Ax}+\boldsymbol{b})$。

解 注：$\boldsymbol{C}$ 应当为 Hermitian 矩阵

(1) 设 $f(\boldsymbol{x})=(\boldsymbol{Ax}+\boldsymbol{b})^{\mathrm{H}}(\boldsymbol{Dx}+\boldsymbol{e})$

$$
\begin{aligned}
&f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{Dx}+\boldsymbol{b}^{\mathrm{H}}\boldsymbol{Dx}+\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{e}+\boldsymbol{b}^{\mathrm{H}}\boldsymbol{e}\\
&\boldsymbol{H}_{\boldsymbol{x}^*,\boldsymbol{x}}=\frac{\partial^2 f(\boldsymbol{x},\boldsymbol{x}^*)}{\partial\boldsymbol{x}^*\partial\boldsymbol{x}^{\mathrm{T}}}=\frac{\partial\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{D}+\boldsymbol{b}^{\mathrm{H}}\boldsymbol{D}}{\partial\boldsymbol{x}^*}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{D}\\
&\boldsymbol{H}_{\boldsymbol{x}^*,\boldsymbol{x}^*}=\frac{\partial^2 f(\boldsymbol{x},\boldsymbol{x}^*)}{\partial\boldsymbol{x}^*\boldsymbol{x}^{\mathrm{H}}}=\frac{\partial\boldsymbol{x}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{A}^*+\boldsymbol{e}^{\mathrm{T}}\boldsymbol{A}^*}{\partial\boldsymbol{x}^*}=\boldsymbol{O}\\
&\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}=\boldsymbol{H}^*_{\boldsymbol{x}^*,\boldsymbol{x}^*}=\boldsymbol{O}\quad \boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}^*}=\boldsymbol{H}^*_{\boldsymbol{x}^*,\boldsymbol{x}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D}^*
\end{aligned}
$$

取 $\boldsymbol{H}_{\boldsymbol{x}^*,\boldsymbol{x}}=\frac{1}{2}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{D}+\boldsymbol{D}^{\mathrm{H}}\boldsymbol{A})$，$\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}^*}=\frac{1}{2}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D}^*+\boldsymbol{D}^{\mathrm{T}}\boldsymbol{A}^*)$

$$
\boldsymbol{H}[f(\boldsymbol{x})]=\begin{bmatrix}\frac{1}{2}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{D}+\boldsymbol{D}^{\mathrm{H}}\boldsymbol{A}) & \boldsymbol{O}\\ \boldsymbol{O} & \frac{1}{2}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D}^*+\boldsymbol{D}^{\mathrm{T}}\boldsymbol{A}^*)\end{bmatrix}
$$

(2) 设 $f(\boldsymbol{x})=(\boldsymbol{Ax}+\boldsymbol{b})^{\mathrm{H}}\boldsymbol{C}(\boldsymbol{Dx}+\boldsymbol{e})$

$$
\begin{aligned}
&f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{CDx}+\boldsymbol{b}^{\mathrm{H}}\boldsymbol{CDx}+\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{Ce}+\boldsymbol{b}^{\mathrm{H}}\boldsymbol{Ce}\\
&\boldsymbol{H}_{\boldsymbol{x}^*,\boldsymbol{x}}=\frac{\partial^2 f(\boldsymbol{x},\boldsymbol{x}^*)}{\partial\boldsymbol{x}^*\partial\boldsymbol{x}^{\mathrm{T}}}=\frac{\partial\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{CD}+\boldsymbol{b}^{\mathrm{H}}\boldsymbol{CD}}{\partial\boldsymbol{x}^*}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{CD}\\
&\boldsymbol{H}_{\boldsymbol{x}^*,\boldsymbol{x}^*}=\frac{\partial^2 f(\boldsymbol{x},\boldsymbol{x}^*)}{\partial\boldsymbol{x}^*\boldsymbol{x}^{\mathrm{H}}}=\frac{\partial\boldsymbol{x}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}^*+\boldsymbol{e}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}^*}{\partial\boldsymbol{x}^*}=\boldsymbol{O}\\
&\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}=\boldsymbol{H}^*_{\boldsymbol{x}^*,\boldsymbol{x}^*}=\boldsymbol{O}\quad \boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}^*}=\boldsymbol{H}^*_{\boldsymbol{x}^*,\boldsymbol{x}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^*\boldsymbol{D}^*
\end{aligned}
$$

取 $\boldsymbol{H}_{\boldsymbol{x}^*,\boldsymbol{x}}=\frac{1}{2}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{CD}+\boldsymbol{D}^{\mathrm{H}}\boldsymbol{CA})$，$\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}^*}=\frac{1}{2}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^*\boldsymbol{D}^*+\boldsymbol{D}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}^*)$

$$
\boldsymbol{H}[f(\boldsymbol{x})]=\begin{bmatrix}\frac{1}{2}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{CD}+\boldsymbol{D}^{\mathrm{H}}\boldsymbol{CA}) & \boldsymbol{O}\\ \boldsymbol{O} & \frac{1}{2}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^*\boldsymbol{D}^*+\boldsymbol{D}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}^*)\end{bmatrix}
$$

(3) 设 $f(\boldsymbol{x})=(\boldsymbol{Ax}+\boldsymbol{b})^{\mathrm{H}}\boldsymbol{C}(\boldsymbol{Ax}+\boldsymbol{b})$

由 (2) 可得

$$
\boldsymbol{H}[f(\boldsymbol{x})]=\begin{bmatrix}\frac{1}{2}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{CA}+\boldsymbol{A}^{\mathrm{H}}\boldsymbol{CA}) & \boldsymbol{O}\\ \boldsymbol{O} & \frac{1}{2}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^*\boldsymbol{A}^*+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}^*)\end{bmatrix}
$$

第三次 作业

✍ **4.3(1)**

解

$$E_w = e^{\mathrm{H}} W e = (y - A\theta)^{\mathrm{H}} W (y - A\theta)$$

因此可以求其关于 $\boldsymbol{\theta}$ 的共轭梯度：

$$\nabla_{\theta^*} = A^{\mathrm{H}} W A \theta - A^{\mathrm{H}} W y$$

令共轭梯度为 0 求得 $\boldsymbol{\theta}$ 的加权最小二乘估计为：

$$\hat{\theta}_{\mathrm{LS}} = (A^{\mathrm{H}} W A)^{-1} A^{\mathrm{H}} W y$$

✍ **4.4**

解 将约束条件 $\mathrm{Re}(w^{\mathrm{H}} x) = b$ 转化为

$$w^{\mathrm{H}} x + x^{\mathrm{H}} w = 2b$$

则可以构造如下的 Lagrange 函数：

$$J(w, \lambda) = w^{\mathrm{H}} R_e w + \lambda(2b - w^{\mathrm{H}} x - x^{\mathrm{H}} w)$$

注意到代价函数 $f(w) = \mathrm{E}(w^{\mathrm{H}} e e^{\mathrm{H}} w)$ 是一个实标量函数，其约束条件也为实方程，因此构造的 Lagrange 函数应为实标量函数，即 $\lambda \in \mathbb{R}$。
将该函数分别关于 w 和 λ 求导并令其为零有：

$$\frac{\partial J(w, \lambda)}{\partial w} = R_e^{\mathrm{T}} w^* - \lambda x^* = 0$$
$$\frac{\partial J(w, \lambda)}{\partial \lambda} = 2b - w^{\mathrm{H}} x - x^{\mathrm{H}} w = 0$$

由于协方差矩阵 R_e 为 Hermitian 矩阵，且 $\lambda \in \mathbb{R}$，则上述两个条件可转化为：

$$R_e w = \lambda x$$
$$w^{\mathrm{H}} x + x^{\mathrm{H}} w = 2b$$

(1)R_e 为非奇异矩阵时：
可以得到 $w = \lambda R_e^{-1} x$，代入约束条件并利用 R_e 为 Hermitian 矩阵有：

$$\lambda x^{\mathrm{H}} R_e^{-1} x + \lambda x^{\mathrm{H}} R_e^{-1} x = 2b$$

因此可以解得

$$\lambda = \frac{b}{x^{\mathrm{H}} R_e^{-1} x}$$

进一步地可以得到最优滤波器 w 为

$$w = \frac{b R_e^{-1} x}{x^{\mathrm{H}} R_e^{-1} x}$$

(2)R_e 为奇异矩阵时：
易知 R_e 存在零特征值 λ_0，因此其对应的特征向量 v_0 满足

$$v_0 \in \ker R_e$$

也即

$$R_e v_0 = 0$$

因此当 $w = v_0$ 时，有

$$f(w) = f(v_0) = v_0^{\mathrm{H}} R_e v_0 = 0$$

故最优滤波器 $\boldsymbol{w}$ 为

$$\boldsymbol{w}=\boldsymbol{v}_0$$

✍ **4.12**

证明 首先证明约束条件的等价性：

$$\begin{aligned}\mathrm{E}\{\boldsymbol{e}-\boldsymbol{\varepsilon}\}&=\mathrm{E}\{\boldsymbol{A}\boldsymbol{y}-\boldsymbol{\varepsilon}\}\\&=\mathrm{E}\{\boldsymbol{A}(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})-\boldsymbol{\varepsilon}\}\\&=\boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta}+\mathrm{E}\{\boldsymbol{A}\boldsymbol{\varepsilon}-\boldsymbol{\varepsilon}\}\\&=\boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{0}\end{aligned}$$

因此 $\mathrm{E}\{\boldsymbol{e}-\boldsymbol{\varepsilon}\}=\boldsymbol{0}$ 等价于

$$\boldsymbol{A}\boldsymbol{X}=\boldsymbol{O}$$

之后证明最优化问题的等价性：

$$\begin{aligned}\mathrm{E}\{(\boldsymbol{e}-\boldsymbol{\varepsilon})^{\mathrm{T}}(\boldsymbol{e}-\boldsymbol{\varepsilon})\}&=\mathrm{E}\{(\boldsymbol{A}\boldsymbol{y}-\boldsymbol{\varepsilon})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{y}-\boldsymbol{\varepsilon})\}\\&=\mathrm{E}\{(\boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{A}\boldsymbol{\varepsilon}-\boldsymbol{\varepsilon})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{A}\boldsymbol{\varepsilon}-\boldsymbol{\varepsilon})\}\\&=\mathrm{E}\{(\boldsymbol{A}\boldsymbol{\varepsilon}-\boldsymbol{\varepsilon})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{\varepsilon}-\boldsymbol{\varepsilon})\}\\&=\mathrm{E}\{\boldsymbol{\varepsilon}^{\mathrm{T}}(\boldsymbol{A}-\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{A}-\boldsymbol{I})\boldsymbol{\varepsilon}\}\\&=\mathrm{E}\{\mathrm{tr}[(\boldsymbol{A}-\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{A}-\boldsymbol{I})\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}]\}\\&=\mathrm{tr}\{(\boldsymbol{A}-\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{A}-\boldsymbol{I})\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}})\}\\&=\mathrm{tr}\{(\boldsymbol{A}-\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{A}-\boldsymbol{I})\sigma^2\boldsymbol{I}\}\\&=\sigma^2[\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})-2\mathrm{tr}(\boldsymbol{A})+\mathrm{tr}(\boldsymbol{I})]\end{aligned}$$

因此最优化问题 $\min\mathrm{E}\{(\boldsymbol{e}-\boldsymbol{\varepsilon})^{\mathrm{T}}(\boldsymbol{e}-\boldsymbol{\varepsilon})\}$ 等价于

$$\min\,[\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})-2\mathrm{tr}(\boldsymbol{A})]$$

✍ **课外习题 1**

解 构造如下的 Lagrange 函数：

$$J(\boldsymbol{x},\boldsymbol{\lambda})=||\boldsymbol{x}||_2^2+\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x})$$

将该函数分别对 $\boldsymbol{x}$ 和 $\boldsymbol{\lambda}$ 求导并令其为零有：

$$\begin{aligned}\frac{\partial J(\boldsymbol{x},\boldsymbol{\lambda})}{\partial\boldsymbol{x}}&=2\boldsymbol{x}-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}=\boldsymbol{0}\\\frac{\partial J(\boldsymbol{x},\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}&=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}\end{aligned}$$

因此可以得到关于 $\boldsymbol{\lambda}$ 和 $\boldsymbol{x}$ 的方程组：

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}=2\boldsymbol{x}\tag{1}$$

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}\tag{2}$$

首先消去 $\boldsymbol{x}$，对 (1) 式两边同时左乘 $\boldsymbol{A}$ 有

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}=2\boldsymbol{A}\boldsymbol{x}=2\boldsymbol{b}$$

则 $\boldsymbol{\lambda}$ 可以表示成如下的形式：

$$\boldsymbol{\lambda}=2(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{\dagger}\boldsymbol{b}+[\boldsymbol{I}-(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}]\boldsymbol{y}$$

将上式代入 (1) 式中有

$$\begin{aligned} \boldsymbol{x}^* &= \frac{1}{2}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} = \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{\dagger}\boldsymbol{b} + \frac{1}{2}\boldsymbol{A}^{\mathrm{T}}[\boldsymbol{I} - (\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}]\boldsymbol{y} \\ &= \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{\dagger}\boldsymbol{b} + \frac{1}{2}[\boldsymbol{A}^{\mathrm{T}} - \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}]\boldsymbol{y} \end{aligned}$$

由 Moore-Penrose 逆矩阵的定义和性质可知：

$$\boldsymbol{A}^{\dagger} = \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{\dagger} \qquad (\boldsymbol{A}^{\mathrm{T}})^{\dagger} = (\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{\dagger}\boldsymbol{A}$$

进一步化简可得:

$$\begin{aligned} \boldsymbol{x}^* &= \boldsymbol{A}^{\dagger}\boldsymbol{b} + \frac{1}{2}[\boldsymbol{A}^{\mathrm{T}} - \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}})^{\dagger}\boldsymbol{A}^{\mathrm{T}}]\boldsymbol{y} \\ &= \boldsymbol{A}^{\dagger}\boldsymbol{b} \end{aligned}$$

课外习题 2

解 首先求出 $f(\boldsymbol{x})$ 的梯度：

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{bmatrix} 4x_1^3 + 6x_1^2 + 4x_1 - 2x_2 \\ -2x_1 + 2x_2 \end{bmatrix}$$

令梯度为 0 可得到如下方程组：

$$\begin{cases} 4x_1^3 + 6x_1^2 + 4x_1 - 2x_2 = 0 \\ -2x_1 + 2x_2 = 0 \end{cases}$$

解得

$$\begin{cases} x_1 = 0 \\ x_2 = 0 \end{cases} \text{或} \begin{cases} x_1 = -\frac{1}{2} \\ x_2 = -\frac{1}{2} \end{cases} \text{或} \begin{cases} x_1 = -1 \\ x_2 = -1 \end{cases}$$

即有三个一阶稳定点

$$\boldsymbol{x}^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{x}^{(2)} = \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}, \boldsymbol{x}^{(3)} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

再求 $f(\boldsymbol{x})$ 的 Hessian 矩阵：

$$\boldsymbol{H}[f(\boldsymbol{x})] = \begin{bmatrix} 12x_1^2 + 12x_1 + 4 & -2 \\ -2 & 2 \end{bmatrix}$$

分别代入 $\boldsymbol{x}^{(1)}$、$\boldsymbol{x}^{(2)}$、$\boldsymbol{x}^{(3)}$ 可得：

$$\boldsymbol{H}[f(\boldsymbol{x}^{(1)})] = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix}, \boldsymbol{H}[f(\boldsymbol{x}^{(2)})] = \begin{bmatrix} 1 & -2 \\ -2 & 2 \end{bmatrix}, \boldsymbol{H}[f(\boldsymbol{x}^{(3)})] = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix}$$

由

$$\mathrm{tr}\{\boldsymbol{H}[f(\boldsymbol{x})]\} = \lambda_1 + \lambda_2, |\boldsymbol{H}[f(\boldsymbol{x})]| = \lambda_1\lambda_2$$

可以通过求 Hessian 矩阵的迹和行列式判断其特征值符号，进而判断其正定性。易知 $\boldsymbol{H}[f(\boldsymbol{x}^{(1)})]$ 和 $\boldsymbol{H}[f(\boldsymbol{x}^{(3)})]$ 正定，$\boldsymbol{H}[f(\boldsymbol{x}^{(2)})]$ 不定，因此 $\boldsymbol{x}^{(2)}$ 为鞍点，$\boldsymbol{x}^{(1)}$ 和 $\boldsymbol{x}^{(3)}$ 都是局部最优点（极小值点），代入数据可得

$$f(\boldsymbol{x}^{(1)}) = f(\boldsymbol{x}^{(3)}) = 0$$

另外容易看出

$$f(\boldsymbol{x}) = x_1^2(x_1 + 1)^2 + (x_1 - x_2)^2 \geq 0$$

因此 $\boldsymbol{x}^{(1)}$ 和 $\boldsymbol{x}^{(3)}$ 都是全局最小值点。

第四次 作业

✍ **6.1** 考虑线性方程 $\boldsymbol{A\theta}+\boldsymbol{\varepsilon}=\boldsymbol{x}$，其中，$\boldsymbol{\varepsilon}$ 为加性有色噪声向量，满足条件 $\mathbb{E}\{\boldsymbol{\varepsilon}\}=\boldsymbol{0}$ 和 $\mathbb{E}\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\}=\boldsymbol{R}$。令 $\boldsymbol{R}$ 已知，并使用加权误差函数 $Q(\boldsymbol{\theta})=\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\varepsilon}$ 作为求参数向量 $\boldsymbol{\theta}$ 最优估计 $\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}$ 的代价函数。这种方法称为加权最小二乘方法。证明

$$\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x}$$

其中，加权矩阵 $\boldsymbol{W}$ 的最优选择为 $\boldsymbol{W}_{\mathrm{opt}}=\boldsymbol{R}^{-1}$。

证明 首先，将 $\boldsymbol{\varepsilon}=\boldsymbol{x}-\boldsymbol{A\theta}$ 代入到 $Q(\boldsymbol{\theta})$ 中得到

$$\begin{aligned}Q(\boldsymbol{\theta})&=\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\varepsilon}\\&=(\boldsymbol{x}-\boldsymbol{A\theta})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{x}-\boldsymbol{A\theta})\\&=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x}-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A\theta}-\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x}+\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A\theta}\end{aligned}$$

考虑到加权矩阵 $\boldsymbol{W}$ 为对称矩阵，因此求 $Q(\boldsymbol{\theta})$ 的梯度并令其为零有

$$\frac{\partial Q(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A\theta}-2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x}=\boldsymbol{0}$$

因此得到最优的参数估计

$$\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x}$$

因此估计误差的协方差矩阵可表示为

$$\begin{aligned}K&=\mathbb{E}\{(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{\mathrm{WLS}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{\mathrm{WLS}})^{\mathrm{T}}\}\\&=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\end{aligned}$$

为了求得最优的加权矩阵，需要使估计误差的均方误差最小，也即最小化 $\boldsymbol{K}$ 的迹，即令

$$\frac{\partial\mathrm{tr}(\boldsymbol{K})}{\partial\boldsymbol{W}}=\boldsymbol{O}$$

求 $\mathrm{tr}(\boldsymbol{K})$ 的微分，有

$$\begin{aligned}\mathrm{dtr}(\boldsymbol{K})&=\mathrm{tr}(\mathrm{d}\boldsymbol{K})\\&=\mathrm{tr}[\mathrm{d}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\cdot\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}+(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\mathrm{d}\boldsymbol{W}\cdot\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\\&\quad+(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{R}\mathrm{d}\boldsymbol{W}\cdot\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}+(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}\mathrm{d}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}]\end{aligned}$$

由 P163 表 3.2.4 可知

$$\mathrm{d}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}=-(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\mathrm{d}\boldsymbol{W}\cdot\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}$$

并由矩阵迹的性质 $\mathrm{tr}(\boldsymbol{AB})=\mathrm{tr}(\boldsymbol{BA})$ 可知

$$\begin{aligned}\mathrm{dtr}(\boldsymbol{K})=&\mathrm{tr}[-\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}\mathrm{d}\boldsymbol{W}+\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}\mathrm{d}\boldsymbol{W}\\&+\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{R}\mathrm{d}\boldsymbol{W}-\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\mathrm{d}\boldsymbol{W}]\end{aligned}$$

为了使过程更加直观，我们记 $\boldsymbol{B}=\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}},\boldsymbol{C}=\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}$，则可化为

$$\mathrm{dtr}(\boldsymbol{K})=\mathrm{tr}[-(\boldsymbol{B}+\boldsymbol{B}^{\mathrm{T}})\mathrm{d}\boldsymbol{W}]+\mathrm{tr}[(\boldsymbol{C}+\boldsymbol{C}^{\mathrm{T}})\mathrm{d}\boldsymbol{W}]$$

因此可以得到其梯度为

$$\frac{\partial\mathrm{tr}(\boldsymbol{K})}{\partial\boldsymbol{W}}=(\boldsymbol{C}+\boldsymbol{C}^{\mathrm{T}})-(\boldsymbol{B}+\boldsymbol{B}^{\mathrm{T}})=\boldsymbol{O}$$

即 $\boldsymbol{C}=\boldsymbol{B}$ 或 $\boldsymbol{C}=\boldsymbol{B}^{\mathrm{T}}$, 展开得

$$\begin{aligned}&\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}\qquad\text{或}\\&\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}\end{aligned}$$

可以看出上述两式均在 $\boldsymbol{W}=\boldsymbol{R}^{-1}$ 时成立，因此加权矩阵 $\boldsymbol{W}$ 的最优选择为 $\boldsymbol{W}_{\mathrm{opt}}=\boldsymbol{R}^{-1}$。

6.4 令 $\lambda>0$，并且 $\boldsymbol{Ax}=\boldsymbol{b}$ 为超定方程。证明：反 Tikhonov 正则化优化问题

$$\min \frac{1}{2}\|\boldsymbol{Ax}-\boldsymbol{b}\|_2^2-\frac{1}{2}\lambda\|\boldsymbol{x}\|_2^2$$

的最优解为

$$\boldsymbol{x}=(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}-\lambda\boldsymbol{I})^{-1}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{b}$$

证明 令代价函数

$$\begin{aligned}
f(\boldsymbol{x})&=\frac{1}{2}\|\boldsymbol{Ax}-\boldsymbol{b}\|_2^2-\frac{1}{2}\lambda\|\boldsymbol{x}\|_2^2\\
&=\frac{1}{2}(\boldsymbol{Ax}-\boldsymbol{b})^{\mathrm{H}}(\boldsymbol{Ax}-\boldsymbol{b})-\frac{1}{2}\lambda\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}\\
&=\frac{1}{2}(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}-\boldsymbol{b}^{\mathrm{H}})(\boldsymbol{Ax}-\boldsymbol{b})-\frac{1}{2}\lambda\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}\\
&=\frac{1}{2}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{Ax}-\frac{1}{2}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{b}-\frac{1}{2}\boldsymbol{b}^{\mathrm{H}}\boldsymbol{Ax}+\frac{1}{2}\boldsymbol{b}^{\mathrm{H}}\boldsymbol{b}-\frac{1}{2}\lambda\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}
\end{aligned}$$

令其梯度为零有

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{Ax}-\boldsymbol{A}^{\mathrm{H}}\boldsymbol{b}-\lambda\boldsymbol{x}=\boldsymbol{0}$$

因此最优解为

$$\boldsymbol{x}=(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}-\lambda\boldsymbol{I})^{-1}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{b}$$

6.7 已知数据点 $(1,3),(3,1),(5,7),(4,6),(7,4)$，分别求总体最小二乘和一般最小二乘的拟合直线，并分析它们的距离平方和。

解 (1) 总体最小二乘：

i) 计算均值向量

$$\bar{\boldsymbol{x}}=\frac{1}{5}\sum_{i=1}^{5}\boldsymbol{x}_i=\frac{1}{5}\left\{\begin{bmatrix}1\\3\end{bmatrix}+\begin{bmatrix}3\\1\end{bmatrix}+\begin{bmatrix}5\\7\end{bmatrix}+\begin{bmatrix}4\\6\end{bmatrix}+\begin{bmatrix}7\\4\end{bmatrix}\right\}=\begin{bmatrix}4\\4.2\end{bmatrix}$$

ii) 构造 5×2 矩阵 $\boldsymbol{M}$

$$\boldsymbol{M}=\begin{bmatrix}(\boldsymbol{x}_1-\bar{\boldsymbol{x}})^{\mathrm{T}}\\ \vdots\\ (\boldsymbol{x}_5-\bar{\boldsymbol{x}})^{\mathrm{T}}\end{bmatrix}=\begin{bmatrix}-3&-1.2\\-1&-3.2\\1&2.8\\0&1.8\\3&-0.2\end{bmatrix}$$

iii) 对 $\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}$ 进行 EVD

$$\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}=\begin{bmatrix}20&9\\9&20.8\end{bmatrix}=\begin{bmatrix}0.6505&-0.7595\\0.7595&0.6505\end{bmatrix}\begin{bmatrix}30.5082&0\\0&12.2918\end{bmatrix}\begin{bmatrix}0.6505&-0.7595\\0.7595&0.6505\end{bmatrix}^{\mathrm{T}}$$

iv) 确定直线方程：$-0.7595(x-4)+0.6505(y-4.2)=0$

v) 计算距离平方和

$$D_{\mathrm{TLS}}=\left\|\begin{bmatrix}-3&-1.2\\-1&-3.2\\1&2.8\\0&1.8\\3&-0.2\end{bmatrix}\begin{bmatrix}-0.7595\\0.6505\end{bmatrix}\right\|_2^2=12.2918$$

(2) 一般最小二乘：

i) 代价函数表示为

$$D_{\mathrm{LS}}^{(1)}=\sum_{i=1}^{5}[m(x_i-\bar{x})^2+(y_i-\bar{y})^2]=(-3m-1.2)^2+(-m-3.2)^2+(m+2.8)^2+1.8^2+(3m-0.2)^2$$

$$\frac{\partial D_{\mathrm{LS}}^{(1)}}{\partial m}=40m+18=0\Rightarrow m=-0.45$$

则直线方程为

$$-0.45(x-4)+(y-4.2)=0$$

距离平方和为

$$d_{\mathrm{LS}}^{(1)}=\frac{D_{\mathrm{LS}}^{(1)}}{1+m^2}=15.5925$$

ii) 代价函数表示为

$$D_{\mathrm{LS}}^{(2)}=\sum_{i=1}^{5}[(x_i-\bar{x})^2+m(y_i-\bar{y})^2]=(-3-1.2m)^2+(-1-3.2m)^2+(1+2.8m)^2+(1.8m)^2+(3-0.2m)^2$$

$$\frac{\partial D_{\mathrm{LS}}^{(2)}}{\partial m}=45.6m+18=0\Rightarrow m=-0.3947$$

则直线方程为

$$(x-4)-0.3947(y-4.2)=0$$

距离平方和为

$$d_{\mathrm{LS}}^{(2)}=\frac{D_{\mathrm{LS}}^{(2)}}{1+m^2}=14.2305$$

7.7 证明以下各题:

(1) 若 $\boldsymbol{A}=\boldsymbol{A}^2=\boldsymbol{A}^{\mathrm{H}}\in\mathbb{C}^{n\times n}$，并且 $\operatorname{rank}(\boldsymbol{A})=r<n$，则存在 $n\times n$ 酉矩阵 $\boldsymbol{V}$，使得

$$\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}=\operatorname{diag}(\boldsymbol{I}_r,\boldsymbol{0})$$

(2) 若 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}\in\mathbb{C}^{n\times n}$，并且 $\boldsymbol{A}^2=\boldsymbol{I}_n$，则存在酉矩阵 $\boldsymbol{V}$，使得

$$\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}=\operatorname{diag}(\boldsymbol{I}_r,-\boldsymbol{I}_{n-r})$$

证明 (1) 由于 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}$，则 $\boldsymbol{A}$ 酉相似于对角阵，即存在 $n\times n$ 酉矩阵 $\boldsymbol{V}$ 和 $n\times n$ 对角矩阵 $\boldsymbol{\Sigma}$，使得

$$\boldsymbol{A}=\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$$

由于 $\boldsymbol{A}=\boldsymbol{A}^2$，因此有 $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}^2$，即对于对角矩阵 $\boldsymbol{\Sigma}$ 上的任意一个对角元素 σ_i，有 $\sigma_i=\sigma_i^2$，即 σ_i 总为 0 或 1。那么总可以交换 $\boldsymbol{V}$ 中列向量的顺序，使得所有的 $\sigma_i=1$ 集中在一起，所有的 $\sigma_i=0$ 集中在一起，由于 $\operatorname{rank}(\boldsymbol{A})=r$，因此总可以得到

$$\boldsymbol{\Sigma}=\operatorname{diag}(\boldsymbol{I}_r,\boldsymbol{0})$$

由 $\boldsymbol{V}$ 是酉矩阵的性质，得到

$$\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}=\operatorname{diag}(\boldsymbol{I}_r,\boldsymbol{0})$$

(2) 与 (1) 同理可以得到

$$\boldsymbol{\Sigma}^2=\boldsymbol{I}_n$$

因此对于对角矩阵 $\boldsymbol{\Sigma}$ 上的任意一个对角元素 σ_i，有 $\sigma_i^2=1$，即 $\sigma_i=\pm1$。那么总可以交换 $\boldsymbol{V}$ 中列向量的顺序，使得所有的 $\sigma_i=1$ 集中在一起，所有的 $\sigma_i=-1$ 集中在一起，因此总可以得到

$$\boldsymbol{\Sigma}=\operatorname{diag}(\boldsymbol{I}_r,-\boldsymbol{I}_{n-r})$$

由 $\boldsymbol{V}$ 是酉矩阵的性质，得到

$$\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}=\operatorname{diag}(\boldsymbol{I}_r,-\boldsymbol{I}_{n-r})$$

✍ **7.17** 已知 $\boldsymbol{u}=[1,1,-1]^{\mathrm{T}}$ 是矩阵 $\boldsymbol{A}=\begin{bmatrix}2 & -1 & 2\\ 5 & a & 3\\ -1 & b & -2\end{bmatrix}$ 的一个特征向量。

(1) 求 a,b 和特征向量 $\boldsymbol{u}$ 对应的特征值。

(2) 矩阵 $\boldsymbol{A}$ 能否相似于对角矩阵？试说明理由。

解 (1) 令 $\boldsymbol{A}\boldsymbol{u}=\lambda\boldsymbol{u}$，有

$$\begin{bmatrix}2 & -1 & 2\\ 5 & a & 3\\ -1 & b & -2\end{bmatrix}\begin{bmatrix}1\\ 1\\ -1\end{bmatrix}=\begin{bmatrix}-1\\ 2+a\\ 1+b\end{bmatrix}=\lambda\begin{bmatrix}1\\ 1\\ -1\end{bmatrix}$$

解得 $\lambda=-1,a=-3,b=0.$

(2) 由 (1) 可知矩阵 $\boldsymbol{A}$ 为

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

令 $\det(\lambda\boldsymbol{I}-\boldsymbol{A})=0$，有

$$(\lambda+1)^3=0$$

即特征值 $\lambda=-1$ 的代数多重度为 3。

考虑如下关于 $\boldsymbol{u}$ 的方程

$$(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{u}\Big|_{\lambda=-1}=\begin{bmatrix}3 & -1 & 2\\ 5 & -2 & 3\\ -1 & 0 & -1\end{bmatrix}\boldsymbol{u}=\boldsymbol{0}$$

其系数矩阵满足

$$\operatorname{rank}(\boldsymbol{A}+\boldsymbol{I})=\operatorname{rank}\begin{bmatrix}3 & -1 & 2\\ 5 & -2 & 3\\ -1 & 0 & -1\end{bmatrix}=2$$

因此其基础解系为 $n-\operatorname{rank}(\boldsymbol{A}+\boldsymbol{I})=1$ 个，从而特征值 $\lambda=-1$ 的几何多重度为 1，因此代数多重度大于几何多重度，矩阵 $\boldsymbol{A}$ 不能相似于对角矩阵。

✍ **7.34** 令 $\boldsymbol{A}$ 为实斜对称矩阵，即其元素 $a_{ij}=-a_{ji}$。证明：

(1) $\boldsymbol{A}$ 的特征值是纯虚数或零。

(2) 若 $\boldsymbol{u}+\mathrm{j}\boldsymbol{v}$ 是与特征值 $\mathrm{j}\mu$(其中 μ 是非零的实数) 对应的特征向量，并且 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 为实向量，则 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 正交。

证明 (1) 由 $\boldsymbol{A}$ 为实斜对称矩阵可知

$$\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{O}$$

设 $\boldsymbol{A}$ 的特征值 λ 及其对应的特征向量 $\boldsymbol{x}$，则有

$$\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$$

取共轭转置，有

$$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}=\lambda^*\boldsymbol{x}^{\mathrm{H}}$$

由 $\boldsymbol{A}$ 为实斜对称矩阵有

$$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{T}}=-\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}=\lambda^*\boldsymbol{x}^{\mathrm{H}}$$

对上式的等式两侧同时右乘向量 $\boldsymbol{x}$ 有

$$-\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}=-\lambda\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}=\lambda^*\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}$$

因此

$$(\lambda+\lambda^*)\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}=0$$

由于 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}=||\boldsymbol{x}||_2^2>0$，因此

$$\lambda+\lambda^*=0$$

即 $\boldsymbol{A}$ 的特征值为纯虚数或零。

(2) 由已知得

$$\boldsymbol{A}(\boldsymbol{u}+\mathrm{j}\boldsymbol{v})=\mathrm{j}\mu(\boldsymbol{u}+\mathrm{j}\boldsymbol{v})=-\mu\boldsymbol{v}+\mathrm{j}\mu\boldsymbol{u}$$

因此

$$\boldsymbol{A}\boldsymbol{u}=-\mu\boldsymbol{v},\boldsymbol{A}\boldsymbol{v}=\mu\boldsymbol{u}$$

左式左乘 $\boldsymbol{u}^{\mathrm{T}}$，右式取转置有

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u}=-\mu\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}$$
$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}=-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}=\mu\boldsymbol{u}^{\mathrm{T}}$$

将下式代入上式有

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u}=\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}=\mu\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}$$

$$\mu\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}=-\mu\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}$$

因此有

$$\mu(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}+\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u})=2\mu(\boldsymbol{u},\boldsymbol{v})=0$$

即 $(\boldsymbol{u},\boldsymbol{v})=0$，$\boldsymbol{u}$ 与 $\boldsymbol{v}$ 正交。

7.36 一滤波器的抽头延迟线的输出为 $y(k)=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}(k)$，其中 $\boldsymbol{a}=[a_0,a_1,\cdots,a_n]^{\mathrm{T}}$ 和 $\boldsymbol{x}(k)=[x(k),x(k-1),\cdots,x(k-n)]^{\mathrm{T}}$。令 $\boldsymbol{R}_{\boldsymbol{x}}\overset{\mathrm{def}}{=}\mathbb{E}\{\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\}=\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}^{\mathrm{T}}$，其中 $\boldsymbol{\Sigma}=\mathrm{diag}(\lambda_0,\lambda_1,\cdots,\lambda_n)$。如果输出序列 $\{y(k)\}$ 的均方值为 $J_{\boldsymbol{a}}=\frac{1}{2}\mathbb{E}\{y^2(k)\}$，证明以下结果：

(1) 在条件 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}=1$ 的约束下，使 $J_{\boldsymbol{a}}$ 最小化等于 $J_{\boldsymbol{w}}=\frac{1}{2}\sum_{i=0}^{n}w_i^2\lambda_i$ 的最小化，其中 $\boldsymbol{w}=[w_0,w_1,\cdots,w_n]^{\mathrm{T}}$，$\sum_{i=0}^{n}w_i^2=1$，且 $\boldsymbol{w}=\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{a}$。

(2) 若取 $\boldsymbol{w}=[\pm1,0,\cdots,0]^{\mathrm{T}}$，则使 $J_{\boldsymbol{a}}$ 最小化的最优向量 $\boldsymbol{a}=\pm\boldsymbol{a}_0$，其中，$\boldsymbol{a}_0$ 是矩阵 $\boldsymbol{R}_{\boldsymbol{x}}$ 相对于最小特征值 λ_0 的特征向量。

证明 (1) 首先证明优化问题的等价性：

$$\begin{aligned}J_{\boldsymbol{a}}&=\frac{1}{2}\mathbb{E}\{y^2(k)\}=\frac{1}{2}\mathbb{E}\{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}(k)\boldsymbol{x}(k)^{\mathrm{T}}\boldsymbol{a}\}\\&=\frac{1}{2}\boldsymbol{a}^{\mathrm{T}}\mathbb{E}\{\boldsymbol{x}(k)\boldsymbol{x}(k)^{\mathrm{T}}\}\boldsymbol{a}\\&=\frac{1}{2}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{a}\\&=\frac{1}{2}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{a})^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{a}\\&=\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{w}\\&=\frac{1}{2}\sum_{i=0}^{n}w_i^2\lambda_i\end{aligned}$$

再说明约束条件的等价性：

由于 $\boldsymbol{Q}$ 为正交矩阵，因此有

$$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}=(\boldsymbol{Q}\boldsymbol{w})^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{w}=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{w}=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}=\sum_{i=0}^{n}w_i^2=1$$

(2) 若使 $J_{\boldsymbol{a}}$ 取得最小值，也即 $J_{\boldsymbol{w}}$ 取得最小值，考虑到约束条件 $\sum_{i=0}^{n} w_i^2 = 1$ 时，$J_{\boldsymbol{w}}$ 的最小值在 $\boldsymbol{w} = [\pm 1, 0, \cdots, 0]^{\mathrm{T}}$ 时取到，设 $\boldsymbol{Q} = [\boldsymbol{a}_0, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_n]^{\mathrm{T}}$，则有

$$\boldsymbol{a} = \boldsymbol{Q}\boldsymbol{w} = \begin{bmatrix} \boldsymbol{a}_0 & \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \end{bmatrix} \begin{bmatrix} \pm 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \pm \boldsymbol{a}_0$$

✍ **7.55** 假定 $n \times n$ 维 Hermitian 矩阵 $\boldsymbol{A}$ 的特征值按照顺序 $\lambda_1(\boldsymbol{A}) \geq \lambda_2(\boldsymbol{A}) \geq \cdots \geq \lambda_n(\boldsymbol{A})$ 排列。用 Rayleigh 商证明：

(1)$\lambda_1(\boldsymbol{A}+\boldsymbol{B}) \geq \lambda_1(\boldsymbol{A}) + \lambda_n(\boldsymbol{B})$。

(2)$\lambda_n(\boldsymbol{A}+\boldsymbol{B}) \geq \lambda_n(\boldsymbol{A}) + \lambda_n(\boldsymbol{B})$

证明 由 Rayleigh 商的性质可知，矩阵的最大和最小特征值分别为其 Rayleigh 商的最大和最小值。

(1)

$$\begin{aligned} \lambda_1(\boldsymbol{A}+\boldsymbol{B}) &= \max \frac{\boldsymbol{x}^{\mathrm{H}}(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \\ &= \max \left(\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} + \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \right) \\ &\geq \max \left(\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} + \lambda_n(\boldsymbol{B}) \right) \\ &= \max \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} + \lambda_n(\boldsymbol{B}) \\ &= \lambda_1(\boldsymbol{A}) + \lambda_n(\boldsymbol{B}) \end{aligned}$$

(2)

$$\begin{aligned} \lambda_n(\boldsymbol{A}+\boldsymbol{B}) &= \min \frac{\boldsymbol{x}^{\mathrm{H}}(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \\ &= \min \left(\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} + \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \right) \\ &\geq \min \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} + \min \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \\ &= \lambda_n(\boldsymbol{A}) + \lambda_n(\boldsymbol{B}) \end{aligned}$$

第五次 作业

5.1 已知矩阵

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

通过计算 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 和 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的特征值和特征向量，求矩阵 $\boldsymbol{A}$ 的奇异值分解。

解 矩阵 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 为

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=\begin{bmatrix}2&2&0\\2&2&0\\0&0&0\end{bmatrix}$$

其特征值为 4,0,0，对应的特征向量为

$$\begin{bmatrix}\frac{\sqrt{2}}{2}\\\frac{\sqrt{2}}{2}\\0\end{bmatrix},\begin{bmatrix}-\frac{\sqrt{2}}{2}\\\frac{\sqrt{2}}{2}\\0\end{bmatrix},\begin{bmatrix}0\\0\\1\end{bmatrix}$$

矩阵 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 为

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\begin{bmatrix}2&2\\2&2\end{bmatrix}$$

其特征值为 4,0，对应的特征向量为

$$\begin{bmatrix}\frac{\sqrt{2}}{2}\\\frac{\sqrt{2}}{2}\end{bmatrix},\begin{bmatrix}-\frac{\sqrt{2}}{2}\\\frac{\sqrt{2}}{2}\end{bmatrix}$$

因此矩阵 $\boldsymbol{A}$ 的奇异值分解为

$$\boldsymbol{A}=\begin{bmatrix}\frac{\sqrt{2}}{2}&-\frac{\sqrt{2}}{2}&0\\\frac{\sqrt{2}}{2}&\frac{\sqrt{2}}{2}&0\\0&0&1\end{bmatrix}\begin{bmatrix}2&0\\0&0\\0&0\end{bmatrix}\begin{bmatrix}\frac{\sqrt{2}}{2}&-\frac{\sqrt{2}}{2}\\\frac{\sqrt{2}}{2}&\frac{\sqrt{2}}{2}\end{bmatrix}^{\mathrm{T}}$$

5.9 令 $\boldsymbol{A}$ 为 $m\times n$ 矩阵，且 $\boldsymbol{P}$ 为 $m\times m$ 正交矩阵。证明 $\boldsymbol{PA}$ 与 $\boldsymbol{A}$ 的奇异值相同。矩阵 $\boldsymbol{PA}$ 与 $\boldsymbol{A}$ 的左、右奇异向量有何关系？

证明 设 $\boldsymbol{A}$ 的奇异值分解为

$$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$$

由于 $\boldsymbol{P}$ 和 $\boldsymbol{U}$ 都是正交矩阵，则显然 $\boldsymbol{PU}$ 也是正交矩阵，因此 $\boldsymbol{PA}$ 的奇异值分解为

$$\boldsymbol{PA}=\boldsymbol{PU}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$$

因此 $\boldsymbol{PA}$ 与 $\boldsymbol{A}$ 的奇异值相同，且 $\boldsymbol{PA}$ 与 $\boldsymbol{A}$ 的右奇异向量也相同，$\boldsymbol{PA}$ 的左奇异向量为 $\boldsymbol{PU}$ 的列向量，$\boldsymbol{A}$ 的左奇异向量为 $\boldsymbol{U}$ 的列向量。

5.10 令 $\boldsymbol{A}$ 是一个 $m\times n$ 矩阵，并且 $\lambda_1,\cdots,\lambda_n$ 是矩阵 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的特征值，相对应的特征向量为 $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n$。证明 $\boldsymbol{A}$ 的奇异值 σ_i 等于范数 $||\boldsymbol{A}\boldsymbol{u}_i||$，即 $\sigma_i=||\boldsymbol{A}\boldsymbol{u}_i||, i=1,\cdots,n$。

证明 由于 $\lambda_i=\sigma_i^2$，有

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u}_i=\sigma_i^2\boldsymbol{u}_i$$

两边同时左乘 $\boldsymbol{u}_i^{\mathrm{T}}$ 得

$$(\boldsymbol{A}\boldsymbol{u}_i)^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{u}_i)=\sigma_i^2\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_i$$

由于 $\boldsymbol{u_i}$ 为单位向量，因此有

$$\sigma_i = ||\boldsymbol{A}\boldsymbol{u_i}||$$

5.15 假定计算机仿真的观测数据为

$$x(n) = \sqrt{20}\sin(2\pi 0.2n) + \sqrt{2}\sin(2\pi 0.215n) + w(n)$$

产生，其中，$w(n)$ 是一高斯白噪声，其均值为 0，方差为 1，并取 $n = 1, 2, \cdots, 128$。试针对 10 次独立的仿真实验数据，分别确定自相关矩阵

$$\boldsymbol{R} = \begin{bmatrix} r(0) & r(-1) & \cdots & r(2p) \\ r(1) & r(0) & \cdots & r(-2p+1) \\ \vdots & \vdots & \ddots & \vdots \\ r(M) & r(M-1) & \cdots & r(M-2p) \end{bmatrix}$$

的有效秩。式中，$r(k) = \frac{1}{128}\sum_{i=1}^{128-k} x(i)x(i+k)$ 表示观测信号的样本自相关函数 (未知的观测数据皆令其等于 0)，并取 $M = 50, p = 10$。

解 采用 P297 的归一化奇异值方法计算有效秩，即有效秩为

$$\hat{r} = \max_i \frac{\sigma_i}{\sigma_1} \geq \varepsilon$$

这里阈值 ε 分别取 3% 和 5%。给出 Python 代码如下：

```python
import numpy as np
import matplotlib.pyplot as plt
import itertools
T = 128
M = 50
p = 10
n = range(1, T +1)
e = [0.03, 0.05]
R = np.zeros((M+1,2*p+1))
def signal_x(n: int) ->float:
"""
Calculate the value of the signal at a given time point.
Args:
n: The time point at which to calculate the signal value.
Returns:
float: The value of the signal at the given time point.
"""
if n in range(1, T +1):
    return (
        np.sqrt(20) *np.sin(2 *np.pi *0.2 *n)
        + np.sqrt(2) *np.sin(2 *np.pi *0.215 *n)
        + np.random.randn()
    )
else:
    return 0
def acf(x, k: int, T: int) ->float:
"""
Calculate the autocorrelation function (ACF) at lag k for a given signal.
Args:
```

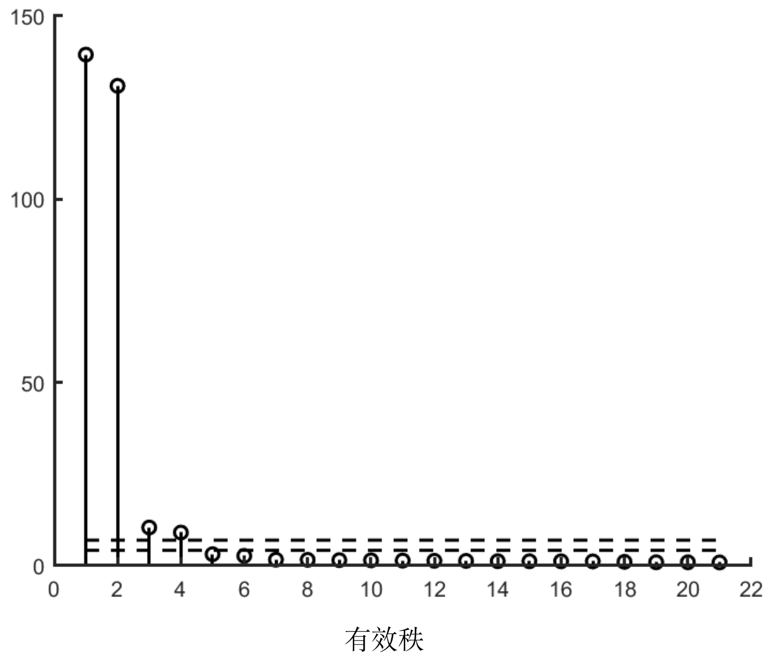
```
x: The signal function.
k: The lag at which to calculate the ACF.
T: The total number of time points in the signal.
Returns:
float: The value of the ACF at the specified lag.
"""
    return sum(x(i) *x(i +k) for i in range(1, T -k +1)) /T
for i, j in itertools.product(range(M+1), range(2*p+1)):
    R[i,j] =acf(signal_x, i-j, T)
u, s, v =np.linalg.svd(R)
s /=s[0]
for e_this in e:
    for rank in range(1,2*p+2):
        if s[rank]/s[0] <e_this:
            break
        else:
            continue
    print(f'选择阈值为{e_this}时的有效秩为{rank}')
```

两个不同的阈值运行结果均显示有效秩为 4.

给出 Matlab 代码如下：

```
clear
clc
T =128;M =50;p =10;
e =0.05;n =1: T;
x =sqrt(20)* sin(2* pi *0.2* n)
    + sqrt(2)* sin(2* pi *0.215* n)
    + randn(1 , T);
r =xcorr(x)/ T ;
R =zeros(M +1 ,2* p +1);
for i =0: M
    R (m +1,:) =r(T +m :-1: T +m -2* p);
end
sigma =svd(R);
for i =1 :length(sigma)
    if sigma(i)/ sigma(1) <e
        r =i -1;
        break;
    end
end
```

两个不同的阈值运行结果均显示有效秩为 4.


150
100
50
0
0
2
4
6
8
10
12
14
16
18
20
22
有效秩