

# 第十二讲

王伟文 暨南大学

## 1 矩阵计算

### 定义 1.1: 矩阵函数

$A \in \mathbb{R}^{n \times n}$  为对称矩阵, 其谱分解为

$$A = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^T,$$

若有函数  $f: \mathbb{R} \rightarrow \mathbb{R}$ , 则

$$f(A) \triangleq \sum_{i=1}^n f(\lambda_i) \mathbf{u}_i \mathbf{u}_i^T \in \mathbb{R}^{n \times n}.$$

例 1.1.

$$\begin{aligned} e^A &= \sum_{i=1}^n e^{\lambda_i} \mathbf{u}_i \mathbf{u}_i^T \\ &= \sum_{i=1}^n \sum_{k=0}^{\infty} \frac{\lambda_i^k}{k!} \mathbf{u}_i \mathbf{u}_i^T \\ &= \sum_{k=0}^{\infty} \sum_{i=1}^n \frac{\lambda_i^k}{k!} \mathbf{u}_i \mathbf{u}_i^T \\ &= I + \sum_{k=1}^{\infty} \frac{A^k}{k!} \end{aligned}$$

### 定义 1.2: 正半定矩阵 (Positive-Semidefinite, PSD)

称矩阵  $A \in \mathbb{R}^{n \times n}$  为正定矩阵当且仅当对任意向量  $\mathbf{u} \in \mathbb{R}^n$  有  $\mathbf{u}^T A \mathbf{u} \geq 0$ , 等价于  $\lambda_i(A) \geq 0, \forall i \in [n]$ .

### 定义 1.3: Löwner 秩

矩阵  $A$  为 PSD, 则  $A \succeq 0$ . 若  $X - Y \succeq 0$ , 则  $X \succeq Y, Y \preceq X$ .

评论. Löwner 秩非完全秩, 即存在两个矩阵不能判断它们之间的 Löwner 秩.

### 命题 1.1: Löwner 秩的性质

- (a) 若  $X \succeq Y$ ,  $A \in \mathbb{R}^{n \times n}$  为任意  $n$  阶矩阵, 则  $A^T X A \succeq A^T Y A$ .
- (b) 若  $X \preceq Y$ , 则  $\lambda_i(X) \leq \lambda_i(Y), \forall i \in [n]$ ; 进一步若  $X$  与  $Y$  可交换, 则逆命题成立.
- (c) 设  $f: \mathbb{R} \rightarrow \mathbb{R}$  为单调递增函数, 若  $X \preceq Y$ , 则  $\text{trace} f(X) \leq \text{trace} f(Y)$ .

**证明.** (a) 给定  $n$  阶矩阵  $A$ , 因为  $X - Y \succeq 0$  为 PSD, 故对任意向量  $\mathbf{u} \in \mathbb{R}^n$  有

$$(\mathbf{A}\mathbf{u})^T (X - Y)(\mathbf{A}\mathbf{u}) \geq 0,$$

即  $\mathbf{u}^T (A^T X A - A^T Y A) \mathbf{u} \geq 0 \Rightarrow A^T X A - A^T Y A \succeq 0$ . 故  $A^T X A \succeq A^T Y A$ .

(b) 设  $\mathbb{S}$  为  $\mathbb{R}^n$  的子空间, 由 Courant-Fisher 最小最大定理

$$\lambda_i(X) = \max_{\dim \mathbb{S}=i} \min_{\substack{\mathbf{u} \in \mathbb{S} \\ \|\mathbf{u}\|=1}} \mathbf{u}^T X \mathbf{u}$$

可直接得到结论.

(c)  $\text{trace}(X) = \sum_{i=1}^n \lambda_i(X)$  结合 (b) 得到. □

### 命题 1.2

若  $X$  可逆且  $0 \preceq X \preceq Y$ , 则  $Y$  可逆且  $X^{-1} \succeq Y^{-1} \succeq 0$ .

**证明.** 因为  $X$  可逆, 故  $\lambda_i \geq 0, \forall i \in [n]$ . 又因为  $Y \succeq X$ , 故  $\lambda_i(Y) \geq \lambda_i(X) > 0, \forall i \in [n]$ . 故  $Y$  可逆.

若  $Y = I$ , 即  $I \succeq X$ , 故有  $X^{-\frac{1}{2}} I X^{-\frac{1}{2}} \succeq X^{-\frac{1}{2}} I X^{-\frac{1}{2}} \Rightarrow X^{-1} \succeq I$ .

若  $Y \succeq X$ ,  $Y^{-\frac{1}{2}} Y Y^{-\frac{1}{2}} \succeq Y^{-\frac{1}{2}} X Y^{-\frac{1}{2}} \Rightarrow I \succeq Y^{-\frac{1}{2}} X Y^{-\frac{1}{2}} \Rightarrow \left( Y^{-\frac{1}{2}} X Y^{-\frac{1}{2}} \right)^{-1} \succeq I$ . 即有

$$Y^{-\frac{1}{2}} X^{-1} Y^{-\frac{1}{2}} \succeq I \Rightarrow X^{-1} \succeq Y^{-1}.$$

□

### 命题 1.3

若  $X$  可逆且  $0 \preceq X \preceq Y$ , 则  $\log X \preceq \log Y$ .

**证明.** 利用恒等式

$$\log x \equiv \int_0^\infty \left( \frac{1}{1+t} - \frac{1}{x+t} \right) dt \quad \forall x > 0$$

$$\begin{aligned}
\log X &= \sum_{i=1}^n \int_0^\infty \left( \frac{1}{1+t} - \frac{1}{\lambda_i(X)+t} \right) dt \cdot \mathbf{u}_i \mathbf{u}_i^T \\
&= \int_0^\infty \sum_{i=1}^n \left( \frac{1}{1+t} \mathbf{u}_i \mathbf{u}_i^T - \frac{1}{\lambda_i(X)+t} \mathbf{u}_i \mathbf{u}_i^T \right) dt \\
&= \int_0^\infty ((I+tI)^{-1} - (tI+X)^{-1}) dt
\end{aligned}$$

因为  $0 \preceq X \preceq Y \Rightarrow tI+X \preceq tI+Y$ , 故  $(tI+X)^{-1} \succeq (tI+Y)^{-1}, \forall t > 0$ .

$$\begin{aligned}
\log Y - \log X &= \int_0^\infty ((I+tI)^{-1} - (tI+Y)^{-1}) dt - \int_0^\infty ((I+tI)^{-1} - (tI+X)^{-1}) dt \\
&= \int_0^\infty ((tI+X)^{-1} - (tI+Y)^{-1}) dt \\
&\succeq 0,
\end{aligned}$$

故  $\log Y \succeq \log X$ . □

## 2 矩阵 Hoeffding 不等式

### 定理 2.1: Hoeffding 不等式

设  $\epsilon_i$  独立等概率地取  $\pm 1$ ,  $a_i \in R$ , 则

$$\mathbb{P} \left[ \left| \sum_{i=1}^n \epsilon_i a_i \right| \geq t \right] \leq 2 \exp \left( -\frac{t^2}{2\sigma^2} \right) \quad t \geq 0.$$

其中  $\sigma^2 = \sum_{i=1}^n a_i^2$ .

证明. 记  $S_n = \sum_{i=1}^n \epsilon_i a_i$ .

$$\begin{aligned}
\mathbb{P}[S_n \geq t] &= \mathbb{P}[e^{\lambda S_n} \geq e^{\lambda t}] \quad \forall \lambda \geq 0. \\
&\leq e^{-\lambda t} \mathbb{E}[e^{\lambda S_n}] = e^{-\lambda t} \mathbb{E} \left[ e^{\lambda \sum_{i=1}^n \epsilon_i a_i} \right] \\
&= e^{-\lambda t} \mathbb{E} \left[ \prod_{i=1}^n e^{\lambda \epsilon_i a_i} \right] = e^{-\lambda t} \prod_{i=1}^n \mathbb{E}[e^{\lambda \epsilon_i a_i}] \\
&= e^{-\lambda t} \prod_{i=1}^n \left( \frac{e^{-\lambda a_i}}{2} + \frac{e^{\lambda a_i}}{2} \right) \\
&\leq e^{-\lambda t} \prod_{i=1}^n e^{\lambda^2 a_i^2 / 2} = \exp \left( -\lambda t + \frac{\lambda^2}{2} \sum_{i=1}^n a_i^2 \right)
\end{aligned}$$

关于  $\lambda \geq 0$  最小化不等式右边得到

$$\mathbb{P}[S_n \geq t] \leq \exp\left(-\frac{t^2}{2\sigma^2}\right) \quad \forall t \geq 0.$$

□

### 定理 2.2: Lieb 不等式

设  $H \in R^{n \times n}$  为  $n$  阶对称矩阵, 定义函数

$$f(X) \triangleq \text{trace}(\exp(H + \log X)),$$

则函数  $f$  在  $\{X : X \succeq 0\}$  为凹函数.

### 推论 2.1

设  $H \in R^{n \times n}$  为  $n$  阶对称实矩阵,  $Z$  为  $n$  阶对称随机矩阵, 则

$$\mathbb{E} \text{trace}(\exp(H + Z)) \leq \text{trace}(\exp(H + \log \mathbb{E} e^Z))$$

证明.  $Z \triangleq \log e^Z$ , 结合 Lieb 不等式和 Jensen 不等式得证.

□

### 推论 2.2

设  $X_i$  为一组相互独立的同阶对称随机矩阵, 则

$$\mathbb{E} \text{trace} \exp\left(\sum_{i=1}^n X_i\right) \leq \text{trace} \exp\left(\sum_{i=1}^n \log \mathbb{E} e^{X_i}\right)$$

证明.

$$\begin{aligned} & \mathbb{E}_{X_i: i \in [n-1]} \left[ \mathbb{E}_{X_n} \text{trace} \exp\left(\sum_{i=1}^{n-1} X_i + X_n\right) \right] \\ & \leq \mathbb{E}_{X_i: i \in [n-1]} \left[ \text{trace} \exp\left(\sum_{i=1}^{n-1} X_i + \log \mathbb{E} e^{X_n}\right) \right] \\ & \leq \dots \\ & \leq \text{trace} \exp\left(\sum_{i=1}^n \log \mathbb{E} e^{X_i}\right) \end{aligned}$$

□

### 定理 2.3: 矩阵 Hoeffding 不等式

设  $\epsilon_i$  独立等概率取  $\pm 1$ ,  $A_i$  为  $d$ -阶对称矩阵, 则

$$\mathbb{P} \left[ \left\| \sum_{i=1}^n \epsilon_i A_i \right\| \geq t \right] \leq 2d \cdot \exp \left( -\frac{t^2}{2\sigma^2} \right) \quad \forall t \geq 0,$$

其中  $\sigma^2 = \left\| \sum_{i=1}^n A_i^2 \right\|$ .

**证明.** 记  $S_n = \sum_{i=1}^n \epsilon_i A_i$ ,  $\|S_n\| = \max(\lambda_1(S_n), \lambda_1(-S_n))$

对任意  $\lambda \geq 0$ ,

$$\begin{aligned} \mathbb{P}[\lambda_1(S_n) \geq t] &= \mathbb{P}[e^{\lambda \cdot \lambda_1(S_n)} \geq e^{\lambda t}] \\ &\leq e^{-\lambda t} \mathbb{E} [e^{\lambda \cdot \lambda_1(S_n)}] \\ &\leq e^{-\lambda t} \mathbb{E} \left[ \sum_{i=1}^n e^{\lambda \cdot \lambda_i(S_n)} \right] = e^{-\lambda t} \mathbb{E} \text{trace} \exp(\lambda S_n) \\ &= e^{-\lambda t} \mathbb{E} \text{trace} \exp \left( \sum_{i=1}^n \lambda \epsilon_i A_i \right) \\ &\leq e^{-\lambda t} \text{trace} \exp \left( \sum_{i=1}^n \log \mathbb{E} e^{\lambda \epsilon_i A_i} \right) \end{aligned}$$

因为

$$\frac{e^{-A} + e^A}{2} \preceq e^{\frac{A^2}{2}}$$

由  $\log$  的单调性, 故有

$$\log \mathbb{E} e^{\lambda \epsilon_i A_i} \preceq \log e^{\frac{\lambda^2 A_i^2}{2}} = \frac{\lambda^2 A_i^2}{2}$$

因此

$$\sum_{i=1}^n \log \mathbb{E} e^{\lambda \epsilon_i A_i} \preceq \sum_{i=1}^n \frac{\lambda^2 A_i^2}{2}.$$

又因为  $\exp(\cdot)$  为单调递增函数, 故

$$\text{trace} \exp \left( \sum_{i=1}^n \log \mathbb{E} e^{\lambda \epsilon_i A_i} \right) \leq \text{trace} \exp \left( \sum_{i=1}^n \frac{\lambda^2 A_i^2}{2} \right)$$

记  $Z = \sum_{i=1}^n A_i^2$ ,  $\sigma^2 = \|Z\|$ , 则有

$$\begin{aligned} \mathbb{P}[\lambda_1(S_n) \geq t] &\leq \exp(-\lambda t) \cdot \text{trace} \exp \left( \frac{\lambda^2}{2} Z \right) \\ &= \exp(-\lambda t) \cdot \sum_{k=1}^d \exp \left( \frac{\lambda^2 \cdot \lambda_k(Z)}{2} \right) \quad (\lambda_k(Z) \leq \sigma^2) \\ &\leq \exp(-\lambda t) \cdot d \exp \left( \frac{\lambda^2 \sigma^2}{2} \right) \end{aligned}$$

不等式右边关于  $\lambda \geq 0$  最小化即

$$\mathbb{P}[\lambda_1(S_n) \geq t] \leq d \exp\left(-\frac{t^2}{2\sigma^2}\right) \quad \forall t \geq 0.$$

□