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

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

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

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

例 1.1.

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

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 $\boldsymbol{u} \in \mathbb{R}^n$ 有 $\boldsymbol{u}^T A \boldsymbol{u} \geq 0$, 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 $X \succeq Y$, $A \in R^{n\times n}$ 为任意 n 阶矩阵, 则 $A^TXA \succeq A^TYA$.

(b) 若 $X \preceq Y$, 则 $\lambda_i(X) \le \lambda(Y), \forall i \in [n]$; 进一步若 X 与 Y 可交换, 则逆命题成立.

(c) 设 $f: \mathbb{R} \to \mathbb{R}$ 为单调递增函数, 若 $X \preceq Y$, 则 $\mathtt{trace} f(X) \le \mathtt{trace} f(Y)$.

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

$$(A\boldsymbol{u})^T(X-Y)(A\boldsymbol{u}) \ge 0,$$

即 $u^T(A^TXA - A^TYA)\boldsymbol{u} \ge 0 \Rightarrow A^TXA - A^TYA \succeq 0$. 故 $A^TXA \succeq A^TYA$.

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

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

可直接得到结论.

(c) $\mathtt{trace}(X) = \sum_{i=1}^{n} \lambda_i(X)$ 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 $\lambda_i \ge 0, \forall i \in [n]$. 又因为 $Y \succeq X$, 故 $\lambda_i(Y) \ge \lambda_i(X) > 0, \forall i \in [n]$. 故 Y 可逆.

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

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

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

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

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

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

故 $\log Y \succeq \log X$. □

2 矩阵 Hoeffing 不等式

定理 2.1: Hoeffding 不等式

设 ϵ_i 独立等概率地取 ± 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 $S_n = \sum_{i=1}^{n} \epsilon_i a_i$.

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

关于 $\lambda \geq 0$ 最小化不等式右边得到

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

□

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

$$f(X) \triangleq \texttt{trace}\left(\exp(H + \log X)\right),$$

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

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

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

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

证明.

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

□

定理 2.3: 矩阵 Hoeffing 不等式

设 ϵ_i 独立等概率取 ± 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 $\sigma^2 = \left\|\sum\limits_{i=1}^{n} A_i^2\right\|$.

证明. 记 $S_n = \sum\limits_{i=1}^{n}\epsilon_i A_i$, $\|S_n\| = \max\left(\lambda_1(S_n), \lambda_1(-S_n)\right)$

对任意 $\lambda\geq 0$,

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

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

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

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

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

不等式右边关于 $\lambda \ge 0$ 最小化即

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

□

3 矩阵 Bernstein 不等式

定理 3.1: 矩阵 Bernstein 不等式

设 $X_1, \ldots, X_N$ 为均值为 0, 相互独立的 d 阶对称随机矩阵, 且 $\forall i$

$$\|X_i\| \le K \quad \text{a.s.},$$

则对任意 $t \ge 0$,

$$\mathbb{P}\left[\left\|\sum_{i=1}^N X_i\right\| \ge t\right] \le 2d \exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right)$$

这里 $\sigma^2 = \left\|\sum_{i=1}^N \mathbb{E}X_i^2\right\|$.

记号 若存在常数 $C > 0$, 使得 $a \le Cb$, 则该不等式简记为 $a \lesssim b$.

推论 3.1

在定理 3.1 条件下,

$$\mathbb{E}\left\|\sum_{i=1}^N X_i\right\| \lesssim \left\|\sum_{i=1}^N \mathbb{E}X_i^2\right\|^{1/2} \sqrt{\log(2d)} + K\log(2d).$$

定理 3.2

设 $\{X_i\}_{i=1}^N$ 为与 X 同一分布的独立 $d(d \ge 2)$-维随机样本, 其中 $\mathbb{E}XX^T = \Sigma$. 记 $\Sigma_N = \frac{1}{N}\sum_{i=1}^N X_i X_i^T$. 若存在 $K \ge 1$, 使得

$$\|X\| \le K(\mathbb{E}\|X\|^2)^{1/2} \quad \text{a.s.},$$

则

$$\mathbb{E}\|\Sigma_N - \Sigma\| \le C\left(\sqrt{\frac{K^2 d \log d}{N}} + \frac{K^2 d \log d}{N}\right)\|\Sigma\| \quad \forall N \in \mathbb{N}_+$$

证明. $X_i X_i^T - \Sigma$ 相互独立且均值为 0, 由推论 3.1

$$\mathbb{E}\left\|\sum_{i=1}^N (X_i X_i^T - \Sigma)\right\| \lesssim \sigma\sqrt{\log d} + M\log d.$$

其中 $\sigma^2 = \left\| \sum_{i=1}^{N} \mathbb{E}(X_i X_i^T - \Sigma)^2 \right\| = N \left\| \mathbb{E}(XX^T - \Sigma)^2 \right\|$, M 满足

$$\|XX^T - \Sigma\| \leq M \quad \text{a.s.}$$

故

$$\mathbb{E}\|\Sigma_N - \Sigma\| = \frac{1}{N}\mathbb{E}\left\| \sum_{i=1}^{N}(X_i X_i^T - \Sigma) \right\| \lesssim \frac{1}{N}\left(\sigma\sqrt{\log d} + M\log d\right)$$

控制 σ 对任意 d 阶对称矩阵 A, 有 $A^2 \equiv A \cdot A$. 故

$$\mathbb{E}(XX^T - \Sigma)^2 = \mathbb{E}(XX^T)^2 - \Sigma^2 \preceq \mathbb{E}(XX^T)^2$$

又因为

$$\begin{aligned}
(XX^T)^2 &\preceq \lambda_{\max}(XX^T)XX^T \\
&= \|X\|^2 XX^T \\
&\leq K^2\mathbb{E}(\|X\|^2)XX^T \quad \text{a.s.}
\end{aligned}$$

即

$$\mathbb{E}(XX^T)^2 \preceq K^2\mathbb{E}(\|X\|^2)\Sigma.$$

综合上述不等式有

$$\|\mathbb{E}(XX^T - \Sigma)^2\| \leq \|\mathbb{E}(XX^T)^2\| \leq K^2\mathbb{E}(\|X\|^2)\|\Sigma\|$$

此外

$$\mathbb{E}\|X\|^2 = \mathbb{E}\texttt{trace}(X^TX) = \texttt{trace}(\Sigma) \leq d\|\Sigma\|$$

故

$$\sigma^2 = N\|\mathbb{E}(XX^T - \Sigma)^2\| \leq NK^2 d\|\Sigma\|^2$$

即

$$\sigma \leq K\sqrt{Nd}\|\Sigma\|$$

控制 M

$$\begin{aligned}
\|XX^T - \Sigma\| &\leq \|XX^T\| + \|\Sigma\| \\
&\leq \|X\|^2 + \|\Sigma\| \\
&\leq K^2\mathbb{E}\|X\|^2 + \|\Sigma\| \\
&\leq 2K^2 d\|\Sigma\| : M \quad \text{a.s.} \quad \forall K \geq 1.
\end{aligned}$$

综上所述

$$\begin{aligned}\mathbb{E}\|\Sigma_N-\Sigma\| &\lesssim \frac{1}{N}\left(\sigma\sqrt{\log(d)}+M\log(d)\right)\\ &\lesssim \left(\sqrt{\frac{K^2 d\log d}{N}}+\frac{K^2 d\log d}{N}\right)\|\Sigma\|\end{aligned}$$

$\square$

评论. $\forall\epsilon\in(0,1)$, $\sqrt{\frac{K^2 d\log d}{N}}+\frac{K^2 d\log d}{N}$ 中 $\sqrt{\frac{K^2 d\log d}{N}}$ 为主项,

$$\mathbb{E}\|\Sigma_N-\Sigma\|\le\epsilon\|\Sigma\|\implies N\asymp\epsilon^{-2}d\log d.$$