第二十一讲 广义特征值与极小极大原理

一、 广义特征值问题

1、定义：设 A、B 为 n 阶方阵，若存在数λ，使得方程 $\mathbf{Ax}=\lambda\mathbf{Bx}$ 存在非零解，则称λ为 A 相对于 B 的广义特征值，x 为 A 相对于 B 的属于广义特征值λ的特征向量。

- 是标准特征值问题的推广，当 B＝I（单位矩阵）时，广义特征值问题退化为标准特征值问题。
- 特征向量是非零的
- 广义特征值的求解

$$(\mathbf{A}-\lambda\mathbf{B})\mathbf{x}=\mathbf{0} \quad \text{或者} \quad (\lambda\mathbf{B}-\mathbf{A})\mathbf{x}=\mathbf{0}$$

$\rightarrow$ 特征方程 $\det(\mathbf{A}-\lambda\mathbf{B})=0$

求得 λ 后代回原方程 $\mathbf{Ax}=\lambda\mathbf{Bx}$ 可求出 x

本课程进一步考虑 A、B 厄米且为正定矩阵的情况。

2、等价表述

（1）B 正定，$\mathbf{B}^{-1}$ 存在 $\rightarrow \mathbf{B}^{-1}\mathbf{Ax}=\lambda\mathbf{x}$，广义特征值问题化为了标准特征值问题，但一般来说，$\mathbf{B}^{-1}\mathbf{A}$ 一般不再是厄米矩阵。

（2）B 厄米，存在 Cholesky 分解，$\mathbf{B}=\mathbf{GG}^{\mathbf{H}}$，G 满秩

$$\mathbf{Ax}=\lambda\mathbf{GG}^{\mathbf{H}}\mathbf{x} \qquad \text{令 } \mathbf{G}^{\mathbf{H}}\mathbf{x}=\mathbf{y}$$

则 $\mathbf{G}^{-1}\mathbf{A}\left(\mathbf{G}^{\mathbf{H}}\right)^{-1}\mathbf{y}=\lambda\mathbf{y}$ 也成为标准特征值问题。

$\mathbf{G^{-1}A\left(G^H\right)^{-1}}$为厄米矩阵，广义特征值是实数，可以按大小顺序排列$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$，一定存在一组正交归一的特征向量，即存在$\mathbf{y_1, y_2 \cdots, y_n}$满足

$$\mathbf{G^{-1}A\left(G^H\right)^{-1} y_i = \lambda y_i}$$

$$\mathbf{y_i^H y_j = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \ne j \end{cases}}$$

还原为$\mathbf{x_i = \left(G^H\right)^{-1} y_i}$ (i=1,2,⋯,n)，则

$$\mathbf{y}_i^H\mathbf{y}_j=\left(\mathbf{x}_i^H\mathbf{G}\right)\left(\mathbf{G}^H\mathbf{x}_j\right)=\mathbf{x}_i^H\mathbf{B}\mathbf{x}_j=\delta_{ij}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases}$$ （带权正交）

二、 瑞利商

A、B 为 n 阶厄米矩阵，且 B 正定，称 $\mathbf{R}(\mathbf{x})=\dfrac{\mathbf{x}^H\mathbf{A}\mathbf{x}}{\mathbf{x}^H\mathbf{B}\mathbf{x}}(\mathbf{x}\neq\mathbf{0})$ 为 A 相对于 B 的瑞利商。

设 λ_i, $\mathbf{x}_i$ 为 A 相对于 B 的广义特征值和特征向量，且 $\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$

$\mathbf{x}_1,\mathbf{x}_2\cdots,\mathbf{x}_n$ 线性无关，所以，$\forall\mathbf{x}\in\mathbf{C}^n$，存在 $\mathbf{a}_1,\mathbf{a}_2\cdots,\mathbf{a}_n\in\mathbf{C}$，

使得 $$\mathbf{x}=\sum_{i=1}^{n}\mathbf{a}_i\mathbf{x}_i$$

$$\mathbf{x}^H\mathbf{B}\mathbf{x}=\left(\sum_{i=1}^{n}\mathbf{a}_i\mathbf{x}_i\right)^H\mathbf{B}\left(\sum_{j=1}^{n}\mathbf{a}_j\mathbf{x}_j\right)=\sum_{i,j=1}^{n}\overline{\mathbf{a}_i}\mathbf{a}_j\mathbf{x}_i^H\mathbf{B}\mathbf{x}_j=\sum_{i=1}^{n}\left|\mathbf{a}_i\right|^2$$

$$\mathbf{x}^H\mathbf{A}\mathbf{x}=\sum_{i,j=1}^{n}\overline{\mathbf{a}_i}\mathbf{a}_j\mathbf{x}_i^H\mathbf{A}\mathbf{x}_j=\sum_{i,j=1}^{n}\overline{\mathbf{a}_i}\mathbf{a}_j\mathbf{x}_i^H\lambda_i\mathbf{B}\mathbf{x}_j=\sum_{i=1}^{n}\lambda_i\left|\mathbf{a}_i\right|^2$$

$$\therefore \mathbf{R}(\mathbf{x})=\frac{\sum_{i=1}^{n}\lambda_i\left|\mathbf{a}_i\right|^2}{\sum_{i=1}^{n}\left|\mathbf{a}_i\right|^2}$$

以下两点成立

● $\min\limits_{x\neq 0} R(x)=\lambda_1 \qquad \max\limits_{x\neq 0} R(x)=\lambda_n$

证明：$R(x)=\dfrac{x^H A x}{x^H B x}=\dfrac{(kx)^H A(kx)}{(kx)^H B(kx)}$　　k 为非零常数

可取 $k=\dfrac{1}{\|x\|}$，　$\|kx\|=1$

$\therefore \quad R(x)=\left.\dfrac{x^H A x}{x^H B x}\right|_{\|x\|=1}$　　（闭区域）

当 $x=x_1$ 或 $a_i=0(i=2,3,\cdots,n)$ 时，$R(x)=\lambda_1$

$\lambda_i \geq \lambda_1$ $\quad R(x) \geq \lambda_1 \dfrac{\sum_{i=1}^{n} |a_i|^2}{\sum_{i=1}^{n} |a_i|^2} = \lambda_1$

$\therefore \quad \min\limits_{x \neq 0} R(x) = \lambda_1$

另一方面，$\lambda_i \leq \lambda_n$ $\quad R(x) \leq \lambda_n \dfrac{\sum_{i=1}^{n} |a_i|^2}{\sum_{i=1}^{n} |a_i|^2} = \lambda_n$

$\therefore \quad \max\limits_{x \neq 0} R(x) = \lambda_n$

[证毕]

当 B＝I 时，标准特征值问题 $\mathbf{Ax}=\lambda\mathbf{x}$ （$\mathbf{A}^H=\mathbf{A}$）

$$\begin{cases}\lambda_1\le\lambda_2\le\cdots\le\lambda_n\\ \mathbf{x}_i^H\mathbf{x}_j=\delta_{ij}\end{cases}$$

则 $$\min_{(x\ne0)}\frac{\mathbf{x}^H\mathbf{Ax}}{\mathbf{x}^H\mathbf{x}}=\lambda_1 \qquad \max_{(x\ne0)}\frac{\mathbf{x}^H\mathbf{Ax}}{\mathbf{x}^H\mathbf{x}}=\lambda_n$$

进一步分析可得

$$\min_{x\ne0}R(\mathbf{x})\Big|_{a_1=0}=\lambda_2 \qquad \max_{x\ne0}R(\mathbf{x})\Big|_{a_n=0}=\lambda_{n-1}$$

$$\vdots \qquad\qquad \vdots$$

$$\min_{x\ne0}R(\mathbf{x})\Big|_{a_1=a_2=\cdots=a_k=0}=\lambda_{k+1} \qquad \max_{x\ne0}R(\mathbf{x})\Big|_{a_n=a_{n-1}=\cdots=a_{n-k}=0}=\lambda_{n-k-1}$$

定理 1.设$\mathbf{L}=\operatorname{span}\left\{\mathbf{x}_r,\mathbf{x}_{r+1},\cdots,\mathbf{x}_s\right\}$ $\left(\lambda_r\le\lambda_{r+1}\le\cdots\le\lambda_s\right)$,则

$$\min_{\substack{x\ne0\\x\in L}}\mathbf{R}(\mathbf{x})=\lambda_r\qquad\qquad \max_{\substack{x\ne0\\x\in L}}\mathbf{R}(\mathbf{x})=\lambda_s$$

这一结果不便于应用，希望对上述结果进行改造，改造成不依赖于$\mathbf{x}_i$的一种表达方式。

$\mathbf{a}_1=\mathbf{0}$和$\mathbf{a}_n=\mathbf{0}$的情况均对应于 x 在（n-1）维的子空间内变动，x 在 L 中变动是在一个(s-r+1)维子空间中变化。

一般的，x 在$\mathbf{C}^n$的(n-1)维子空间$\mathbf{V}_{n-1}$中变动时，

$$\min_{\substack{x\ne0\\x\in V_{n-1}}}\mathbf{R}(\mathbf{x})\le\lambda_2\qquad\qquad \max_{\substack{x\ne0\\x\in V_{n-1}}}\mathbf{R}(\mathbf{x})\ge\lambda_{n-1}$$

即，对于不同的$\mathbf{V}_{n-1}$， $\mathbf{R}(\mathbf{x})$的最小值及最大值有可能不同，其中各个

最小值中最大者为λ_2，各个最大值中的最小者为λ_{n-1}

$$\max_{V_{n-1}\in C^n}\left[\min_{\substack{x\neq 0\\ x\in V_{n-1}}} R(x)\right]=\lambda_2 \qquad \min_{V_{n-1}\in C^n}\left[\max_{\substack{x\neq 0\\ x\in V_{n-1}}} R(x)\right]=\lambda_{n-1}$$

定理 2. 设V_k是C^n的一个 k 维子空间，则

$$\max_{V_k\in C^n}\left[\min_{\substack{x\neq 0\\ x\in V_k}} R(x)\right]=\lambda_{n-k+1} \qquad \min_{V_k\in C^n}\left[\max_{\substack{x\neq 0\\ x\in V_k}} R(x)\right]=\lambda_k$$

以上两式称为广义特征值的极小极大原理。

● B＝I 时，标准特征值问题同样存在上述关系。

● 矩阵奇异值问题：$\left[\sigma(A)\right]^2=\lambda\left(A^H A\right)$ （非零）

$$R(\mathbf{x}) = \frac{\mathbf{x}^H\left(A^H A\right)\mathbf{x}}{\mathbf{x}^H\mathbf{x}} = \frac{\|A\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2}$$

$$\sigma_{n-k+1} = \max_{V_k \in C^n}\left[\min_{\substack{x \neq 0 \\ x \in V_k}} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}\right]$$

$$\sigma_k = \min_{V_k \in C^n}\left[\max_{\substack{x \neq 0 \\ x \in V_k}} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}\right]$$