六 矩阵特征值问题的数值方法

1. 特征值的估计和扰动

定理 6.1.1 (Gershgorin) $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 则 A 的特征值 $\lambda\in\bigcup_{i=1}^{n}D_i$，其中 D_i 为复平面上以 a_{ii} 为中心 $r_i=\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}|$ 为半径的圆盘，即

$$D_i=\{z\mid |z-a_{ii}|\le r_i,\ z\in\mathbb{C}\},\quad i=1,2,\cdots,n$$

证明：令 $D=\mathrm{diag}\{a_{11},a_{22},\cdots,a_{nn}\}$，设 $\lambda\in\sigma(A)$ x 为相应的特征向量，则有

$$(A-D)x=(\lambda I-D)x$$

不妨设 $\lambda\neq a_{ii}$，则有

$$(\lambda I-D)^{-1}(A-D)x=x$$

$$\Rightarrow\quad \|x\|_\infty\le\|(\lambda I-D)^{-1}(A-D)\|_\infty\|x\|_\infty$$

$$\Rightarrow\quad \|(\lambda I-D)^{-1}(A-D)\|_\infty\ge 1$$

$$\Rightarrow\quad \max_{1\le i\le n}\frac{r_i}{|\lambda-a_{ii}|}\ge 1$$

定理 6.1.2 设定理 6.1.1 中的 n 个圆盘中，有 m 个圆盘构成了一个连通区域 S，且 S 与其余 $n-m$ 个圆盘严格分离，则在 S 中有且只有 A 的 m 个特征值

证明：记 $D=\mathrm{diag}\{a_{11}\cdots a_{nn}\}$，$A_\theta=D+\theta(A-C)$ $\theta\in[0,1]$.

A_θ 的特征多项式的系数是 θ 的多项式，从而 A_θ 的特征值是 θ 的连续函数．令

$D_i(\theta)=\{z \mid |z-a_{ii}|\le\theta r_i\}$，$i=1,2,\cdots,n$

有 $D_i(\theta)\subset D_i$．不失一般性，设前 m 个圆盘连通

即 $S=\bigcup_{i=1}^{m} D_i$，令 $S(\theta)=\bigcup_{i=1}^{m} D_i(\theta)$

$\bar{S}(\theta)=\bigcup_{i=m+1} D_i(\theta)$ 则有 $S(\theta)$ 与 $\bar{S}(\theta)$ 严格分离

利用 A_θ 的特征值关于 θ 连续 易证定理结论

例 $A=\begin{bmatrix}0.9 & 0.01 & 0.12\\ 0.01 & 0.8 & 0.13\\ 0.01 & 0.02 & 0.4\end{bmatrix}$

A 对应的圆盘为 $D_1=\{|z-0.9|\le 0.13\}$

$D_2=\{|z-0.8|\le 0.14\}$，$D_3=\{|z-0.4|\le 0.03\}$

D_3 与 $D_1\cup D_2$ 严格分离 则 D_3 中包含 A 的一个特征值 记为 λ_3 且 λ_3 为实特征值

令 $B=\mathrm{diag}(1,1,0.1)$，则

$B^{-1}AB=\begin{bmatrix}0.9 & 0.01 & 0.012\\ 0.01 & 0.8 & 0.013\\ 0.1 & 0.2 & 0.4\end{bmatrix}$

$\bar{D}_1=\{|z-0.9|\le 0.022\}$，$\bar{D}_2=\{|z-0.8|\le 0.023\}$

$\bar{D}_3=\{|z-0.4|\le 0.3\}$

$\bar{D}_1$，$\bar{D}_2$，$\bar{D}_3$ 均严格分离 则有

$|\lambda_1 - 0.9| \le 0.022$, $|\lambda_2 - 0.8| \le 0.023$

$|\lambda_3 - 0.4| \le 0.03$

定理 6.1.3 设 μ 是矩阵 $A+E \in \mathbb{C}^{n\times n}$ 的一个特征值

且存在 $X \in \mathbb{C}^{n\times n}$，$X^{-1}AX = D = \mathrm{diag}(\lambda_1, \cdots \lambda_n)$ 则

$$\min_{\lambda \in \sigma(A)} |\lambda - \mu| \le \|X^{-1}\|_p \|X\|_p \|E\|_p$$

其中 $\|\cdot\|_p$ 为矩阵的 p 范数，$p=1,2,\infty$

证明：设 $\mu \notin \sigma(A)$ 则有

$$A+E-\mu I = X[D-\mu I + X^{-1}EX]X^{-1}$$

μ 为 $A+E$ 的特征值，所以 $D-\mu I + X^{-1}EX$ 奇异

存在 $y \in \mathbb{C}^n$ 使得

$$(D-\mu I)y = -(X^{-1}EX)y$$

$$\Rightarrow \quad y = -(D-\mu I)^{-1}(X^{-1}EX)y$$

$$\Rightarrow \quad \|(D-\mu I)^{-1}\|_p \|X^{-1}\|_p \|E\|_p \|X\|_p \ge 1$$

$$\Rightarrow \quad \|X^{-1}\|_p \|X\|_p \|E\|_p \ge \|(D-\mu I)^{-1}\|_p^{-1} = \min_{\lambda \in \sigma(A)} |\lambda - \mu|$$

$\|X^{-1}\| \|X\|$ 称为矩阵 A 关于特征值问题的条件数

设 $A \in \mathbb{C}^{n\times n}$，$\lambda$ 为 A 的一个单特征值，x, y 分别为 A 的左右特征向量. 即

$$Ax = \lambda x, \quad y^H A = \lambda y^H$$

设 $\|x\|_2 = \|y\|_2 = 1$. 可以证明存在可微函数 $x(\varepsilon)$, $\lambda(\varepsilon)$ 使得

$$(A+\varepsilon F)\, x(\varepsilon) = \lambda(\varepsilon)\, x(\varepsilon), \quad \|F\|_2 = 1$$

其中 $\lambda(0) = \lambda$, $x(0) = x$. 对 ε 求导得

$$Fx + A\dot{x}(0) = \dot{\lambda}(0)x + \lambda\dot{x}(0)$$

$$\Rightarrow y^H F x + y^H A \dot{x}(0) = \dot{\lambda}(0) y^H x + \lambda y^H \dot{x}(0)$$

$\boxed{y^H A = \lambda y^H}$ $\Rightarrow \dot{\lambda}(0) = -\dfrac{y^H F x}{y^H x}$

$$\Rightarrow |\dot{\lambda}(0)| \le \frac{1}{|y^H x|}$$

$\dfrac{1}{|y^H x|}$ 称为 A关于特征值 λ 的条件数

2. 幂法和逆幂法.

设 $z = (z_1 \cdots z_n)^T \in \mathbb{R}^n$，定义

$$\max(z) = z_i, \quad \text{其中 } |z_i| = \|z\|_\infty$$

幂法为

$$\boxed{\begin{aligned} & v^{(0)} \in \mathbb{R}^n \\ & k=1,2,\cdots \\ & \quad z^{(k)} = A v^{(k-1)} \\ & \quad m_k = \max(z^{(k)}), \\ & \quad v^{(k)} = z^{(k)}/m_k \end{aligned}}$$

定理 6.2.1 $A \in \mathbb{R}^{n\times n}$ 存在n个线性无关的特征向量 $x_1, x_2, \cdots x_n$ 对应的特征值满足

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$$

且 $v^{(0)}$ 在 x_1 方向投影非零 则有

$$\lim_{k\to\infty} v^{(k)} = \frac{x_1}{\max(x_1)}$$

$$\lim_{k\to\infty} m_k = \lambda_1$$

证明: $v^{(k)} = \dfrac{A v^{(k-1)}}{m_k} = \cdots = \dfrac{A^k v^{(0)}}{m_k m_{k-1} \cdots m_1}$

由 $v^{(k)}$ 的最大分量为 1 可得

$$v^{(k)} = \frac{A^k v^{(0)}}{\max(A^k v^{(0)})}$$

设 $v^{(0)} = \sum_{i=1}^{n} \alpha_i x_i$

$$\Rightarrow A^k v^{(0)} = \lambda_1^k \left[\alpha_1 x_1 + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i\right]$$

$$\Rightarrow \lim_{k\to\infty} \frac{A^k v^{(0)}}{\lambda_1^k \alpha_1} = x_1$$

$$\Rightarrow \lim_{k\to\infty} \frac{A^k v^{(0)}}{\max(A^k v^{(0)})} = \frac{x_1}{\max(x_1)}$$

另一方面

$$m_k = \max(z^{(k)}) = \max(A v^{(k-1)})$$

$$= \frac{\max(A^k v^{(0)})}{\max(A^{k-1} v^{(0)})}$$

$$= \lambda_1 \frac{\max\left(\alpha_1 x_1 + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i\right)}{\max\left(\alpha_1 x_1 + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^{k-1} x_i\right)}$$

$$\Rightarrow \lim_{k\to\infty} m_k = \lambda_1$$

- 逆幂法.

$A \in \mathbb{R}^{n\times n}$ 非奇异 有 n 个线性无关的特征向量

$x_1 \cdots x_n$ 对应的特征值 满足

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_{n-1}| > |\lambda_n| > 0$$

将幂法用于 A^{-1} 可得 逆幂法.

$$\boxed{\begin{aligned} A z^{(k)} &= v^{(k-1)} \\ m_k &= \max(z^{(k)}) \\ v^{(k)} &= \frac{z^{(k)}}{m_k} \end{aligned}}$$

在逆幂法基础上选择参数 $q \neq \lambda_i \quad i=1\cdots n$
将幂法用于 $(A-qI)^{-1}$ 可得 原点位移的逆幂法

$$\boxed{\begin{aligned} (A-qI)z^{(k)} &= v^{(k-1)} \\ m_k &= \max(z^{(k)}) \\ v^{(k)} &= \frac{z^{(k)}}{m_k} \end{aligned}}$$

- 收缩方法

设 $|\lambda_1| > |\lambda_2| > |\lambda_3| \geq \cdots \geq |\lambda_n|$ 收缩方法为求出 λ_1 后求 λ_2 的方法

方法 1. 设 λ_1, x_1 已知，计算 Householder 矩阵 H，使得 $Hx_1 = e_1$ （设 $\|x_1\|_2 = 1$）

由 $Ax_1 = \lambda_1 x_1$ 得

$$HAH^{-1}e_1 = \lambda_1 e_1$$

即

$$A_2 = HAH^{-1} = \begin{bmatrix} \lambda_1 & b_1^T \\ 0 & B_2 \end{bmatrix}$$

其中 $B_2 \in \mathbb{R}^{(n-1)\times(n-1)}$ 具有特征值 $\lambda_2 \cdots \lambda_n$

利用幂法求 B_2 的主特征值 λ_2 及特征向量 y_2 满足 $B_2 y_2 = \lambda_2 y_2$

则 A_2 的特征向量为 $z_2 = \begin{bmatrix} \frac{b_1^T y_2}{\lambda_2 - \lambda_1} \\ y_2 \end{bmatrix}$

A 相应的特征向量为 $H^{-1}z_2$

- 定理 6.2.2 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$ 特征值为 $\lambda_1\cdots\lambda_n$，$x_1\cdots x_n$ 为相应的特征向量. 且 λ_1 的重数为 1，$v\in\mathbb{R}^n$ 满足 $v^Tx_1=1$ 则

$$B=A-\lambda_1x_1v^T$$

的特征值为 $0,\lambda_2\cdots\lambda_n$ 对应特征向量为 $x_1,y_2\cdots y_n$ 且

$$x_i=(\lambda_i-\lambda_1)y_i+\lambda_1(v^Ty_i)x_1,\quad i=2\cdots n$$

- Wielandt 收缩方法、

设 $x_1=(x_{11}\cdots x_{1n})^T$ 令

$$v=\frac{1}{\lambda_1x_{1i}}(a_{i1}\cdots a_{in})^T$$

x_{1i} 为 x_1 的非零分量. 则有

$$v^Tx_1=\frac{1}{\lambda_1x_{1i}}\sum_{j=1}^{n}a_{ij}x_{1j}=1$$

并且

$$B=A-\lambda_1x_1v^T$$

第 i 行为零

若 y 为 B 的特征向量，且 $\lambda\neq0$ 则

$$By=\lambda y$$

$\Rightarrow$ $y_i=0$，$y=(y_1\cdots y_n)$

可删去 B 的第 i 行第 i 列得到 $B' \in R^{(n-1)\times(n-1)}$.

采用幂法计算 B' 的主特征值 λ_2 及特征向量 y_2'. 在 y_2' 第 i 个分量位置插入 0 得到 y_2 再利用定理 6.2.2 中的公式可以得到 x_2