特征值问题的数值方法

包承龙

丘成桐数学科学中心

本章研究对象及目标

对于矩阵$\mathbf{A}\in\mathbb{R}^{n\times n}$, 求特征值$\lambda\in\mathbb{C}$及$\mathbf{x}\in\mathbb{C}^n$,满足

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

特征多项式

$$p(\lambda)=\det(\lambda\mathbf{I}-\mathbf{A})=\lambda^n+a_{n-1}\lambda^{n-1}+\cdots+a_1+a_0:=q(\lambda)$$

$q(\lambda)$是矩阵 $\underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ & 1 & \cdots & 0 & -a_2 \\ & & \ddots & \vdots & \vdots \\ & & & 1 & -a_{n-1} \end{bmatrix}}_{\text{Hessenberg矩阵–QR方法求解}}$ 的特征多项式。

目录

特征值估计

设$\sigma(\mathbf{A})$为$\mathbf{A}$的所有特征值集合：在复平面中，$\sigma(\mathbf{A})$位于半径为$\|\mathbf{A}\|$的圆盘内。

定理 (Gershgorin定理)

设$\mathbf{A} = [a_{ij}] \in \mathbb{C}^{n\times n}$，则对任意$\lambda \in \sigma(\mathbf{A})$，有$\lambda \in \cup_{i=1}^{n} D_i$，其中$D_i$是复平面上以$a_{ij}$为中心，$r_i = \sum_{j=1,j\neq i} |a_{ij}|$为半径的圆盘，即 行和

$$D_i = \{z | |z - a_{ii}| \le r_i, z \in \mathbb{C}\} \tag{1}$$

设$\mathbf{D} = \mathrm{Diag}(a_{11}, a_{22}, \ldots, a_{nn})$，有

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

定理

在由(1)所表示的n个圆盘中，有m个圆盘构成一个连通区域S，且与其余$n-m$个圆盘严格分离，则S中恰有m个特征值。

若$m=1$,则说明每个孤立的圆盘内恰有一个特征值。

取$\mathbf{B}=\mathrm{Diag}(b_1,b_2,\ldots,b_n)$, 其中$b_i\neq 0$, 则$\mathbf{A}$与$\mathbf{B}^{-1}\mathbf{A}\mathbf{B}$相似，且$\mathbf{B}^{-1}\mathbf{A}\mathbf{B}$的Gershgorin圆盘的半径为

$$\sum_{j=1,j\neq i}^{n}\left|\frac{a_{ij}b_j}{b_i}\right|$$

例：$\mathbf{A}=\begin{bmatrix}0.9 & 0.01 & 0.12\\ 0.01 & 0.8 & 0.13\\ 0.01 & 0.02 & 0.4\end{bmatrix}\Rightarrow|\lambda_1-0.9|\leq 0.022, |\lambda_2-0.8|\leq 0.023, |\lambda_3-0.4|\leq 0.03$

特征值扰动

定理

设$\mu \in \sigma(\mathbf{A}+\mathbf{E})$, 且 $\mathbf{X}^{-1}\mathbf{A}\mathbf{X} = \mathbf{D} = \mathrm{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$,$\mathbf{X} \in \mathbb{C}^{n\times n}$,则

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

设$\mu \notin \sigma(\mathbf{A})$, 则有

$$\mathbf{A}+\mathbf{E}-\mu\mathbf{I} = \mathbf{X}[\mathbf{D}-\mu\mathbf{I}+\mathbf{X}^{-1}\mathbf{E}\mathbf{X}]\mathbf{X}^{-1} \Rightarrow \mathbf{D}-\mu\mathbf{I}+\mathbf{X}^{-1}\mathbf{E}\mathbf{X}\cancel{\text{非}}\text{奇异}$$

存在非零向量y，使得 $\mathbf{y} = -(\mathbf{D}-\mu\mathbf{I})^{-1}(\mathbf{X}^{-1}\mathbf{E}\mathbf{X})\mathbf{y}$

$\|\mathbf{X}\|\|\mathbf{X}^{-1}\|$: 特征值问题的条件数

$\mathbf{A}$为正规矩阵($\mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A}$), 则可取X为酉矩阵，$\|\mathbf{X}\|_2\|\mathbf{X}^{-1}\|_2 = 1$。

若$\mathbf{A} \in \mathbb{C}^{n\times n}$为非正规矩阵，则$\mathbf{A}$可能对某些特征值敏感。

设$\lambda \in \sigma(\mathbf{A})$, $\mathbf{x}$, $\mathbf{y}$分别为右、左特征向量：

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}, \quad \mathbf{y}^H\mathbf{A} = \lambda\mathbf{y}^H$$

存在可微函数$\mathbf{x}(\epsilon)$与$\lambda(\epsilon)$满足$\mathbf{x}(0) = \mathbf{x}$, $\lambda(0) = \lambda$,

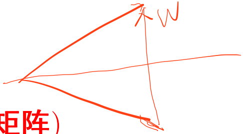
$$\begin{aligned}
& (\mathbf{A} + \epsilon\mathbf{F})\mathbf{x}(\epsilon) = \lambda(\epsilon)\mathbf{x}(\epsilon), \quad \|\mathbf{F}\|_2 = 1 \\
\Rightarrow \quad & \mathbf{A}\dot{\mathbf{x}}(0) + \mathbf{F}\mathbf{x} = \dot{\lambda}(0)\mathbf{x} + \lambda\dot{\mathbf{x}}(0) \\
\Rightarrow \quad & |\dot{\lambda}(0)| = \frac{\mathbf{y}^H\mathbf{F}\mathbf{x}}{\mathbf{y}^H\mathbf{x}} \le \underbrace{\frac{1}{|\mathbf{y}^H\mathbf{x}|}}_{\frac{1}{S(\lambda)}}
\end{aligned}$$

$\frac{1}{|y^H x|}$

$S(\lambda)$称为关于特征值λ的条件数。

目录

Householder变换



设$\mathbf{w} \in \mathbb{R}^n$且$\|\mathbf{w}\| = 1$，称矩阵

$$\mathbf{P} = \mathbf{I} - 2\mathbf{w}\mathbf{w}^\top, \quad \text{(Householder 矩阵)}$$

- $\mathbf{P}$对称，且正交 $\quad \Rightarrow \|\mathbf{P}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$
- 设S为过原点且与$\mathbf{w}$垂直的平面，$\forall \mathbf{v} \in \mathbb{R}^n$，存在$\mathbf{v}_1 \in S$，$\mathbf{v}_2 \in S^\perp$，满足$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$，有

$$\mathbf{P}\mathbf{v}_1 = \mathbf{v}_1, \quad \mathbf{P}\mathbf{v}_2 = -\mathbf{v}_2 \quad \Rightarrow \quad \mathbf{P}\mathbf{v} = \underbrace{\mathbf{v}_1 - \mathbf{v}_2}_{\text{初等镜面反射}}$$

- 任给$\mathbf{x} \neq \mathbf{y}$且$\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2$，存在Household矩阵$\mathbf{P}$，使得$\mathbf{y} = \mathbf{P}\mathbf{x}$

Givens变换

$$\begin{pmatrix} y_i \\ y_k \end{pmatrix} = \begin{pmatrix} c & s \\ -s & c \end{pmatrix}\begin{pmatrix} x_i \\ x_k \end{pmatrix}$$

记$c=\cos\theta$，$s=\sin\theta$，旋转矩阵$\mathbf{J}=\begin{bmatrix} c & s \\ -s & c \end{bmatrix}$

$\mathbf{J}(i,k,\theta)\in\mathbb{R}^{n\times n}$是旋转矩阵，且有$\mathbf{y}=\mathbf{J}(i,k,\theta)\mathbf{x}$满足

$$y_j = x_j, j\neq i,k, \quad y_i = cx_i + sx_k, \quad y_k = -sx_i + cx_k$$

不妨设$k>i$。可以通过选取c，s，使得$y_k=0$ 。

- 若$x_k=0$，则取$c=1$，$s=0$
- 若$x_k\neq 0$，则有

$$\begin{cases} t=\frac{x_i}{x_k}, s=\operatorname{sign}(x_k)(1+t^2)^{-\frac{1}{2}}, c=st, & |x_k|\geq|x_i|\text{时}, \\ t=\frac{x_k}{x_i}, c=\operatorname{sign}(x_k)(1+t^2)^{-\frac{1}{2}}, s=ct, & |x_k|<|x_i|\text{时}. \end{cases}$$

QR因式分解

定理

设$\mathbf{A} \in \mathbb{R}^{n\times n}$, 则存在正交矩阵$\mathbf{P}$, 使得 $\mathbf{PA} = \mathbf{R}$, 其中$\mathbf{R}$是上三角矩阵。

定理

设$\mathbf{A} \in \mathbb{R}^{n\times n}$，存在正交矩阵$\mathbf{Q}$与上三角矩阵$\mathbf{R}$，使得

$$\mathbf{A} = \mathbf{QR}$$

若$\mathbf{A}$是非奇异的,且若规定$\mathbf{R}$的对角元素大于*0*，则这种分解是唯一的。

QR分解计算：Householder变换、Givens变换、Gram-Schmidt正交化。

Shur分解

定理

设$\mathbf{A} \in \mathbb{C}^{n\times n}$, 则存在酉矩阵$\mathbf{Q}$,使得$\mathbf{Q}^H\mathbf{A}\mathbf{Q} = \mathbf{T}$为上三角矩阵。

定理 (实Shur分解定理)

设$\mathbf{A} \in \mathbb{R}^{n\times n}$, 则存在正交矩阵$\mathbf{Q} \in \mathbb{R}^{n\times n}$，使得

$$\mathbf{Q}^\top\mathbf{A}\mathbf{Q} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \cdots & \mathbf{R}_{1n} \\ & \mathbf{R}_{22} & \cdots & \mathbf{R}_{2n} \\ & & \ddots & \vdots \\ & & & \mathbf{R}_{nn} \end{bmatrix}$$

其中$\mathbf{R}_{ii}$是1×1或者2×2的，为$\mathbf{A}$的特征值或者一对共轭的复特征值。

目录

幂迭代法

设$\mathbf{A} \in \mathbb{R}^{n\times n}$且可对角化，则有$n$个线性无关的特征向量$\mathbf{x}_i$满足

$$\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i, \quad i = 1, 2, \ldots, n$$

对任意的$\mathbf{x}^{(0)}$，存在$\alpha_i \in \mathbb{R}$，$i = 1, 2, \ldots, n$ 使得

$$\mathbf{x}^{(0)} = \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \cdots + \alpha_n \mathbf{x}_n$$

若$\alpha_1 \neq 0$，有

$$\mathbf{A}^k \mathbf{x}^{(0)} = \lambda_1^k \left[\alpha_1 \mathbf{x}_1 + \sum_{i=2}^{n} \underbrace{\alpha_i \left(\frac{\lambda_i}{\lambda_1} \right)^k \mathbf{x}_i}_{\to 0, k\to\infty, \text{若} |\lambda_i| < |\lambda_1|} \right] {\color{red}\approx \lambda_1^k \alpha_1 \mathbf{x}_1}$$

给定$\mathbf{v}^{(0)} \in \mathbb{R}^n$, 幂法迭代公式为

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

其中$\max(\mathbf{z}) = z_i, \ |z_i| = \|\mathbf{z}\|_\infty$.

定理

若$\mathbf{A}$有n个线性无关的特征向量，且对应的特征值满足

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

若$\mathbf{v}^{(0)}$在$\mathbf{x}_1$方向的投影非零，则有

$$\lim_{k\to\infty} \mathbf{v}^{(k)} = \frac{\mathbf{x}_1}{\max(\mathbf{x}_1)}, \quad \lim_{k\to\infty} m_k = \lambda_1$$

λ_1称为$\mathbf{A}$的主特征值，$\mathbf{x}_1$称为主特征向量。

幂迭代法的收敛速度

$$|m_k - \lambda_1| \approx K \left|\frac{\lambda_2}{\lambda_1}\right|^k, \quad \frac{|m_{k+1} - \lambda_1|}{|m_k - \lambda_1|} \approx \left|\frac{\lambda_2}{\lambda_1}\right|$$

对于一般的矩阵$\mathbf{A}$, 幂迭代法的分析则较为复杂。

加速技术：类似于Aitken方法，构造$\bar{\lambda}_1^{(k)}$

$$\bar{\lambda}_1^{(k)} = \frac{m_k m_{k+2} - m_{k+1}^2}{m_{k+2} - 2m_{k+1} + m_k}$$

逆幂迭代法

将幂迭代法作用于$\mathbf{A}^{-1}$：

$$\underbrace{\mathbf{z}^{(k)} = \mathbf{A}^{-1}\mathbf{v}^{(k-1)}}_{\text{求解}\mathbf{A}\mathbf{z}^{(k)}=\mathbf{v}^{(k-1)}}, \quad m_k = \max(\mathbf{z}^{(k)}), \quad \mathbf{v}^{(k)} = \frac{\mathbf{z}^{(k)}}{m_k}$$

在幂迭代定理的条件下，若

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

则有

$$\lim_{k\to\infty} \mathbf{v}^{(k)} = \frac{\mathbf{x}_n}{\max(\mathbf{x}_n)}, \quad \lim_{k\to\infty} m_k = \lim_{k\to\infty} \frac{1}{\lambda_n}\left[1 + O\left(\left|\frac{\lambda_n}{\lambda_1}\right|^k\right)\right] = \frac{1}{\lambda_n}$$

原点位移逆幂迭代

如何通过逆幂迭代，找到特征值离q最近的特征值？

$$\sigma(\mathbf{A}-q\mathbf{I})=\{\lambda-q|\lambda\in\sigma(\mathbf{A})\}$$

离q最近的特征值等价于$\mathbf{A}-q\mathbf{I}$模长最小的特征值，对矩阵$(\mathbf{A}-q\mathbf{I})^{-1}$做逆幂迭代，有

$$\underbrace{\mathbf{z}^{(k)}=(\mathbf{A}-q\mathbf{I})^{-1}\mathbf{v}^{(k-1)}}_{\text{求解}(\mathbf{A}-q\mathbf{I})\mathbf{z}^{(k)}=\mathbf{v}^{(k-1)}},\quad m_k=\max(\mathbf{z}^{(k)}),\quad \mathbf{v}^{(k)}=\frac{\mathbf{z}^{(k)}}{m_k}$$

$$\Rightarrow \lim_{k\to\infty}\mathbf{v}^{(k)}=\frac{\mathbf{x}_i}{\max(\mathbf{x}_i)},\quad \lim_{k\to\infty}m_k=\frac{1}{\lambda_i}$$

q值的估计可以从Gerschgorin圆盘定理得到。

收缩方法

设$|\lambda_1| > |\lambda_2| > |\lambda_3| \geq \cdots \geq |\lambda_n|$，若已经计算完$\lambda_1$，如何计算$\lambda_2$？

存在Householder矩阵$\mathbf{H}$，使得 $\mathbf{H}\mathbf{x}_1 = k\mathbf{e}_1$，其中$k = \|\mathbf{x}_1\|_2$，则有

$$\mathbf{H}\mathbf{A}\mathbf{H}^{-1}\mathbf{e}_1 = \lambda\mathbf{e}_1 \quad \Rightarrow \quad \mathbf{A}_2 = \mathbf{H}\mathbf{A}\mathbf{H}^{-1} = \begin{bmatrix} \lambda_1 & \mathbf{b}_1^\top \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix}$$

且$\sigma(\mathbf{B}_2) = \{\lambda_2, \lambda_3, \ldots, \lambda_n\}$，对$\mathbf{B}_2 \in \mathbb{R}^{n-1 \times n-1}$做幂迭代，有 $\mathbf{B}_2\mathbf{y}_2 = \lambda_2\mathbf{y}_2$. 设$\mathbf{z}_2 = [\alpha, \mathbf{y}^\top]^\top$，由

$$\mathbf{A}_2\mathbf{z}_2 = \lambda_2\mathbf{z}_2 \Rightarrow \begin{cases} \lambda_1\alpha + \mathbf{b}_1^\top\mathbf{y} = \lambda\alpha \\ \mathbf{B}_2\mathbf{y} = \lambda_2\mathbf{y} \end{cases} \Rightarrow \mathbf{y} = \mathbf{y}_2, \alpha = \frac{\mathbf{b}_1^\top y_2}{\lambda_2 - \lambda_1}$$

有$\mathbf{x}_2 = \mathbf{H}^{-1}\mathbf{z}_2$是$\mathbf{A}$对应$\lambda_2$的特征向量。

设λ_1的重数为1，向量$\mathbf{v}$满足$\mathbf{v}^\top\mathbf{x}_1=1$，则矩阵$\mathbf{B}=\mathbf{A}-\lambda_1\mathbf{x}_1\mathbf{v}^\top$ 有特征值$0,\lambda_2,\lambda_3,\cdots,\lambda_n$，对应的特征向量为$\mathbf{x}_1,\mathbf{y}_2,\mathbf{y}_3,\cdots,\mathbf{y}_n$，其中$\mathbf{x}_i$与$\mathbf{y}_i$满足如下关系

$$\mathbf{x}_i=(\lambda_i-\lambda_1)\mathbf{y}_i+\lambda_1(\mathbf{v}^\top\mathbf{y}_i)\mathbf{x}_1, i=2,3,\ldots,n$$

Wielandt收缩方法：

$$\mathbf{v}=\frac{1}{\lambda_1 x_{1i}}(a_{i1},a_{i2},\ldots,a_{in})^\top,\quad \mathbf{x}_1=(x_{11},x_{12},\ldots,x_{1n})^\top$$

有

$$\mathbf{v}^\top\mathbf{x}_1=\frac{1}{\lambda_1 x_{1i}}\sum_{i=1}^{n}a_{ij}x_{1j}\Rightarrow \mathbf{B}\text{第}i\text{行元素全是零}$$

则可以删去B的第i行与第i列。

目录

QR方法

令$\mathbf{A}_1 = \mathbf{A}$，有如下迭代：

$$\mathbf{A}_k = \mathbf{Q}_k\mathbf{R}_k \ (QR\text{分解}), \quad \mathbf{A}_{k+1} = \mathbf{R}_k\mathbf{Q}_k \tag{2}$$

QR迭代(2)有如下性质：

- $\mathbf{A}_{k+1} = \mathbf{Q}_k^\top \mathbf{A}_k \mathbf{Q}_k$
- 设$\bar{\mathbf{Q}}_k = \mathbf{Q}_1\mathbf{Q}_2\cdots\mathbf{Q}_k$，$\bar{\mathbf{R}}_k = \mathbf{R}_k\mathbf{R}_{k-1}\cdots\mathbf{R}_1$，有

$$\mathbf{A}_{k+1} = \bar{\mathbf{Q}}_k^\top \mathbf{A}\bar{\mathbf{Q}}_k, \quad \mathbf{A}^k = \bar{\mathbf{Q}}_k\bar{\mathbf{R}}_k$$

基本收敛到上三角矩阵：矩阵序列$\{\mathbf{A}_k\}$对角元素均收敛且严格下三角元素收敛到0

引理

设$\mathbf{M}_k = \mathbf{Q}_k\mathbf{R}_k$, 其中$\mathbf{Q}_k$正交，$\mathbf{R}_k$是正对角元素的上三角矩阵。若$\mathbf{M}_k \to \mathbf{I}$, 则有$\mathbf{Q}_k \to \mathbf{I}$, $\mathbf{R}_k \to \mathbf{I}$

定理

设$\mathbf{A}$的特征值满足$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| > 0$, 矩阵$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$的每一列是所对应的特征向量。设$\mathbf{X}^{-1} = \mathbf{L}\mathbf{U}$, 其中$\mathbf{L}$为单位下三角，$\mathbf{U}$为上三角，则*QR*方法产生的序列$\{\mathbf{A}_k\}$基本收敛到上三角矩阵，且有

$$\lim_{k\to\infty} a_{ii}^{(k)} = \lambda_i, \quad i = 1, 2, \ldots, n$$

一般情况下的QR方法收敛性较为复杂。

上Hessenberg矩阵

上Hessenberg矩阵： $b_{ij} = 0, \forall i > j+1$. 若$b_{i,i-1} \neq 0$, 则$\mathbf{B}$为不可约的上Hessenberg矩阵。

$$\mathbf{B} = \begin{bmatrix} \star & \star & \cdots & \star & \star \\ \star & \star & \cdots & \star & \star \\ & \star & \cdots & \star & \star \\ & & \ddots & \vdots & \vdots \\ & & & \star & \star \end{bmatrix}$$

定理

设$\mathbf{A} \in \mathbb{R}^{n\times n}$，则存在正交矩阵$\mathbf{Q} \in \mathbb{R}^{n\times n}$，有$\mathbf{B} = \mathbf{Q}^\top \mathbf{A} \mathbf{Q}$为一个上*Hessenberg*矩阵。

上Hessenberg矩阵形式不唯一，但具有如下性质。

定理 (隐式Q定理)

设$\mathbf{Q}$, $\mathbf{V}$均为正交矩阵，且有$\mathbf{Q}^\top\mathbf{A}\mathbf{Q}=\mathbf{H}$和$\mathbf{V}^\top\mathbf{A}\mathbf{V}=\mathbf{G}$均为上*Hessenberg*矩阵。记$k$为$h_{k+1,k}=0$为最小整数，当$\mathbf{H}$不可约时规定$k=n$。如果$\mathbf{q}_1=\mathbf{v}_1$, 则对$i=2,\ldots,k$有

$$\mathbf{q}_i=\pm\mathbf{v}_i,\quad |h_{i,i-1}|=|g_{i,i-1}|$$

若$k<n$，则$g_{k+1,k}=0$

$\mathbf{G}$与$\mathbf{H}$“本质上相同”，即$\mathbf{G}=\mathbf{D}^{-1}\mathbf{H}\mathbf{D}$, $\mathbf{D}$为对角矩阵且对角元素为1或者-1

上Hessenberg矩阵的QR方法

用$n-1$次Givens变换将$\mathbf{H}$下三角元素变成0，即

$$\mathbf{U}^\top = \mathbf{J}(n-1,n,\theta_{n-1})\cdots\mathbf{J}(1,2,\theta_1), \qquad \mathbf{U}^\top\mathbf{H} = \mathbf{R}$$

可以验证，

$$\mathbf{H}_2 = \mathbf{R}\mathbf{U}$$

仍为上Hessenberg矩阵，即QR迭代法保持了上Hessenberg矩阵的结构形式。

带原点位移的QR方法

对$\mathbf{H}$的QR迭代方法的收敛速度取决于$\max\left|\frac{\lambda_{j+1}}{\lambda_j}\right|$, 设$\mathbf{H}_1=\mathbf{H}$，有

$$\mathbf{H}_k-\mu\mathbf{I}=\mathbf{U}_k\mathbf{R}_k\ (\mathbf{H}_k-\mu\mathbf{I}\text{的QR分解}),\quad \mathbf{H}_{k+1}=\mathbf{R}_k\mathbf{U}_k+\mu\mathbf{I}$$

上述迭代有如下关系:

$$\begin{aligned} &\mathbf{H}_{k+1}=\mathbf{R}_k\mathbf{U}_k+\mu\mathbf{I}=\mathbf{U}_k^\top(\mathbf{U}_k\mathbf{R}_k+\mu\mathbf{I})\mathbf{U}_k=\mathbf{U}_k^\top\mathbf{H}_k\mathbf{U}_k\\ \Rightarrow\quad &\mathbf{H}_k\sim\mathbf{H},\ \forall k\in\mathbb{N}\end{aligned}$$

若$\mu\in\sigma(\mathbf{A})$, 则$\mathbf{H}_k-\mu\mathbf{I}$奇异，从而$\mathbf{R}_k$对角线上必有零元素

可以证明$|r_{ii}^{(k)}|\geq|h_{i+1,i}^{(k)}|$, $i=1,2,\cdots,n-1$. 若$\mathbf{H}_k$不可约，则有$r_{nn}^{(k)}=0$, 从而$\mathbf{H}_{k+1}$的最后一行为$(0,\cdots,0,\mu)$

- 第一种带原点位移的QR迭代

$$\text{取}\mu_k = h_{nn}^{(k)}, \quad \mathbf{H}_k - \mu_k \mathbf{I} = \mathbf{U}_k \mathbf{R}_k, \quad \mathbf{H}_{k+1} = \mathbf{R}_k \mathbf{U}_k + \mu_k \mathbf{I}$$

取$\bar{\mathbf{U}}_k = \mathbf{U}_1 \mathbf{U}_2 \cdots \mathbf{U}_k$, $\bar{\mathbf{R}}_k = \mathbf{R}_k \mathbf{R}_{k-1} \cdots \mathbf{R}_1$,有

$$\mathbf{H}_{k+1} = \bar{\mathbf{U}}_k^\top \mathbf{H}_1 \bar{\mathbf{U}}_k, \quad (\mathbf{H}_k\text{相似于}\mathbf{H}_1)$$
$$(\mathbf{H}_1 - \mu_{k+1}\mathbf{I})(\mathbf{H}_1 - \mu_k \mathbf{I}) \cdots (\mathbf{H}_1 - \mu_1 \mathbf{I}) = \bar{\mathbf{U}}_{k+1} \mathbf{R}_{k+1}$$

- 第二种带原点位移的QR迭代

$$\mu_k \in \sigma\left(\begin{bmatrix} h_{n-1,n-1}^{(k)} & h_{n-1,n}^{(k)} \\ h_{n,n-1}^{(k)} & h_{nn}^{(k)} \end{bmatrix}\right)$$

且μ_k更接近于$h_{nn}^{(k)}$

- 无法处理复共轭特征根

双重步QR方法

$$\mathbf{G} = \begin{bmatrix} h_{n-1,n-1} & h_{n-1,n} \\ h_{n,n-1} & h_{nn} \end{bmatrix}, \quad \mu_1, \mu_2 \in \sigma(\mathbf{G})$$

若$\mu_1 = \bar{\mu}_2 \in \mathbb{C}$,

$$\begin{aligned} &\mathbf{H} - \mu_1\mathbf{I} = \mathbf{U}_1\mathbf{R}_1 \ (\mathbf{H} - \mu_1\mathbf{I}\text{复的QR分解}), \quad \mathbf{H}_1 = \mathbf{R}_1\mathbf{U}_1 + \mu_1\mathbf{I} \\ &\mathbf{H}_1 - \mu_2\mathbf{I} = \mathbf{U}_2\mathbf{R}_2 \ (\mathbf{H}_1 - \mu_2\mathbf{I}\text{复的QR分解}), \quad \mathbf{H}_2 = \mathbf{R}_2\mathbf{U}_2 + \mu_2\mathbf{I} \end{aligned} \tag{3}$$

设$s = \mu_1 + \mu_2 \in \mathbb{R}$, $t = \mu_1\mu_2 \in \mathbb{R}$, 则(3)等价于

$$\text{计算}\mathbf{M} = \mathbf{H}^2 - s\mathbf{H} + t\mathbf{I}, \quad \mathbf{M} = \mathbf{Z}\mathbf{R}(\text{实QR分解}) \quad , \mathbf{H}_2 = \mathbf{Z}^\top\mathbf{H}\mathbf{Z} \tag{4}$$

定理：若$\mathbf{H}$为上Hessenberg矩阵，$\mu_1, \mu_2 \notin \sigma(\mathbf{H})$, 且$\mathbf{H}_2$由(4)得到，

则$\mathbf{H}_2$为上Hessenberg矩阵。

目录

对称矩阵的性质

若$\mathbf{A} \in \mathbb{R}^{n\times n}$且有$\mathbf{A} = \mathbf{A}^\top$

性质： 存在正交矩阵$\mathbf{Q} \in \mathbb{R}^{n\times n}$，使得

$$\mathbf{Q}^\top \mathbf{A}\mathbf{Q} = \mathrm{Diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

对称矩阵可正交相似于对角阵

定理

设$\mathbf{A} = \mathbf{A}^\top$, 且$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, 则有

$$\lambda_i = \max_{\dim(\mathbf{W})=i} \min_{\mathbf{x}\in\mathbf{W},\mathbf{x}\neq\mathbf{0}} \frac{(\mathbf{A}\mathbf{x},\mathbf{x})}{(\mathbf{x},\mathbf{x})} = \min_{\mathbf{x}\in\mathrm{Span}\{\mathbf{x}_1,\dots,\mathbf{x}_i\},\mathbf{x}\neq\mathbf{0}} \frac{(\mathbf{A}\mathbf{x},\mathbf{x})}{(\mathbf{x},\mathbf{x})}$$

$$\lambda_i = \min_{\dim(\mathbf{W})=n-i+1} \max_{\mathbf{x}\in\mathbf{W},\mathbf{x}\neq\mathbf{0}} \frac{(\mathbf{A}\mathbf{x},\mathbf{x})}{(\mathbf{x},\mathbf{x})} = \max_{\mathbf{x}\in\mathrm{Span}\{\mathbf{x}_i,\dots,\mathbf{x}_n\},\mathbf{x}\neq\mathbf{0}} \frac{(\mathbf{A}\mathbf{x},\mathbf{x})}{(\mathbf{x},\mathbf{x})}$$

定理 (扰动性质)

设$\mathbf{A}$, $\mathbf{E}$为对称矩阵，有

$$\sigma(\mathbf{A}) = \{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n\}$$
$$\sigma(\mathbf{E}) = \{v_1 \geq v_2 \geq \cdots \geq v_n\}$$
$$\sigma(\mathbf{A}+\mathbf{E}) = \{\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n\}$$

则有$\lambda_i + v_n \leq \mu_i \leq \lambda_i + v_1$, $\forall i = 1, 2, \ldots, n$

推论：$|\lambda_i - \mu_i| \leq \rho(\mathbf{E}) \leq \|\mathbf{E}\|_2$, $\forall i = 1, 2, \ldots, n$

定理

设$\mathbf{A} = \mathbf{A}^\top$, $\mathbf{B} = \mathbf{P}^\top\mathbf{A}\mathbf{P}$, 其中$\mathbf{P}$是正交矩阵，则有$\|\mathbf{B}\|_F = \|\mathbf{A}\|_F$

Rayleigh商迭代法

$\mathbf{A}\in\mathbb{C}^{n\times n}$， $\mathbf{x}\in\mathbb{C}^n$, $\mathbf{x}\neq\mathbf{0}$, 定义

$$\mathbf{R}(\mathbf{x})=\frac{(\mathbf{A}\mathbf{x},\mathbf{x})}{(\mathbf{x},\mathbf{x})},\quad \text{(Rayleigh商)}$$

性质： $\mathbf{R}(\mathbf{x})=\overset{\text{arg min}}{\cancel{\min_{\lambda\in\mathbb{C}}}}\|(\mathbf{A}-\lambda\mathbf{I})\mathbf{x}\|_2$ (由于 $=\arg\min\|(A-\lambda I)x\|_2^2$)

因此，若$\mathbf{v}$近似一个特征向量，则$\mathbf{R}(\mathbf{v})$近似相应的特征值。

Rayleigh商迭代法： 给定$\mathbf{v}^{(0)}\in\mathbb{R}^n$且$\|\mathbf{v}^{(0)}\|_2=1$

$$\mu_k=\mathbf{R}(\mathbf{v}^{(k)}),\quad 解(\mathbf{A}-\mu_k\mathbf{I})\mathbf{y}^{(k+1)}=\mathbf{v}^{(k)},\quad \mathbf{v}^{(k+1)}=\mathbf{y}^{(k+1)}/\|\mathbf{y}^{(k+1)}\|_2 \tag{5}$$

可以证明，(5)是三次收敛的。

古典Jacobi方法

利用Givens变换（Jacobi旋转矩阵）,使得off(A)变为0。 A对称的情况

$$\mathrm{off}(\mathbf{A}) = \sum_{\mathrm{i,j=1,i\neq j}}^{\mathrm{n}} \mathrm{a}_{\mathrm{ij}}^2 = \|\mathbf{A}\|_{\mathrm{F}}^2 - \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{a}_{\mathrm{ii}}^2$$

$J^{-1} = J^T$

考虑$\mathbf{J} = \mathbf{J}(p, q, \theta)$, 则$\mathbf{B} = \mathbf{JAJ}^{-1}$有

只考虑 a_{pp} a_{qq} . a_{pq}
b_{pp} b_{pq} b_{qq}

$$\begin{aligned} b_{pp} &= a_{pp}c^2 + a_{qq}s^2 + a_{pq}\sin 2\theta, \\ b_{qq} &= a_{pp}s^2 + a_{qq}c^2 - a_{pq}\sin 2\theta \\ b_{pq} &= b_{qp} = \frac{1}{2}(a_{qq} - a_{pp})\sin 2\theta + a_{pq}\cos 2\theta \end{aligned} \Rightarrow \begin{aligned} &b_{pp}^2 + b_{qq}^2 + 2b_{pq}^2 \\ =&a_{pp}^2 + a_{qq}^2 + 2a_{pq}^2 \end{aligned}$$

$$\begin{aligned} b_{ip} &= b_{pi} = a_{ip}c + a_{iq}s, \quad i \neq p, q \\ b_{iq} &= b_{qi} = -a_{ip}s + a_{iq}c, \quad i \neq p, q \Rightarrow b_{ip}^2 + b_{iq}^2 = a_{ip}^2 + a_{iq}^2 \\ b_{ij} &= a_{ij}, \quad \text{otherwise} \end{aligned}$$

$$
\begin{aligned}
\mathrm{off}(\mathbf{B}) &= \|\mathbf{B}\|_F^2 - \sum_{i=1}^{n} b_{ii}^2 = \|\mathbf{A}\|_F^2 - \sum_{i\neq p,q}^{n} a_{ii}^2 - (b_{pp}^2 + b_{qq}^2) \\
&= \|\mathbf{A}\|_F^2 - \sum_{i=1}^{n} a_{ii}^2 + (a_{pp}^2 + a_{qq}^2 - b_{pp}^2 - b_{qq}^2) \\
&= \mathrm{off}(\mathbf{A}) - 2\mathrm{a}_{\mathrm{pq}}^2 + 2\mathrm{b}_{\mathrm{pq}}^2
\end{aligned}
$$

选取合适的θ，使得

$$
b_{pq} = \frac{1}{2}(a_{qq} - a_{pp})\sin 2\theta + a_{pq}\cos 2\theta = 0 \Rightarrow \cot 2\theta = \frac{a_{pp} - a_{qq}}{2a_{pq}} = \tau
$$

令$t = \tan\theta$, 则t满足

$$
\begin{aligned}
t^2 + 2t\tau - 1 = 0 \quad &\Rightarrow \quad t = \frac{\mathrm{sign}\tau}{|\tau| + \sqrt{1+\tau^2}}, (|\theta| \le \pi/4) \\
&\Rightarrow \quad c = \frac{1}{\sqrt{1+t^2}}, \quad s = tc
\end{aligned} \tag{6}
$$

设$\mathbf{A}_1 = \mathbf{A}$，在$\mathbf{A}_k$中选取非对角元素绝对值最大的元素$a_{pq}^{(k)}$，按照公式(6)选取合适的$c$和$s$，构造相应的矩阵$\mathbf{J}_k = \mathbf{J}(p, q, \theta_k)$,使得

$$\mathbf{A}_{k+1} = \mathbf{J}_k \mathbf{A}_k \mathbf{J}_k^\top$$

则有

$$\mathrm{off}(\mathbf{A}_{k+1}) = (\mathrm{off})(\mathbf{A}_k) - 2\mathrm{a}_{\mathrm{pq}}^{(\mathrm{k})2} \leq 1 - \frac{1}{\mathrm{N}}\mathrm{off}(\mathbf{A}_k),$$

其中$N = \frac{n(n-1)}{2}$。

更进一步，

$$\mathbf{J}_m \cdots \mathbf{J}_1 \mathbf{A} \mathbf{J}_1^\top \cdots \mathbf{J}_m^\top \approx \mathbf{D}$$

$\mathbf{D}$为一个对角矩阵，$\mathbf{Q}_m = \mathbf{J}_1^\top \cdots \mathbf{J}_m^\top$为近似特征向量。

注：合理选择更新次序，平行化实现，实现迭代加速。

对称矩阵的QR方法

若$\mathbf{A} = \mathbf{A}^\top$，存在正交矩阵$\mathbf{Q}$，使得$\mathbf{Q}^\top\mathbf{A}\mathbf{Q} = \mathbf{H}$为实对称三对角矩阵。

设$\mathbf{T}$为实对称三对角矩阵，$\mathbf{T}$的QR分解具有如下性质：

- $\mathbf{T} = \mathbf{Q}\mathbf{R}$，正交矩阵$\mathbf{Q}$下带宽为$1$，上三角矩阵$\mathbf{R}$上带宽为$2$，而且$\mathbf{R}\mathbf{Q} = \mathbf{Q}^\top\mathbf{T}\mathbf{Q}$也是对称三对角矩阵。
- $s \in \mathbb{R}$，$\mathbf{T} - s\mathbf{T} = \mathbf{Q}\mathbf{R}$，则$\mathbf{R}\mathbf{Q} + s\mathbf{I} = \mathbf{Q}^\top\mathbf{T}\mathbf{Q}$也是三对角矩阵
- 设$\mathbf{T}$不可约，对一切$s \in \mathbb{R}$，有$\mathbf{T} - s\mathbf{I}$前$n-1$列线性无关。若$s \in \sigma(\mathbf{T})$，$\mathbf{Q}\mathbf{R} = \mathbf{T} - s\mathbf{I}$，则有$r_{nn} = 0$且$\mathbf{R}\mathbf{Q} + s\mathbf{I}$的最后一列等于$s\mathbf{e}_n$

迭代格式：

$$\mathbf{T}_k - \mu_k\mathbf{I} = \mathbf{Q}_k\mathbf{R}_k \ (\text{QR分解}), \quad \mathbf{T}_{k+1} = \mathbf{R}_k\mathbf{Q}_k + \mu_k\mathbf{I}$$

μ_k的选取：Wilkinson位移。