第 9 章 矩阵的特征值与特征向量计算

姚遥

南京大学智能科学与技术学院

本章目录

- 引言
- 幂法
- 反幂法
- Householder方法
- QR算法

特征值与特征向量 | Householder方法

前面几节讨论的是求解矩阵最大（小）特征值及其对应特征向量的方法，若要求出所有特征值及其特征向量，应该用什么方法？

思路：

- 将矩阵 $\boldsymbol{A} \in \mathbf{R}^{n\times n}$ 约化为更易求解特征值/向量的矩阵。具体的，可通过正交相似变换来做约化
  - 一般实矩阵： Hessenberg阵
  - 对称矩阵：对称三对角阵

特征值与特征向量 | 正交相似变换

定理9.10 设 $\boldsymbol{A} \in \mathbf{R}^{\mathbf{n}\times\mathbf{n}}$，则存在一个正交阵 $\boldsymbol{R}$，使

$$\boldsymbol{R}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{R} = \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1s} \\ & T_{22} & \cdots & T_{2s} \\ & & \ddots & \vdots \\ & & & T_{ss} \end{bmatrix}$$

其中对角块为一阶或二阶矩阵

- 每个一阶对角块即为 $\boldsymbol{A}$ 的实特征值
- 每个二阶对角块的两个特征值是 $\boldsymbol{A}$ 的一对共轭复特征值
- 参考Golub和Loan的《Matrix Computations》

特征值与特征向量 | 上Hessenberg（海森堡）阵

定义9.2 一方阵 $\boldsymbol{B}$，如果当 $i > j+1$ 时有 $b_{ij} = 0$，则称 $\boldsymbol{B}$ 为上Hessenberg阵，即

$$\boldsymbol{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ & \ddots & \ddots & \vdots \\ & & b_{n,n-1} & b_{nn} \end{bmatrix}$$

本节讨论如下两个问题

- 用正交相似变换约化一般实矩阵为上Hessenberg阵
- 用正交相似变换约化对称阵为对称三对角阵，就变成求转换后矩阵的特征值问题
- 从最基本的向量“约化”（初等反射阵）做起

特征值与特征向量 | 初等反射阵

定义9.3 设向量 $\boldsymbol{w}$ 满足 $\|\boldsymbol{w}\|_2 = 1$，矩阵 $\boldsymbol{H} = \boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^T$ 称为初等反射阵，记作 $H(\boldsymbol{w})$，即

$$\boldsymbol{H}(\boldsymbol{w}) = \begin{bmatrix} 1-2w_1^2 & -2w_1w_2 & \cdots & -2w_1w_n \\ -2w_2w_1 & 1-2w_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & -2w_{n-1}w_n \\ -2w_nw_1 & \cdots & -2w_nw_{n-1} & 1-2w_n^2 \end{bmatrix}$$

其中，$\boldsymbol{w} = (w_1, w_2, \cdots, w_n)^T$

定理9.11 初等反射阵 $\boldsymbol{H}$ 是对称阵 $(\boldsymbol{H}^T = \boldsymbol{H})$、正交阵 $(\boldsymbol{H}^T\boldsymbol{H} = \boldsymbol{I})$ 和对合阵 $(\boldsymbol{H}^2 = I)$

特征值与特征向量 | 初等反射阵

- 只证 $\boldsymbol{H}$ 的正交性，其他显然

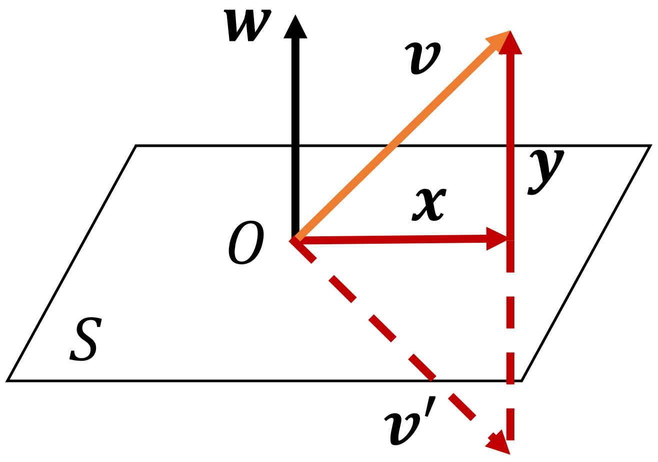
$$
\begin{aligned}
\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H} &= \boldsymbol{H}^2 \\
&= \left(\boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}\right)\left(\boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}\right) \\
&= \boldsymbol{I} - 4\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}} + 4\boldsymbol{w}\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}\right)\boldsymbol{w}^{\mathrm{T}} \\
&= \boldsymbol{I}
\end{aligned}
$$

- 设向量 $\boldsymbol{u} \neq \boldsymbol{0}$，则显然 $\boldsymbol{H} = \boldsymbol{I} - 2\dfrac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\|\boldsymbol{u}\|_2^2}$ 是一个初等反射阵

特征值与特征向量 | 初等反射阵的几何含义

考虑以 w 为法向量过原点 O 的超平面

$$S:\ \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}=0$$



- 设任意向量 $\boldsymbol{v}\in\mathbf{R}^n$，则

$$\boldsymbol{v}=\boldsymbol{x}+\boldsymbol{y}$$

  其中 $\boldsymbol{x}\in S,\ \ \boldsymbol{y}\in S^{\perp}$

- 根据 $\boldsymbol{x}\in S$，可得

$$\boldsymbol{Hx}=\left(\boldsymbol{I}-2\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}\right)\boldsymbol{x}=\boldsymbol{x}-2\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{x}$$

- $\boldsymbol{y}\in S^{\perp}$ 意味着 $\boldsymbol{y}$ 与 $\boldsymbol{w}$ 共线，可得

$$\boldsymbol{Hy}=\left(\boldsymbol{I}-2\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}\right)\boldsymbol{y}=\boldsymbol{y}-\boldsymbol{2y}=-\boldsymbol{y}$$

- 因此 $\boldsymbol{Hv}=\boldsymbol{x}-\boldsymbol{y}=\boldsymbol{v}'$，$\boldsymbol{v}'$ 为 $\boldsymbol{v}$ 关于平面 S 的镜面反射

定理9.12 设 $\boldsymbol{x},\boldsymbol{y}$ 为两个不相等的 n 维向量，$\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2$，则存在一个初等反射阵 $\boldsymbol{H}$，使 $\boldsymbol{Hx}=\boldsymbol{y}$

- 令 $\boldsymbol{w}=\frac{\boldsymbol{x}-\boldsymbol{y}}{\|\boldsymbol{x}-\boldsymbol{y}\|_2}$，则得到一个初等反射阵

$$\boldsymbol{H}=\boldsymbol{I}-2\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}=\boldsymbol{I}-2\frac{(\boldsymbol{x}-\boldsymbol{y})}{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}\left(\boldsymbol{x}^{\mathrm{T}}-\boldsymbol{y}^{\mathrm{T}}\right)$$

- 易得

$$\begin{aligned}\boldsymbol{Hx}&=\boldsymbol{x}-2\frac{\boldsymbol{x}-\boldsymbol{y}}{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}\left(\boldsymbol{x}^{\mathrm{T}}-\boldsymbol{y}^{\mathrm{T}}\right)\boldsymbol{x}\\&=\boldsymbol{x}-2\frac{(\boldsymbol{x}-\boldsymbol{y})\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}\right)}{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}\end{aligned}$$

特征值与特征向量 | Householder方法

- 根据 $\|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2$，可得

$$\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 = (\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y}) = 2\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}\right)$$

- 所以

$$\boldsymbol{H}\boldsymbol{x} = \boldsymbol{x}-(\boldsymbol{x}-\boldsymbol{y}) = \boldsymbol{y}$$

易知，$\boldsymbol{w} = \frac{\boldsymbol{x}-\boldsymbol{y}}{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}$ 是使 $\boldsymbol{H}\boldsymbol{x} = \boldsymbol{y}$ 成立的唯一长度等于 1 的向量（不计符号）

特征值与特征向量 | Householder：推论

初等反射阵在计算上的意义是它能用来约化矩阵，例如设向量 $\boldsymbol{a} \neq \boldsymbol{0}$，可选择一初等反射阵 H 使 $H\boldsymbol{a} = \sigma\boldsymbol{e}_1$，这种约化矩阵的方法称为Householder方法

推论 设向量 $\boldsymbol{x} \in \mathbf{R}^n$ $(x \neq 0)$，$\sigma = \pm\|\boldsymbol{x}\|_2$，且 $\boldsymbol{x} \neq -\sigma\boldsymbol{e}_1$，则存在一个初等反射阵

$$\boldsymbol{H} = \boldsymbol{I} - 2\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\|\boldsymbol{u}\|_2^2} \equiv \boldsymbol{I} - \rho^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$$

使 $\boldsymbol{H}\boldsymbol{x} = -\sigma\boldsymbol{e}_1,\ \boldsymbol{u} = \boldsymbol{x} + \sigma\boldsymbol{e}_1,\ \rho = \|\boldsymbol{u}\|_2^2/2$

- 应用定理9.12，令 $\boldsymbol{y} = -\sigma\boldsymbol{e}_1$，则 $\|\boldsymbol{y}\|_2 = \|\boldsymbol{x}\|_2$，$\boldsymbol{x} - \boldsymbol{y} = \boldsymbol{x} + \sigma\boldsymbol{e}_1$

特征值与特征向量 | Householder：推论

如何选择 $\sigma = \pm\|\boldsymbol{x}\|_2$ ？

- 设 $\boldsymbol{x} = (\alpha_1, \alpha_2, \cdots, \alpha_n)^{\mathrm{T}} \neq \mathbf{0}$，则

$$\boldsymbol{u} = \boldsymbol{x} + \sigma \boldsymbol{e}_1 = (\alpha_1 + \sigma, \alpha_2, \cdots, \alpha_n)^{\mathrm{T}}$$

- 因此

$$\begin{aligned}\rho = \frac{\|\boldsymbol{u}\|_2^2}{2} &= \frac{(\alpha_1+\sigma)^2 + \alpha_2^2 + \cdots + \alpha_n^2}{2} \\ &= \frac{\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2 + \sigma^2}{2} + \alpha_1\sigma \\ &= \sigma^2 + \alpha_1\sigma \quad = \sigma(\sigma + \alpha_1) \qquad (\sigma = \pm\|\boldsymbol{x}\|_2)\end{aligned}$$

- 若 σ、α_1 异号，则计算 $(\sigma + \alpha_1)$ 时有效数字可能损失，故取 σ、α_1 有相同符号

$$\sigma = \operatorname{sgn}(\alpha_1)\,\|\boldsymbol{x}\|_2$$

特征值与特征向量 | Householder：推论

如何选择 $\sigma = \pm\|\boldsymbol{x}\|_2$ ？

- 设 $\boldsymbol{x} = (\alpha_1, \alpha_2, \cdots, \alpha_n)^{\mathrm{T}} \neq \mathbf{0}$，则

$$\boldsymbol{u} = \boldsymbol{x} + \sigma \boldsymbol{e}_{\mathbf{1}} = (\alpha_1 + \sigma, \alpha_2, \cdots, \alpha_n)^{\mathrm{T}}$$

- 因此

$$\begin{aligned}\rho = \frac{\|\boldsymbol{u}\|_2^2}{2} &= \frac{(\alpha_1 + \sigma)^2 + \alpha_2^2 + \cdots + \alpha_n^2}{2} \\ &= \frac{\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2 + \sigma^2}{2} + \alpha_1 \sigma \\ &= \sigma^2 + \alpha_1 \sigma \quad = \sigma(\sigma + a_1) \qquad (\sigma = \pm\|\boldsymbol{x}\|_2)\end{aligned}$$

- 若 σ、α_1 异号，则计算 $(\sigma + a_1)$ 时有效数字可能损失，故取 σ、α_1 有相同符号

$$\sigma = \operatorname{sgn}(\alpha_1)\, \|\boldsymbol{x}\|_2$$

特征值与特征向量 | Householder：算法1

已知向量 $x=(\alpha_1,\alpha_2,\cdots,\alpha_n)^{\mathrm{T}}\neq \mathbf{0}$，本算法算出 σ、ρ 及 $\boldsymbol{u}$，使 $\left(\boldsymbol{I}-\rho^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\right)\boldsymbol{x}=-\sigma\boldsymbol{e}_1$，$\boldsymbol{u}$ 的分量冲掉 x 的分量

1. 计算 $\sigma=\operatorname{sgn}(\alpha_1)\left(\sum_{i=1}^{n}\alpha_i^2\right)^{\frac{1}{2}}$
2. $\alpha_1\leftarrow u_1=\alpha_1+\sigma$
3. $\rho=\sigma(\sigma+a_1)=\sigma u_1$

$$\left.\begin{array}{l}\alpha_1\leftarrow u_1=\alpha_1+\sigma\\ \rho=\sigma(\sigma+a_1)=\sigma u_1\end{array}\right\}\boldsymbol{H}=\left(\boldsymbol{I}-\rho^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\right)$$

在计算 σ 时，可能上溢或下溢，为了避免溢出，将 x 规范化

$$\eta=\max_i|\alpha_i|,\qquad \boldsymbol{x}'=\frac{\boldsymbol{x}}{\eta}$$

- 显然

$$\sigma'=\sigma/\eta,\qquad \boldsymbol{H}'=\boldsymbol{H}$$

特征值与特征向量 | Householder：算法2

已知 $\boldsymbol{x}=(\alpha_1,\alpha_2,\cdots,\alpha_n)^{\mathrm{T}}\neq\mathbf{0}$，本算法算出 $\boldsymbol{H}$ 及 σ 使 $\boldsymbol{Hx}\neq-\sigma\boldsymbol{e}_1$， $\boldsymbol{u}$ 的分量冲掉 x 的分量

1. $\eta=\max\limits_i|\alpha_i|$
2. $\alpha_i\leftarrow u_i=\dfrac{\alpha_i}{\eta}\quad(i=1,2,\cdots,n)$
3. $\sigma=\mathrm{sgn}(u_1)\left(\sum_{i=1}^n u_i^2\right)^{\frac{1}{2}}$

$$\left.\begin{array}{l}4.\ \ u_1\leftarrow u_1+\sigma\\[2ex] 5.\ \ \rho=\sigma u_1\end{array}\right\}\ \boldsymbol{H}=\left(\boldsymbol{I}-\rho^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\right)$$

6. $\sigma\leftarrow\eta\sigma$ （原始 $\boldsymbol{x}$ 对应的 σ ）

特征值与特征向量 | Householder：矩阵计算

关于 HA 的计算，设 $A=(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n)$，其中 $\boldsymbol{a}_i$ 为 A 的第 i 列向量，则

$$\boldsymbol{HA}=(\boldsymbol{Ha}_1,\boldsymbol{Ha}_2,\cdots,\boldsymbol{Ha}_n)$$

因此计算 HA 就是要计算 n 次

$$\begin{aligned}\boldsymbol{Ha_i}&=\left(\boldsymbol{I}-\rho^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\right)\boldsymbol{a}_i\\&=\boldsymbol{a}_i-\left(\rho^{-1}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{a}_i\right)\boldsymbol{u}\quad(i=1,2,\cdots,n)\end{aligned}$$

- 于是计算 $\boldsymbol{Ha}_i$，只需要计算两向量的数量积和两向量的加法

计算 HA 共需要 $2n^2$ 次乘法运算

特征值与特征向量 | 用正交相似变换约化矩阵

设

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \equiv \begin{bmatrix} a_{11} & \boldsymbol{A}_{12}^{(1)} \\ \boldsymbol{a}_{21}^{(1)} & \boldsymbol{A}_{22}^{(1)} \end{bmatrix}$$

- 步 1：设 $\boldsymbol{a}_{21}^{(1)} \neq \boldsymbol{0}$，否则这一步不需约化，选择初等反射阵 $\boldsymbol{R}_1$ 使 $\boldsymbol{R}_1\boldsymbol{a}_{21}^{(1)} = -\sigma_1\boldsymbol{e}_1$，其中

$$\begin{cases} \sigma_1 = \operatorname{sgn}(a_{21})\left(\sum_{i=2}^{n} a_{i1}^2\right)^{1/2} \\ \boldsymbol{u}_1 = \boldsymbol{a}_{21}^{(1)} + \sigma_1\boldsymbol{e}_1 \\ \rho_1 = \dfrac{1}{2}\|\boldsymbol{u}_1\|_2^2 = \sigma_1(\sigma_1 + a_{21}) \\ \boldsymbol{R}_1 = \boldsymbol{I} - \rho_1^{-1}\boldsymbol{u}_1\boldsymbol{u}_1^{\mathrm{T}} \end{cases}$$

特征值与特征向量 | 用正交相似变换约化矩阵

- 令 $U_1 = \begin{bmatrix} I & O \\ O & R_1 \end{bmatrix}$，则

$$A_2 = U_1 A_1 U_1 = \begin{bmatrix} a_{11} & A_{12}^{(1)} R_1 \\ R_1 a_{21}^{(1)} & R_1 A_{22}^{(1)} R_1 \end{bmatrix}$$

$$= \begin{bmatrix} A_{11}^{(2)} & a_{21}^{(2)} & A_{13}^{(2)} \\ O & a_{22}^{(2)} & A_{23}^{(2)} \end{bmatrix}$$

其中

$$A_{11}^{(2)} \in \mathbf{R}^{2\times 1}, \quad a_{22}^{(2)} \in \mathbf{R}^{n-2}, \quad A_{23}^{(2)} \in \mathbf{R}^{(n-2)\times(n-2)}$$

特征值与特征向量 | 用正交相似变换约化矩阵

- 步 k：设对 $\boldsymbol{A}$ 已进行了第 $k-1$ 步正交相似约化，即 $\boldsymbol{A}_k$ 有形式

$$
\boldsymbol{A}_k = \boldsymbol{U}_{k-1}\boldsymbol{A}_{k-1}\boldsymbol{U}_{k-1} = \begin{bmatrix}
a_{11}^{(1)} & a_{12}^{(2)} & \cdots & a_{1k}^{(k)} & a_{1,k+1}^{(k)} & \cdots & a_{1n}^{(k)} \\
-\sigma_1 & a_{22}^{(2)} & \cdots & a_{2k}^{(k)} & a_{2,k+1}^{(k)} & \cdots & a_{2n}^{(k)} \\
 & \ddots & \ddots & \vdots & \vdots & & \vdots \\
 & & -\sigma_{k-1} & a_{kk}^{(k)} & a_{k,k+1}^{(k)} & \cdots & a_{k,n}^{(k)} \\
 & & & {\color{red}a_{k+1,k}^{(k)}} & a_{k+1,k+1}^{(k)} & \cdots & a_{k+1,n}^{(k)} \\
 & & & {\color{red}\vdots} & \vdots & & \vdots \\
 & & & {\color{red}a_{nk}^{(k)}} & a_{n,k+1}^{(k)} & \cdots & a_{nn}^{(k)}
\end{bmatrix}
$$

$$
= \begin{bmatrix}
\boldsymbol{A}_{11}^{(k)} & \boldsymbol{a}_{12}^{(k)} & \boldsymbol{A}_{13}^{(k)} \\
\boldsymbol{O} & {\color{red}\boldsymbol{a}_{22}^{(k)}} & \boldsymbol{A}_{23}^{(k)}
\end{bmatrix}
$$

其中 $\boldsymbol{A}_{11}^{(k)} \in \mathbf{R}^{k\times(k-1)}$，$\boldsymbol{a}_{22}^{(2)} \in \mathbf{R}^{n-k}$，$\boldsymbol{A}_{23}^{(k)} \in \mathbf{R}^{(n-k)\times(n-k)}$

特征值与特征向量 | 用正交相似变换约化矩阵

- 设 $\boldsymbol{a}_{22}^{(k)} \neq \mathbf{0}$，选择初等反射阵 $\boldsymbol{R}_k$ ，使 $\boldsymbol{R}_k\boldsymbol{a}_{22}^{(k)} = -\sigma_k\boldsymbol{e}_1$，其中

$$\begin{cases} \sigma_k = \operatorname{sgn}\left(a_{k+1,k}^{(k)}\right)\left(\sum_{i=k+1}^{n} a_{ik}^2\right)^{1/2} \\ \boldsymbol{u}_k = \boldsymbol{a}_{22}^{(k)} + \sigma_k\boldsymbol{e}_1 \\ \rho_k = \dfrac{1}{2}\|\boldsymbol{u}_k\|_2^2 = \sigma_k(\sigma_k + a_{k+1,n}^{(k)}) \\ \boldsymbol{R}_k = \boldsymbol{I} - \rho_k^{-1}\boldsymbol{u}_k\boldsymbol{u}_k{}^{\mathrm{T}} \end{cases}$$

- 设$\boldsymbol{U}_k = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{R}_k \end{bmatrix}$，则

$$\boldsymbol{A}_{k+1} = \boldsymbol{U}_k\boldsymbol{A}_k\boldsymbol{U}_k = \begin{bmatrix} \boldsymbol{A}_{11}^{(k)} & \boldsymbol{a}_{12}^{(k)} & \boldsymbol{A}_{13}^{(k)}\boldsymbol{R}_k \\ \boldsymbol{O} & \boldsymbol{R}_k\boldsymbol{a}_{22}^{(k)} & \boldsymbol{R}_k\boldsymbol{A}_{23}^{(k)}\boldsymbol{R}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11}^{(k)} & \boldsymbol{a}_{12}^{(k)} & \boldsymbol{A}_{13}^{(k)}\boldsymbol{R}_k \\ \boldsymbol{O} & -\sigma_k\boldsymbol{e}_1 & \boldsymbol{R}_k\boldsymbol{A}_{23}^{(k)}\boldsymbol{R}_k \end{bmatrix}$$

特征值与特征向量 | 用正交相似变换约化矩阵

- 由上式知，$\boldsymbol{A}_{k+1}$ 的左上角 $k+1$ 阶子阵为上Hessenberg阵，从而约化又进了一步

- 重复这过程，直到

$$\boldsymbol{A}_{n-1} = \boldsymbol{U}_{n-2} \cdots \boldsymbol{U}_2 \boldsymbol{U}_1 \boldsymbol{A} \boldsymbol{U}_1 \boldsymbol{U}_2 \cdots \boldsymbol{U}_{n-2}$$

$$= \begin{bmatrix} a_{11} & \times & \times & \cdots & \times \\ -\sigma_1 & a_{22}^{(2)} & \times & \cdots & \times \\ & -\sigma_2 & a_{33}^{(3)} & \ddots & \vdots \\ & & \ddots & \ddots & \times \\ & & & -\sigma_{n-1} & a_{nn}^{(n-1)} \end{bmatrix}$$

特征值与特征向量 | 用正交相似变换约化矩阵

定理9.13 设 $\boldsymbol{A} \in \mathbf{R}^{n\times n}$，则存在初等反射阵 $\boldsymbol{U}_1, \boldsymbol{U}_2, \cdots, \boldsymbol{U}_{n-2}$ 使

$$\boldsymbol{U}_{n-2}\cdots\boldsymbol{U}_2\boldsymbol{U}_1\boldsymbol{A}\boldsymbol{U}_1\boldsymbol{U}_2\cdots\boldsymbol{U}_{n-2} = \boldsymbol{C} \text{（上Hessenberg阵）}$$

在 $\boldsymbol{A}_k \to \boldsymbol{A}_{k+1} = \boldsymbol{U}_k\boldsymbol{A}_k\boldsymbol{U}_k$ 的进一步约化中，需要计算 $\boldsymbol{R}_k$ 和 $\boldsymbol{A}_{13}^{(k)}\boldsymbol{R}_k$，$\boldsymbol{R}_k\boldsymbol{A}_{23}^{(k)}\boldsymbol{R}_k$

$$\boldsymbol{A}_{k+1} = \boldsymbol{U}_k\boldsymbol{A}_k\boldsymbol{U}_k = \begin{bmatrix} \boldsymbol{A}_{11}^{(k)} & \boldsymbol{a}_{12}^{(k)} & \boldsymbol{A}_{13}^{(k)}\boldsymbol{R}_k \\ \boldsymbol{O} & -\sigma_k\boldsymbol{e}_1 & \boldsymbol{R}_k\boldsymbol{A}_{23}^{(k)}\boldsymbol{R}_k \end{bmatrix}$$

- 用初等反射阵正交相似约化 $\boldsymbol{A}$ 为上Hessenberg阵，大约需要 $\frac{5}{3}n^3$ 次乘法运算

特征值与特征向量｜特征值求解

由于 U_k 都是正交阵，所以$A_1 \sim A_2 \sim \cdots \sim A_{n-1}$，求 A 的特征值问题，就转化为求上Hessenberg阵 C 的特征值问题

- 由定理9.13，记 $\boldsymbol{P} = \boldsymbol{U}_{n-2} \cdots \boldsymbol{U}_2 \boldsymbol{U}_1$ 则

$$\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{C}$$

- $\boldsymbol{y}$ 是 $\boldsymbol{C}$ 的对应特征值 λ 的特征向量，则 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{y}$ 为 $\boldsymbol{A}$ 的对应特征值 λ 的特征向量，且

$$\begin{aligned}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{y} &= \boldsymbol{U}_1\boldsymbol{U}_2 \cdots \boldsymbol{U}_{n-2}\boldsymbol{y} \\ &= \left(\boldsymbol{I} - \lambda_1^{-1}\boldsymbol{u}_1\boldsymbol{u}_1^{\mathrm{T}}\right) \cdots \left(\boldsymbol{I} - \lambda_{n-2}^{-1}\boldsymbol{u}_{n-2}\boldsymbol{u}_{n-2}^{\mathrm{T}}\right)\boldsymbol{y}\end{aligned}$$

特征值与特征向量 | 约化对称矩阵

定理9.14 设 $A \in \mathbf{R}^{n\times n}$ 为对称阵，存在初等反射阵 $\boldsymbol{U}_1, \boldsymbol{U}_2, \cdots, \boldsymbol{U}_{n-2}$ 使

$$\boldsymbol{A}_{n-1} = \boldsymbol{U}_{n-2}\cdots\boldsymbol{U}_1\boldsymbol{A}\boldsymbol{U}_1\cdots\boldsymbol{U}_{n-2}$$

$$= \begin{bmatrix} c_1 & b_1 & & & \\ b_1 & c_2 & b_2 & & \\ & \ddots & \ddots & \ddots & \\ & & b_{n-2} & c_{n-1} & b_{n-1} \\ & & & b_{n-1} & c_n \end{bmatrix} \equiv \boldsymbol{C}$$

- 由定理9.13，存在初等反射阵 $\boldsymbol{U}_1, \boldsymbol{U}_2, \cdots, \boldsymbol{U}_{n-2}$，使 $\boldsymbol{A}_{n-1}$ 为上Hessenberg阵，但 $\boldsymbol{A}_{n-1}$ 又为对称阵，因此 $\boldsymbol{A}_{n-1}$ 为对称三对角阵

特征值与特征向量 | 约化对称矩阵

对于约化过程

$$\boldsymbol{A}_{k+1}=\boldsymbol{U}_k\boldsymbol{A}_k\boldsymbol{U}_k=\begin{bmatrix}\boldsymbol{A}_{11}^{(k)} & \boldsymbol{a}_{12}^{(k)} & \boldsymbol{A}_{13}^{(k)}\boldsymbol{R}_k\\ \boldsymbol{O} & -\sigma_k\boldsymbol{e}_1 & \boldsymbol{R}_k\boldsymbol{A}_{23}^{(k)}\boldsymbol{R}_k\end{bmatrix}$$

- 由于对称性，在 $\boldsymbol{A}_k\to\boldsymbol{A}_{k+1}=\boldsymbol{U}_k\boldsymbol{A}_k\boldsymbol{U}_k$ 的计算过程中，只需计算 $\boldsymbol{R}_k$ 和 $\boldsymbol{R}_k\boldsymbol{A}_{23}^{(k)}\boldsymbol{R}_k$
- 更进一步，只计算 $\boldsymbol{R}_k\boldsymbol{A}_{23}^{(k)}\boldsymbol{R}_k$ 对角线下面的元素
- 注意到

$$\begin{aligned}\boldsymbol{R}_k\boldsymbol{A}_{23}^{(k)}\boldsymbol{R}_k&=\left(\boldsymbol{I}-\rho_k^{-1}\boldsymbol{u}_k\boldsymbol{u}_k^{\mathrm{T}}\right)\left(\boldsymbol{A}_{23}^{(k)}-\rho_k^{-1}\boldsymbol{A}_{23}^{(k)}\boldsymbol{u}_k\boldsymbol{u}_k^{\mathrm{T}}\right)\\&=\boldsymbol{A}_{23}^{(k)}-\rho_k^{-1}\boldsymbol{u}_k\boldsymbol{u}_k^{\mathrm{T}}\boldsymbol{A}_{23}^{(k)}-\rho_k^{-1}\boldsymbol{A}_{23}^{(k)}\boldsymbol{u}_k\boldsymbol{u}_k^{\mathrm{T}}+\rho_k^{-1}\boldsymbol{u}_k\boldsymbol{u}_k^{\mathrm{T}}\rho_k^{-1}\boldsymbol{A}_{23}^{(k)}\boldsymbol{u}_k\boldsymbol{u}_k^{\mathrm{T}}\end{aligned}$$

特征值与特征向量 | 约化对称矩阵

$$= \boldsymbol{A}_{23}^{(k)} - \boldsymbol{u}_k \left(\rho_k^{-1}\boldsymbol{A}_{23}^{(k)}\boldsymbol{u}_k\right)^{\mathrm{T}} - \left(\rho_k^{-1}\boldsymbol{A}_{23}^{(k)}\boldsymbol{u}_k\right)\boldsymbol{u}_k^{\mathrm{T}} + \rho_k^{-1}\left(\boldsymbol{u}_k^{\mathrm{T}}\rho_k^{-1}\boldsymbol{A}_{23}^{(k)}\boldsymbol{u}_k\right)\boldsymbol{u}_k\boldsymbol{u}_k^{\mathrm{T}}$$

$$= \boldsymbol{A}_{23}^{(k)} - \boldsymbol{u}_k \left(\rho_k^{-1}\boldsymbol{A}_{23}^{(k)}\boldsymbol{u}_k - \frac{\rho_k^{-1}}{2}\boldsymbol{u}_k^{\mathrm{T}}\rho_k^{-1}\boldsymbol{A}_{23}^{(k)}\boldsymbol{u}_k\boldsymbol{u}_k\right)^{\mathrm{T}}$$

$$-\left(\rho_k^{-1}\boldsymbol{A}_{23}^{(k)}\boldsymbol{u}_k - \frac{\rho_k^{-1}}{2}\boldsymbol{u}_k^{\mathrm{T}}\rho_k^{-1}\boldsymbol{A}_{23}^{(k)}\boldsymbol{u}_k\boldsymbol{u}_k\right)\boldsymbol{u}_k^{\mathrm{T}}$$

$$= \boldsymbol{A}_{23}^{(k)} - \boldsymbol{u}_k\boldsymbol{t}_k^{\mathrm{T}} - \boldsymbol{t}_k\boldsymbol{u}_k^{\mathrm{T}}$$

其中 $\boldsymbol{r}_k = \rho_k^{-1}\boldsymbol{A}_{23}^{(k)}\boldsymbol{u}_k$, $\boldsymbol{t}_k = \boldsymbol{r}_k - \frac{\rho_k^{-1}}{2}(\boldsymbol{u}_k^{\mathrm{T}}\boldsymbol{r}_k)\boldsymbol{u}_k$

特征值与特征向量 | 约化对称矩阵：讨论

- 将对称阵 $\boldsymbol{A}$ 用初等反射阵正交相似约化为对称三对角阵约需做 $\frac{2}{3}n^3$ 次乘法运算
- 将一般实矩阵 $\boldsymbol{A}$ 用初等反射阵正交相似约化为上Hessenberg阵，大约需要 $\frac{5}{3}n^3$ 次乘法运算
- 用正交矩阵进行约化，有一些特点，如构造的 $\boldsymbol{U}_k$ 容易求逆，且 $\boldsymbol{U}_k$ 的元素数量级不大，因此这个算法十分稳定的

例9.5 用Householder方法将下述矩阵化为Hessenberg阵

$$A_1 = A = \begin{bmatrix} -4 & -3 & -7 \\ 2 & 3 & 2 \\ 4 & 2 & 7 \end{bmatrix}$$

1. 对于 $k = 1$，确定变换

$$U_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & & \\ 0 & & R_1 \end{bmatrix}, \qquad a_{21}^{(1)} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

其中 R_1 为初等反射阵且满足

$$R_1 a_{21}^{(1)} = -\sigma_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \sigma_1 = \left\| a_{21}^{(1)} \right\|_2 = \sqrt{20} \approx 4.472136$$

特征值与特征向量 | Householder方法举例

$$\boldsymbol{u}_1 = \boldsymbol{a}_{21}^{(1)} + \sigma_1\rho_1 = \begin{pmatrix} 2+\sqrt{20} \\ 4 \end{pmatrix} \approx \begin{pmatrix} 6.472136 \\ 4 \end{pmatrix}$$

$$\rho_1 = \sigma_1(\sigma_1 + a_{21}) = \sqrt{20}\left(\sqrt{20}+2\right) \approx 28.94427$$

$$\boldsymbol{R}_1 = \boldsymbol{I} - \rho_1^{-1}\boldsymbol{u}_1\boldsymbol{u}_1^{\mathrm{T}}$$

2. 计算 $\boldsymbol{R}_1\boldsymbol{A}_{22}^{(1)}$，记

$$\boldsymbol{A}_{22}^{(1)} = \begin{bmatrix} 3 & 2 \\ 2 & 7 \end{bmatrix} \equiv (\boldsymbol{a}_1, \boldsymbol{a}_2)$$

$$\boldsymbol{R}_1\boldsymbol{A}_{22}^{(1)} = (\boldsymbol{R}_1\boldsymbol{a}_1, \boldsymbol{R}_1\boldsymbol{a}_2) = \begin{bmatrix} -3.130496 & -7.155419 \\ -1.788855 & 1.341640 \end{bmatrix}$$

$$\boldsymbol{R}_1\boldsymbol{a}_i = \left(\boldsymbol{I} - \rho_1^{-1}\boldsymbol{u}_1\boldsymbol{u}_1^{\mathrm{T}}\right)\boldsymbol{a}_i = \boldsymbol{a}_i - \left(\rho_1^{-1}\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{a}_i\right)\boldsymbol{u}_1 \quad (i = 1,2)$$

特征值与特征向量 | Householder方法举例

3. 计算 $\boldsymbol{A}_{12}^{(1)}\boldsymbol{R}_1$ 及 $\left(\boldsymbol{R}_1\boldsymbol{A}_{22}^{(1)}\right)\boldsymbol{R}_1$，即计算

$$\begin{bmatrix} \boldsymbol{A}_{12}^{(1)} \\ \left(\boldsymbol{R}_1\boldsymbol{A}_{22}^{(1)}\right) \end{bmatrix}\boldsymbol{R}_1 \equiv \begin{bmatrix} \boldsymbol{b}_1^{\mathrm{T}} \\ \boldsymbol{b}_2^{\mathrm{T}} \\ \boldsymbol{b}_3^{\mathrm{T}} \end{bmatrix}\boldsymbol{R}_1 = \begin{bmatrix} \boldsymbol{b}_1^{\mathrm{T}}\boldsymbol{R}_1 \\ \boldsymbol{b}_2^{\mathrm{T}}\boldsymbol{R}_1 \\ \boldsymbol{b}_3^{\mathrm{T}}\boldsymbol{R}_1 \end{bmatrix}$$

$$= \begin{bmatrix} 7.602634 & -0.447212 \\ 7.800003 & -0.399999 \\ -0.399999 & 2.200000 \end{bmatrix}$$

其中

$$\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{R}_1 = \boldsymbol{b}_i^{\mathrm{T}} - \left(\rho_1^{-1}\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{u}_1\right)\boldsymbol{u}_1^{\mathrm{T}} \quad (i = 1,2,3)$$

特征值与特征向量 | Householder方法举例

4. 得到 $\boldsymbol{A}_2 = \boldsymbol{U}_1\boldsymbol{A}_1\boldsymbol{U}_1$

$$\boldsymbol{A}_2 = \begin{bmatrix} -4 & \boldsymbol{A}_{12}^{(1)}\boldsymbol{R}_1 \\ -\sigma_1 & \boldsymbol{R}_1\boldsymbol{A}_{22}^{(1)}\boldsymbol{R}_1 \\ 0 & \end{bmatrix}$$

$$= \begin{bmatrix} -4 & 7.602634 & -0.447212 \\ -4.472136 & 7.800003 & -0.399999 \\ 0 & -0.399999 & 2.200000 \end{bmatrix}$$

为上Hessenberg阵

本章目录

- 引言
- 幂法
- 反幂法
- Householder方法
- QR算法

特征值与特征向量 | QR算法

Householder方法：上Hessenberg阵/对称三对角阵特征值/向量计算？

Francis（在1961年、1962年）利用矩阵的 QR 分解建立了计算矩阵特征值的 QR 方法，是计算一般矩阵（中小型矩阵）全部特征值问题的最有效的方法之一，主要用来计算

- 上Hessenberg阵的全部特征值问题
- 对称三对角阵的全部特征值问题

QR方法具有收敛快、算法稳定等特点

特征值与特征向量 | QR算法

对于一般矩阵 $\boldsymbol{A} \in \mathbf{R}^{n\times n}$（或对称阵）

- 首先用Householder方法将 $\boldsymbol{A}$ 化为上Hessenberg阵 $\boldsymbol{B}$（或对称三对角阵）
- 然后再用QR方法计算 $\boldsymbol{B}$ 的全部特征值问题
- 注意： QR方法可以直接处理任意矩阵 $\boldsymbol{A}$

平面旋转矩阵来约化矩阵

- 同样是一种用正交相似变换约化矩阵的方法
- 是QR方法的基础

特征值与特征向量 | 平面旋转矩阵（Givens变换）

$$\boldsymbol{P}_{i,j} = \begin{bmatrix} 1 & & & & & & & & & & \\ & \ddots & & & & & & & & & \\ & & 1 & & & & & & & & \\ & & & c & & & & s & & & \\ & & & & 1 & & & & & & \\ & & & & & \ddots & & & & & \\ & & & & & & 1 & & & & \\ & & & -s & & & & c & & & \\ & & & & & & & & 1 & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{bmatrix}$$

其中 $c = \cos\theta$，$s = \sin\theta$

引理1 设 $\boldsymbol{x} = (\alpha_1, \alpha_2, \cdots, \alpha_i, \cdots, \alpha_j, \cdots, \alpha_n)^T$，其中 α_i, α_j 不全为零，可选一平面旋转矩阵 $\boldsymbol{P}_{ij}$ 使 $\boldsymbol{y} \equiv \boldsymbol{P}_{ij}\boldsymbol{x} = \left(\alpha_1, \alpha_2, \cdots, \alpha_i^{(1)}, \cdots, \alpha_j^{(1)}, \cdots, \alpha_n\right)^T$，且

$$\alpha_i^{(1)} = \sqrt{\alpha_i^2 + \alpha_j^2} \tag{9.4.2}$$

$$\alpha_j^{(1)} = 0 \tag{9.4.3}$$

$$\begin{cases} c = \alpha_i \Big/ \sqrt{\alpha_i^2 + \alpha_j^2} \\ s = \alpha_j \Big/ \sqrt{\alpha_i^2 + \alpha_j^2} \end{cases} \tag{9.4.4}$$

注意到，$P_{ij}x$ 只改变 x 的第 i 个及第 j 个元素

特征值与特征向量 | 平面旋转算法

用平面旋转阵进行左变换可产生一算法，设给定$x=\begin{bmatrix}\alpha\\ \beta\end{bmatrix}$**，计算**$c=\cos\theta, s=\sin\theta$**，**

$v=\sqrt{\alpha^2+\beta^2}$**，使**$\boldsymbol{P}_{ij}\boldsymbol{x}=\begin{bmatrix}c & s\\ -s & c\end{bmatrix}\begin{bmatrix}\alpha\\ \beta\end{bmatrix}=\begin{bmatrix}v\\ 0\end{bmatrix}$

- 为了防止溢出，将 $\boldsymbol{x}$ 规范化，有

$$\eta\equiv\|\boldsymbol{x}\|_\infty=\max\{|\alpha|,|\beta|\}\neq 0,\quad \boldsymbol{x}'=\boldsymbol{x}/\eta=\begin{bmatrix}\alpha/\eta\\ \beta/\eta\end{bmatrix}$$

于是

$$\begin{cases}c'=c, \quad s'=s\\ v'=v/\eta\end{cases}$$

特征值与特征向量 | 平面旋转：算法1

设 $x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$，计算 c, s 及 v

1. 计算 $\eta = \max\{|\alpha|, |\beta|\}$
2. 如果 $\eta = 0$，则 $c \leftarrow 1, s \leftarrow 0$，转步 8
3. $\alpha' = \alpha/\eta$
4. $\beta' = \beta/\eta$
5. $v' = \sqrt{\alpha'^2 + \beta'^2}$
6. $c = \alpha'/v', s = \beta'/v'$
7. $v = \eta v'$
8. 计算终止

定理9.15 如果 $\boldsymbol{A}$ 为非奇异矩阵，则存在正交矩阵 $\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_{n-1}$（即一系列平面旋转矩阵）使

$$\boldsymbol{P}_{n-1}, \cdots, \boldsymbol{P}_2 \boldsymbol{P}_1 \boldsymbol{A} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix} \equiv \boldsymbol{R} \tag{9.4.5}$$

且 $r_{ii} > 0 \quad (i = 1,2, \cdots, n-1)$

- 由于 $\boldsymbol{A}$ 的第 1 列一定存在 $a_{j1} \neq 0$，于是，如果$a_{j1} \neq 0$ $(j = 2,3, \cdots, n)$，应用算法1，即存在平面旋转矩阵 $\boldsymbol{P}_{12}, \boldsymbol{P}_{13}, \cdots, \boldsymbol{P}_{1n}$，使

特征值与特征向量 | 用平面旋转矩阵约化矩阵

$$\boldsymbol{P}_{1n}\cdots\boldsymbol{P}_{13}\boldsymbol{P}_{12}\boldsymbol{A}=\begin{bmatrix} r_{11} & a_{12}^{(2)} & \cdots & a_{1n}^{(2)} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2}^{(2)} & \cdots & a_{nn}^{(2)} \end{bmatrix}\equiv\boldsymbol{A}^{(2)}$$

- 记 $\boldsymbol{P}_{1n}\cdots\boldsymbol{P}_{12}=\boldsymbol{P}_1$，则

$$\boldsymbol{P}_1\boldsymbol{A}=\boldsymbol{A}^{(2)}$$

- 同理如果 $a_{2j}^{(2)}\neq 0\ (j=3,\cdots,n)$，应用算法1，存在平面旋转矩阵 $\boldsymbol{P}_{23},\cdots,\boldsymbol{P}_{2n}$ (记 $\boldsymbol{P}_2=\boldsymbol{P}_{2n}\cdots\boldsymbol{P}_{23}$)，使

特征值与特征向量 | 用平面旋转矩阵约化矩阵

$$P_2P_1A=\begin{bmatrix} r_{11} & a_{12}^{(2)} & a_{13}^{(2)} & \cdots & a_{1n}^{(2)} \\ & r_{22} & a_{23}^{(3)} & \cdots & a_{2n}^{(3)} \\ & & a_{33}^{(3)} & \cdots & a_{3n}^{(3)} \\ & & \vdots & & \vdots \\ & & a_{n3}^{(3)} & \cdots & a_{nn}^{(3)} \end{bmatrix}$$

- 注意到，构造 $\boldsymbol{P}_2$ 的矩阵数量减少了 1 个
- 重复上述过程，最后得到：存在正交阵$\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_{n-1}$，使式(9.4.5)成立

特征值与特征向量 | 矩阵的QR分解

定理9.16 如果 $\boldsymbol{A} \in \mathrm{R}^{n\times n}$ 为非奇异矩阵，则 $\boldsymbol{A}$ 可分解为一正交阵 $\boldsymbol{Q}$ 与上三角阵 $\boldsymbol{R}$ 的乘积，即 $\boldsymbol{A} = \boldsymbol{QR}$，且当 $\boldsymbol{R}$ 对角元素都为正数时分解唯一

- 由定理9.12，存在正交阵 $\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_{n-1}$，使

$$\boldsymbol{P}_{n-1} \cdots \boldsymbol{P}_2 \boldsymbol{P}_1 \boldsymbol{A} = \boldsymbol{R} \tag{9.4.6}$$

  为上三角阵

- 记

$$\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{P}_{n-1} \cdots \boldsymbol{P}_2 \boldsymbol{P}_1$$

  于是式(9.4.6)为 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{R}$，即 $\boldsymbol{A} = \boldsymbol{QR}$，其中 $\boldsymbol{Q} = \boldsymbol{P}_1^{\mathrm{T}}\boldsymbol{P}_2^{\mathrm{T}} \cdots \boldsymbol{P}_{n-1}^{\mathrm{T}}$ 为正交阵

定理9.16 如果 $\boldsymbol{A}\in\mathbf{R}^{n\times n}$ 为非奇异矩阵，则 $\boldsymbol{A}$ 可分解为一正交阵 $\boldsymbol{Q}$ 与上三角阵 $\boldsymbol{R}$ 的乘积，即 $\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{R}$，且当 $\boldsymbol{R}$ 对角元素都为正数时分解唯一

- 唯一性证明：设有 $\boldsymbol{A}=\boldsymbol{Q}_1\boldsymbol{R}_1=\boldsymbol{Q}_2\boldsymbol{R}_2$ ，其中 $\boldsymbol{R}_1$，$\boldsymbol{R}_2$ 为上三角阵（显然为非奇异阵）且对角元素都为正数，$\boldsymbol{Q}_1,\boldsymbol{Q}_2$ 为正交阵
- 于是

$$\boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{Q}_1\boldsymbol{R}_1\boldsymbol{R}_1^{-1}=\boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{Q}_2\boldsymbol{R}_2\boldsymbol{R}_1^{-1} \quad \Rightarrow \boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{Q}_1=\boldsymbol{R}_2\boldsymbol{R}_1^{-1}$$

- 由于 $\boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{Q}_1$ 是正交阵，因此 $\boldsymbol{R}_2\boldsymbol{R}_1^{-1}$ 为正交阵
- 同时由于 $\boldsymbol{R}_2\boldsymbol{R}_1^{-1}$ 为上三角阵， $\boldsymbol{R}_2\boldsymbol{R}_1^{-1}$ 必然为对角阵，即

$$\boldsymbol{R}_2\boldsymbol{R}_1^{-1}=\boldsymbol{D}=\mathrm{diag}(d_1,d_2,\cdots,d_n)$$

定理9.16 如果 $\boldsymbol{A} \in \mathbf{R}^{n\times n}$ 为非奇异矩阵，则 $\boldsymbol{A}$ 可分解为一正交阵 $\boldsymbol{Q}$ 与上三角阵 $\boldsymbol{R}$ 的乘积，即 $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$，且当 $\boldsymbol{R}$ 对角元素都为正数时分解唯一

- 再次利用 $\boldsymbol{R}_2\boldsymbol{R}_1^{-1} = \boldsymbol{D}$ 是正交阵：

$$\boldsymbol{D}^2 = \boldsymbol{I} \Rightarrow d_i = \pm 1\ (i = 1,2,\cdots,n)$$

- 又因 $\boldsymbol{R}_1$，$\boldsymbol{R}_2$ 对角元素都为正数，故 $d_i > 0\ (i = 1,2,\cdots,n)$，即 $\boldsymbol{D} = \boldsymbol{R}_2\boldsymbol{R}_1^{-1} = \boldsymbol{I}$

- 于是 $\boldsymbol{R}_2 = \boldsymbol{R}_1$，由式(9.4.7)得到 $\boldsymbol{Q}_2 = \boldsymbol{Q}_1$

$$\boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{Q}_1 = \boldsymbol{R}_2\boldsymbol{R}_1^{-1} \qquad (9.4.7)$$

- 故分解唯一

特征值与特征向量 | QR迭代算法

设 $A=\left(a_{ij}\right)\in R^{n\times n}$ 且对 A 进行QR分解，即 $A=QR$，其中 R 为上三角阵，Q 为正交阵，于是可得到一新矩阵

$$\boldsymbol{B}=\boldsymbol{R}\boldsymbol{Q}=\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}\qquad(\boldsymbol{R}=\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A})$$

- 显然，$\boldsymbol{B}$ 是由 $\boldsymbol{A}$ 经过正交相似变换得到，因此 $\boldsymbol{B}$ 与 $\boldsymbol{A}$ 特征值相同
- 再对 $\boldsymbol{B}$ 进行QR分解，又可得一新的矩阵，重复这过程可得到矩阵序列
- 设 $\boldsymbol{A}_1=\boldsymbol{A}$，QR分解得 $\boldsymbol{A}_1=\boldsymbol{Q}_1\boldsymbol{R}_1$，作矩阵 $\boldsymbol{A}_2=\boldsymbol{R}_1\boldsymbol{Q}_1=\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{A}_1\boldsymbol{Q}_1\cdots$ ，将 $\boldsymbol{A}_k$ 进行QR分解，得 $\boldsymbol{A}_k=\boldsymbol{Q}_k\boldsymbol{R}_k$ 上，作矩阵 $\boldsymbol{A}_{k+1}=\boldsymbol{R}_k\boldsymbol{Q}_k=\boldsymbol{Q}_k^{\mathrm{T}}\boldsymbol{A}_k\boldsymbol{Q}_k\cdots$

定理9.17（基本QR方法）设 $\boldsymbol{A}_1 = \boldsymbol{A} \in \mathbf{R}^{n\times n}$ ，QR算法为

$$\begin{cases} \boldsymbol{A}_k = \boldsymbol{Q}_k\boldsymbol{R}_k \ \left(\boldsymbol{Q}_k^{\mathrm{T}}\boldsymbol{Q}_k = \boldsymbol{I}, \boldsymbol{R}_k \text{ 为上三角阵}\right) \\ \boldsymbol{A}_{k+1} = \boldsymbol{R}_k\boldsymbol{Q}_k \ (k = 1,2,\dots) \end{cases}$$

记 $\widetilde{\boldsymbol{Q}}_k \equiv \boldsymbol{Q}_1\boldsymbol{Q}_2\cdots\boldsymbol{Q}_k, \widetilde{\boldsymbol{R}}_k \equiv \boldsymbol{R}_k\cdots\boldsymbol{R}_2\boldsymbol{R}_1$ 则有

1. $\boldsymbol{A}_{k+1}$ 相似于 $\boldsymbol{A}_k$ ，即 $\boldsymbol{A}_{k+1} = \boldsymbol{Q}_k^{\mathrm{T}}\boldsymbol{A}_k\boldsymbol{Q}_k$
2. $\boldsymbol{A}_{k+1} = (\boldsymbol{Q}_1\boldsymbol{Q}_2\cdots\boldsymbol{Q}_k)^{\mathrm{T}}\boldsymbol{A}_1(\boldsymbol{Q}_1\boldsymbol{Q}_2\cdots\boldsymbol{Q}_k) = \widetilde{\boldsymbol{Q}}_k^{\mathrm{T}}\boldsymbol{A}_1\widetilde{\boldsymbol{Q}}_k$
3. $\boldsymbol{A}^k$ 的QR分解式为 $\boldsymbol{A}^k = \widetilde{\boldsymbol{Q}}_k\widetilde{\boldsymbol{R}}_k$

特征值与特征向量 | QR迭代算法

为什么要迭代？

定理9.18 设 $\boldsymbol{A}=(a_{ij})\in\mathbf{R}^{n\times n}$，如果 $\boldsymbol{A}$ 的特征值满足 $|\lambda_1|>|\lambda_2|>\cdots>|\lambda_n|>0$，且 $\boldsymbol{A}$ 有标准形 $\boldsymbol{A}=\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-1}$，其中 $\boldsymbol{D}=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$，则由QR算法产生的 $\{\boldsymbol{A}_k\}$ 本质上收敛于上三角阵，即

$$\boldsymbol{A}_k \xrightarrow{\text{本质上}} \boldsymbol{R}=\begin{bmatrix}\lambda_1 & \times & \cdots & \times\\ & \lambda_2 & \cdots & \vdots\\ & & \ddots & \times\\ & & & \lambda_n\end{bmatrix}\quad(k\to\infty)$$

- 计算量为 $O(n^4)$
- 实用方法：先采用 Householder 变换，通过相似变换，将矩阵 A 转化为 上 Hessenberg 矩阵 H，运算量 $O(n^3)$；再进行隐式 QR 迭代，每步运算量$O(n^3)$

特征值与特征向量 | QR迭代算法：原点位移

在分析QR方法收敛性时发现： $a_{nn}^{(k)} \to \lambda_n \ (k \to \infty)$ **速度依赖于比值** $r_n = |\lambda_n/\lambda_{n-1}|$

- 当 r_n 很小时，收敛较快

如果 s **为** λ_n **一个估计，对** $A - sI$ **运用QR算法**

- 则 $(n, n-1)$ 元素将以收敛因子 $\left|\frac{\lambda_n - s}{\lambda_{n-1} - s}\right|$ 线性收敛于零，(n, n) 元素将比在基本算法中收敛更快

为了加速收敛，选择数列 $\{s_k\}$ **，按下述方法构造矩阵序列** $\{A_k\}$ **称为带原点位移的QR算法**

算法流程

- 设 $\boldsymbol{A}_1 = \boldsymbol{A} \in \mathbf{R}^{n\times n}$

1. 将 $\boldsymbol{A}_k - s_k\boldsymbol{I}$ 进行QR分解，即

$$\boldsymbol{A}_k - s_k\boldsymbol{I} = \boldsymbol{Q}_k\boldsymbol{R}_k, k = 1,2,\cdots$$

2. 构造新矩阵

$$\boldsymbol{A}_{k+1} = \boldsymbol{R}_k\boldsymbol{Q}_k + s_k\boldsymbol{I} = \boldsymbol{Q}_k^{\mathrm{T}}\boldsymbol{A}_k\boldsymbol{Q}_k \qquad （与平移前一样）$$

记 $\widetilde{\boldsymbol{Q}}_k \equiv \boldsymbol{Q}_1\boldsymbol{Q}_2\cdots\boldsymbol{Q}_k, \widetilde{\boldsymbol{R}}_k \equiv \boldsymbol{R}_k\cdots\boldsymbol{R}_2\boldsymbol{R}_1$ 则有

1. $\boldsymbol{A}_{k+1} = \widetilde{\boldsymbol{Q}}_k^{\mathrm{T}}\boldsymbol{A}\widetilde{\boldsymbol{Q}}_k$
2. 矩阵 $(\boldsymbol{A} - s_1\boldsymbol{I})(\boldsymbol{A} - s_2\boldsymbol{I})\cdots(\boldsymbol{A} - s_k\boldsymbol{I}) \equiv \varphi(\boldsymbol{A})$ 有QR分解式

$$\varphi(\boldsymbol{A}) = \widetilde{\boldsymbol{Q}}_k\widetilde{\boldsymbol{R}}_k$$

特征值与特征向量 | QR迭代算法：原点位移算法

带位移QR方法变换一步的计算

- 首先用正交变换（=左变换）将$\boldsymbol{A}_k - s_k\boldsymbol{I}$化为上三角阵，即

$$\boldsymbol{Q}_k^{\mathrm{T}}(\boldsymbol{A}_k - s_k\boldsymbol{I}) = \boldsymbol{R}_k$$

其中 $\boldsymbol{Q}_k^{\mathrm{T}} = \boldsymbol{P}_{n-1}\cdots\boldsymbol{P}_2\boldsymbol{P}_1$ 为一系列平面旋转矩阵的乘积

- 接着进行右变换完成迭代

$$\boldsymbol{A}_{k+1} = \boldsymbol{P}_{n-1}\cdots\boldsymbol{P}_2\boldsymbol{P}_1(\boldsymbol{A}_k - s_k\boldsymbol{I})\boldsymbol{P}_1^{\mathrm{T}}\boldsymbol{P}_2^{\mathrm{T}}\cdots\boldsymbol{P}_{n-1}^{\mathrm{T}} + s_k\boldsymbol{I}$$

选择适当的位移策略，平均 2 到 3 步就能收敛到一个特征值，因此总迭代步数约 $2n$ 到 $3n$；叠加 Householder 变换总共只需 $O(n^3)$ 运算量

特征值与特征向量 | 总结

9.1 引言

- 特征多项式、特征值、特征向量
- 特征值的性质

9.2 幂法

- 幂法的流程、幂法的收敛性
- 规范化幂法
- 原点平移法、Rayleigh商加速法

9.3 反幂法

- 反法的流程、反幂法的收敛性、反幂法的加速

特征值与特征向量 | 总结

9.4 Householder方法

- 正交相似变换、上Hessenberg阵
- 初等反射阵、Householder方法
- 用正交相似变换约化矩阵、约化对称矩阵

9.5 QR算法

- 平面旋转矩矩阵的QR分解
- 阵、用平面旋转矩阵约化矩阵
- 基本QR方法、QR方法的收敛性
- 带原点位移的QR方法

特征值与特征向量 | 作业

第九章

习题 1，3，6，10

祝同学们期末考试顺利！