第9章 矩阵的特征值与特征向量计算

张利军
zlj@nju.edu.cn
http://cs.nju.edu.cn/zlj

目录

- 引言
- 幂法
- 反幂法
- Householder方法

引言

- 物理、力学和工程技术中的很多问题在数学上都归结为求矩阵特征值的问题
  - 振动问题（桥梁的振动、机械的振动、电磁振荡、地震引起的建筑物的振动等）
  - 物理学中某些临界值的确定问题
  - 理论物理中的一些问题

- 矩阵特征值
  - 已知$\boldsymbol{A}=(a_{ij})_{n\times n}$，要求代数方程

$$\varphi(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})=0 \tag{9.1.1}$$

的根，$\varphi(\lambda)$称为$\boldsymbol{A}$的特征多项式

引言（续）

- (9.1.1)式展开即有

$$\varphi(\lambda)=\lambda^n+c_1\lambda^{n-1}+\cdots+c_n=0$$

一般$\varphi(\lambda)$有n个零点，称为$\boldsymbol{A}$的特征值

矩阵特征向量

- 设λ为$\boldsymbol{A}$的特征值，要求相应的齐次方程组

$$(\lambda \boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=0$$

的非零解（即求$\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$的非零解）

- 上式的非零解$\boldsymbol{x}$称为矩阵$\boldsymbol{A}$的对应于λ的特征向量

特征值的性质

- **定理9.1** 如果$\lambda_i(i=1,\cdots,n)$是矩阵$\boldsymbol{A}$的特征值，则有
  1. $$\sum_{i=1}^{n}\lambda_i=\sum_{i=1}^{n}a_{ii}=\mathrm{tr}(\boldsymbol{A})$$
  2. $\det(\boldsymbol{A})=\lambda_1\lambda_2\cdots\lambda_n$

- **定理9.2** 设$\boldsymbol{A}$与$\boldsymbol{B}$为相似矩阵（即存在非奇异阵$\boldsymbol{T}$使$\boldsymbol{B}=\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$），则
  1. $\boldsymbol{A}$与$\boldsymbol{B}$有相同的特征值
  2. 若$\boldsymbol{x}$是$\boldsymbol{B}$的一个特征向量，则$\boldsymbol{T}\boldsymbol{x}$是$\boldsymbol{A}$的特征向量

特征值的性质（续）

- **定理9.3**（戈氏Gerschgorin's定理） 设$\boldsymbol{A}=(a_{ij})_{n\times n}$，则$\boldsymbol{A}$的每一个特征值必属于下述某个圆盘之中：

$$|\lambda - a_{ii}| \le \sum_{\substack{j=1 \\ j\ne i}}^{n} |a_{ij}| \quad (i=1,2,\dots,n)$$

- 设λ为$\boldsymbol{A}$的任意一个特征值，$\boldsymbol{x}$为对应的特征向量，即

$$(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = 0$$

- 记$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}} \ne \boldsymbol{0}$及$|x_i| = \max_k |x_k|$，$x_i \ne 0$

定理9.3的证明

- 从$(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = 0$的第$i$个方程

$$(\lambda - a_{ii})x_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ij}x_j$$

及$\left|x_j/x_i\right| \leq 1(j \neq i)$，得

$$|\lambda - a_{ii}| \leq \sum_{j\neq i}|a_{ij}|\left|\frac{x_j}{x_i}\right| \leq \sum_{j\neq i}|a_{ij}|$$

- 这说明λ属于复平面上以a_{ii}为圆心、$\sum_{j\neq i}|a_{ij}|$为半径的一个圆盘
- 证明过程表明：若一个特征向量的第i个分量最大，则对应的特征值一定属于第i个圆盘中

特征值的性质（续）

- **定义9.1** 设$\boldsymbol{A}$为n阶实对称矩阵，对于任一非零向量$\boldsymbol{x}$，$R(\boldsymbol{x}) = \dfrac{(\boldsymbol{Ax},\boldsymbol{x})}{(\boldsymbol{x},\boldsymbol{x})}$为对应于$\boldsymbol{x}$的Rayleigh商

- **定理9.4** 设$\boldsymbol{A} \in \mathbf{R}^{n\times n}$为对称矩阵（特征值记作$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$，对应的特征向量$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$组成规范化正交组，即$(\boldsymbol{x}_i, \boldsymbol{x}_j) = \delta_{ij}$），则

1. $\lambda_n \leq \dfrac{(\boldsymbol{Ax},\boldsymbol{x})}{(\boldsymbol{x},\boldsymbol{x})} \leq \lambda_1$（对于任何非零$\boldsymbol{x} \in \mathbf{R}^n$）

2. $\lambda_1 = \max\limits_{\substack{\boldsymbol{x}\in\mathbf{R}^n \\ \boldsymbol{x}\neq 0}} \dfrac{(\boldsymbol{Ax},\boldsymbol{x})}{(\boldsymbol{x},\boldsymbol{x})}$, $\qquad \lambda_n = \min\limits_{\substack{\boldsymbol{x}\in\mathbf{R}^n \\ \boldsymbol{x}\neq 0}} \dfrac{(\boldsymbol{Ax},\boldsymbol{x})}{(\boldsymbol{x},\boldsymbol{x})}$

定理9.4的证明

- 仅证明结论1。设$\boldsymbol{x} \neq \mathbf{0}$为$\mathbf{R}^n$中任一向量，则有展开式

$$\boldsymbol{x} = \sum_{i=1}^{n} a_i \boldsymbol{x}_i , \qquad (\boldsymbol{x}, \boldsymbol{x}) = \sum_{i=1}^{n} a_i^2 \neq 0$$

- 因此 $$\frac{(\boldsymbol{Ax}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})} = \frac{(\sum_{i=1}^{n} a_i \boldsymbol{Ax}_i , \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})} = \frac{(\sum_{i=1}^{n} a_i \lambda_i \boldsymbol{x}_i , \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})}$$

$$= \frac{(\sum_{i=1}^{n} a_i \lambda_i \boldsymbol{x}_i , \sum_{i=1}^{n} a_i \boldsymbol{x}_i)}{(\boldsymbol{x}, \boldsymbol{x})} = \frac{\sum_{i=1}^{n} a_i^2 \lambda_i}{\sum_{i=1}^{n} a_i^2}$$

- 显然

$$\lambda_n = \frac{\sum_{i=1}^{n} a_i^2 \lambda_n}{\sum_{i=1}^{n} a_i^2} \leq \frac{(\boldsymbol{Ax}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})} \leq \frac{\sum_{i=1}^{n} a_i^2 \lambda_1}{\sum_{i=1}^{n} a_i^2} = \lambda_1$$

数值算法

- 关于计算矩阵$\boldsymbol{A}$的特征值问题
  - 当$n=2,3$时，还可按行列式展开的办法求$\varphi(\lambda)=0$的根
  - 但当n较大时，如果按展开行列式的办法，首先求出$\varphi(\lambda)$的系数，再求$\varphi(\lambda)$的根，工作量就非常大了，用这种办法求矩阵特征值是不切实际的
  - 需要研究求$\boldsymbol{A}$的特征值及特征向量的数值方法
- 本章将介绍计算机上常用的两类方法
  - 一类是幂法及反幂法（迭代法）
  - 另一类是正交相似变换的方法（变换法）

目录

- 引言
- 幂法
- 反幂法
- Householder方法

研究动机

- 在一些工程、物理问题中，通常只需要求出矩阵的按模最大的特征值（称为$\boldsymbol{A}$的主特征值）和相应的特征向量，对于解这种特征值问题，应用幂法是合适的

- 幂法是一种计算实矩阵$\boldsymbol{A}$的主特征值的一种迭代法
  - 最大的优点是方法简单，对于稀疏矩阵较合适
  - 但有时收敛速度很慢

幂法

- 设实矩阵$\boldsymbol{A}=(a_{ij})_{n\times n}$有一个完全的特征向量组，其特征值为$\lambda_1,\lambda_2,\cdots,\lambda_n$，相应的特征向量为$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$
  - 已知的主特征值是实根，且满足条件

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$$

- 幂法的基本思想是任取一个非零的初始向量$\boldsymbol{v}_0$，由矩阵$\boldsymbol{A}$构造一向量序列

$$\begin{cases} \boldsymbol{v}_1 = \boldsymbol{A}\boldsymbol{v}_0 \\ \boldsymbol{v}_2 = \boldsymbol{A}\boldsymbol{v}_1 = \boldsymbol{A}^2\boldsymbol{v}_0 \\ \qquad \vdots \\ \boldsymbol{v}_{k+1} = \boldsymbol{A}\boldsymbol{v}_k = \boldsymbol{A}^{k+1}\boldsymbol{v}_0 \end{cases} \tag{9.2.2}$$

幂法（续）

- 上述迭代值被称为迭代向量
- $\boldsymbol{A} = (a_{ij})_{n\times n}$有一个完全的特征向量组，因此

$$\boldsymbol{v}_0 = a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_n\boldsymbol{x}_n \quad （设 a_1 \neq 0)$$

- 容易得到

$$\boldsymbol{v}_k = \boldsymbol{A}\boldsymbol{v}_{k-1} = \boldsymbol{A}^k\boldsymbol{v}_0$$

$$= a_1\lambda_1^k\boldsymbol{x}_1 + a_2\lambda_2^k\boldsymbol{x}_2 + \cdots + a_n\lambda_n^k\boldsymbol{x}_n$$

$$= \lambda_1^k\left[a_1\boldsymbol{x}_1 + \sum_{i=2}^{n} a_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\boldsymbol{x}_i\right]$$

$$= \lambda_1^k(a_1\boldsymbol{x}_1 + \boldsymbol{\varepsilon}_k)$$

其中$\boldsymbol{\varepsilon}_k = \sum_{i=2}^{n} a_i(\lambda_i/\lambda_1)^k\boldsymbol{x}_i$

幂法（续）

- 由假设$|\lambda_i/\lambda_1| < 1 \ (i = 2,3,\cdots,n)$，故

$$\boldsymbol{\varepsilon}_k = \sum_{i=2}^{n} a_i \left(\frac{\lambda_i}{\lambda_1}\right)^k \boldsymbol{x}_i \to \boldsymbol{0} \ (k \to \infty)$$

- 从而

$$\lim_{k\to\infty} \frac{\boldsymbol{v}_k}{\lambda_1^k} = \frac{\lambda_1^k(a_1\boldsymbol{x}_1 + \boldsymbol{\varepsilon}_k)}{\lambda_1^k} = a_1\boldsymbol{x}_1 + \boldsymbol{\varepsilon}_k \quad = a_1\boldsymbol{x}_1$$

- 这说明序列 $\frac{\boldsymbol{v}_k}{\lambda_1^k}$ 越来越接近$\boldsymbol{A}$的对应于λ_1的特征向量$\boldsymbol{x}_1$，或者说当k充分大时

$$\boldsymbol{v}_k \approx a_1\lambda_1^k\boldsymbol{x}_1 \qquad (9.2.5)$$

即迭代向量$\boldsymbol{v}_k$为$\boldsymbol{x}_1$的近似向量（除一个因子外）

幂法（续）

- 下面再考虑主特征值λ_1的计算

$$\boldsymbol{v}_k = \lambda_1^k(a_1\boldsymbol{x}_1 + \boldsymbol{\varepsilon}_k)$$

  - 用$(\boldsymbol{v}_k)_i$ 表示$\boldsymbol{v}_k$的第i个分量，则

$$\frac{(\boldsymbol{v}_{k+1})_i}{(\boldsymbol{v}_k)_i} = \frac{\lambda_1^{k+1}(a_1(\boldsymbol{x}_1)_i + (\boldsymbol{\varepsilon}_{k+1})_i)}{\lambda_1^k(a_1(\boldsymbol{x}_1)_i + (\boldsymbol{\varepsilon}_k)_i)} = \lambda_1 \frac{a_1(\boldsymbol{x}_1)_i + (\boldsymbol{\varepsilon}_{k+1})_i}{a_1(\boldsymbol{x}_1)_i + (\boldsymbol{\varepsilon}_k)_i} \quad (9.2.6)$$

  - 由于$\boldsymbol{\varepsilon}_k \to \mathbf{0}\ (k \to \infty)$，得

$$\lim_{k\to\infty} \frac{(\boldsymbol{v}_{k+1})_i}{(\boldsymbol{v}_k)_i} = \lambda_1 \frac{a_1(\boldsymbol{x}_1)_i}{a_1(\boldsymbol{x}_1)_i} = \lambda_1 \quad (9.2.7)$$

即两相邻迭代向量分量的比值收敛到主特征值

幂法（续）

$$\boldsymbol{\varepsilon}_k = \sum_{i=2}^{n} a_i \left(\frac{\lambda_i}{\lambda_1}\right)^k \boldsymbol{x}_i$$

- 这种由已知非零向量$\boldsymbol{v}_0$及矩阵$\boldsymbol{A}$的乘幂$\boldsymbol{A}^k$构造向量序列$\{\boldsymbol{v}_k\}$以计算$\boldsymbol{A}$的主特征值λ_1及相应特征向量$\boldsymbol{x}_1$的方法称为幂法

$$\boldsymbol{v}_k \approx a_1 \lambda_1^k \boldsymbol{x}_1 \tag{9.2.5}$$

$$\lim_{k\to\infty} \frac{(\boldsymbol{v}_{k+1})_i}{(\boldsymbol{v}_k)_i} = \lambda_1 \frac{a_1(\boldsymbol{x}_1)_i}{a_1(\boldsymbol{x}_1)_i} = \lambda_1 \tag{9.2.7}$$

- 根据(9.2.6)，$(\boldsymbol{v}_{k+1})_i/(\boldsymbol{v}_k)_i \to \lambda_1$的收敛速度由比值$r = \lambda_2/\lambda_1$来确定，$r$越小收敛越快，但当$r = \lambda_2/\lambda_1 \approx 1$时收敛可能就很慢

$$\frac{(\boldsymbol{v}_{k+1})_i}{(\boldsymbol{v}_k)_i} = \lambda_1 \frac{a_1(\boldsymbol{x}_1)_i + (\boldsymbol{\varepsilon}_{k+1})_i}{a_1(\boldsymbol{x}_1)_i + (\boldsymbol{\varepsilon}_k)_i} \tag{9.2.6}$$

幂法的收敛性

- **定理9.5** 设$\boldsymbol{A} \in \mathbf{R}^{n\times n}$有$n$个线性无关的特征向量，主特征值$\lambda_1$满足
$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$$
则对于任何非零初始向量$\boldsymbol{v}_0(a_1 \neq 0)$，
$$\lim_{k\to\infty} \frac{\boldsymbol{v}_k}{\lambda_1^k} = a_1\boldsymbol{x}_1 、 \lim_{k\to\infty} \frac{(\boldsymbol{v}_{k+1})_i}{(\boldsymbol{v}_k)_i} = \lambda_1$$成立

- 当$\boldsymbol{A}$的主特征值是实的重根时，定理9.5的结论还是正确的
  - 设$\boldsymbol{A}$的主特征值为实重根，即$\lambda_1 = \lambda_2 = \cdots = \lambda_r$，且$|\lambda_r| > |\lambda_{r+1}| \geq \cdots \geq |\lambda_n|$

幂法的收敛性（续）

- 设$\boldsymbol{A}$有n个线性无关的特征向量，λ_1对应的r个线性无关特征向量$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r$

- 重复$\boldsymbol{v}_k$的推导，易得

$$\boldsymbol{v}_k = \boldsymbol{A}^k \boldsymbol{v}_0 = \lambda_1^k \left\{ \sum_{i=1}^{r} a_i \boldsymbol{x}_i + \sum_{i=r+1}^{n} a_i \left(\frac{\lambda_i}{\lambda_1} \right)^k \boldsymbol{x}_i \right\}$$

- 因此，当$\sum_{i=1}^{r} a_i \boldsymbol{x}_i \neq \boldsymbol{0}$

$$\lim_{k \to \infty} \frac{\boldsymbol{v}_k}{\lambda_1^k} = \sum_{i=1}^{r} a_i \boldsymbol{x}_i + \sum_{i=r+1}^{n} a_i \left(\frac{\lambda_i}{\lambda_1} \right)^k \boldsymbol{x}_i \quad = \sum_{i=1}^{r} a_i \boldsymbol{x}_i$$

幂法的不足

- 应用幂法计算$\boldsymbol{A}$的主特征值λ_1及对应的特征向量时，如果$|\lambda_1| > 1$（或$|\lambda_1| < 1$），迭代向量$\boldsymbol{v}_k$的各个不等于零的分量将随$k \to \infty$而趋于无穷（或趋于零），这样在用计算机计算时就可能“溢出”

$$\boldsymbol{v}_k = a_1\lambda_1^k\boldsymbol{x}_1 + a_2\lambda_2^k\boldsymbol{x}_2 + \cdots + a_n\lambda_n^k\boldsymbol{x}_n$$

- 为了克服该缺点， 需要将迭代向量规范化
  - 设有一向量$\boldsymbol{v} \neq \boldsymbol{0}$，将其规范化得到向量$\boldsymbol{u} = \frac{\boldsymbol{v}}{\max(\boldsymbol{v})}$，其中$\max(\boldsymbol{v})$表示向量$\boldsymbol{v}$的绝对值最大的分量

规范化幂法

- 任取一初始向量$\boldsymbol{v}_0 \neq \boldsymbol{0}\ (a_1 \neq 0)$构造向量序列

$$
\begin{cases}
\boldsymbol{v}_1 = \boldsymbol{A}\boldsymbol{u}_0 = \boldsymbol{A}\boldsymbol{v}_0, & \boldsymbol{u}_1 = \dfrac{\boldsymbol{v}_1}{\max(\boldsymbol{v}_1)} = \dfrac{\boldsymbol{A}\boldsymbol{v}_0}{\max(\boldsymbol{A}\boldsymbol{v}_0)}, \\[2ex]
\boldsymbol{v}_2 = \boldsymbol{A}\boldsymbol{u}_1 = \dfrac{\boldsymbol{A}^2\boldsymbol{v}_0}{\max(\boldsymbol{A}\boldsymbol{v}_0)}, & \boldsymbol{u}_2 = \dfrac{\boldsymbol{v}_2}{\max(\boldsymbol{v}_2)} = \dfrac{\boldsymbol{A}^2\boldsymbol{v}_0}{\max(\boldsymbol{A}^2\boldsymbol{v}_0)}, \\[2ex]
\boldsymbol{v}_k = \dfrac{\boldsymbol{A}^k\boldsymbol{v}_0}{\max(\boldsymbol{A}^{k-1}\boldsymbol{v}_0)}, & \boldsymbol{u}_k = \dfrac{\boldsymbol{v}_k}{\max(\boldsymbol{v}_k)} = \dfrac{\boldsymbol{A}^k\boldsymbol{v}_0}{\max(\boldsymbol{A}^k\boldsymbol{v}_0)},
\end{cases}
$$

- 根据前面的推导，

$$
\boldsymbol{A}^k\boldsymbol{v}_0 = \sum_{i=1}^{n} a_i\lambda_i^k\boldsymbol{x}_i = \lambda_1^k\left[a_1\boldsymbol{x}_1 + \sum_{i=2}^{n} a_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\boldsymbol{x}_i\right]
$$

规范化幂法（续）

- 可得

$$\boldsymbol{u}_k=\frac{\boldsymbol{A}^k\boldsymbol{v}_0}{\max(\boldsymbol{A}^k\boldsymbol{v}_0)}=\frac{\lambda_1^k\left[a_1\boldsymbol{x}_1+\sum_{i=2}^n a_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\boldsymbol{x}_i\right]}{\max\left[\lambda_1^k\left(a_1\boldsymbol{x}_1+\sum_{i=2}^n a_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\boldsymbol{x}_i\right)\right]}$$

$$=\frac{\left[a_1\boldsymbol{x}_1+\sum_{i=2}^n a_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\boldsymbol{x}_i\right]}{\max\left[a_1\boldsymbol{x}_1+\sum_{i=2}^n a_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\boldsymbol{x}_i\right]}$$

- 由假设$|\lambda_i/\lambda_1|<1\ \ (i=2,3,\cdots,n)$

$$\boldsymbol{u}_k\to\frac{\boldsymbol{x}_1}{\max(\boldsymbol{x}_1)}\ \ (k\to\infty)$$

规范化幂法（续）

- 因此，规范化向量序列收敛到主特征值对应的特征向量
- 同理，可得到

$$\boldsymbol{v}_k = \frac{\boldsymbol{A}^k \boldsymbol{v}_0}{\max(\boldsymbol{A}^{k-1}\boldsymbol{v}_0)} = \frac{\lambda_1^k \left[a_1 \boldsymbol{x}_1 + \sum_{i=2}^{n} a_i \left(\frac{\lambda_i}{\lambda_1}\right)^k \boldsymbol{x}_i \right]}{\lambda_1^{k-1} \max\left[a_1 \boldsymbol{x}_1 + \sum_{i=2}^{n} a_i \left(\frac{\lambda_i}{\lambda_1}\right)^{k-1} \boldsymbol{x}_i \right]}$$

$$\max(\boldsymbol{v}_k) = \frac{\lambda_1 \max\left[a_1 \boldsymbol{x}_1 + \sum_{i=2}^{n} a_i \left(\frac{\lambda_i}{\lambda_1}\right)^k \boldsymbol{x}_i \right]}{\max\left[a_1 \boldsymbol{x}_1 + \sum_{i=2}^{n} a_i \left(\frac{\lambda_i}{\lambda_1}\right)^{k-1} \boldsymbol{x}_i \right]} \to \lambda_1 \ \ (k \to \infty)$$

收敛速度由比值$r = \lambda_2/\lambda_1$确定

规范化幂法的收敛性

- **定理9.6** 设$\boldsymbol{A} \in \mathbf{R}^{n\times n}$有$n$个线性无关的特征向量，主特征值$\lambda_1$满足$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$，则对于任何非零初始向量$\boldsymbol{v}_0 = \boldsymbol{u}_0\ (a_1 \neq 0)$，按下述方法构造的向量序列

$$\begin{cases} \boldsymbol{u}_0 = \boldsymbol{v}_0 \neq \boldsymbol{0}, \\ \boldsymbol{v}_k = \boldsymbol{A}\boldsymbol{u}_{k-1}, \\ \boldsymbol{u}_k = \dfrac{\boldsymbol{v}_k}{\max(\boldsymbol{v}_k)} \end{cases} \quad (k = 1,2,\cdots) \quad (9.2.9)$$

有

$$\lim_{k\to\infty} \boldsymbol{u}_k = \frac{\boldsymbol{x}_1}{\max(\boldsymbol{x}_1)}, \qquad \lim_{k\to\infty} \max(\boldsymbol{v}_k) = \lambda_1$$

例9.1

- 用幂法计算$\boldsymbol{A}=\begin{bmatrix}1.0 & 1.0 & 0.5\\1.0 & 1.0 & 0.25\\0.5 & 0.25 & 2.0\end{bmatrix}$的主特征值和相应的特征向量
  - 下一页的结果是用8位浮点数字进行运算得到的，$\boldsymbol{u}_k$的分量值是舍入值
  - 于是得到

$$\lambda \approx 2.5365323$$

    及相应特征向量$(0.7482, 0.6497, 1)^{\mathrm{T}}$

  - λ_1和相应的特征向量真值（8位数字）为

$$\lambda_1 = 2.5365258, \qquad \tilde{\boldsymbol{x}}_1 = (0.74822116, 0.64966116, 1)^{\mathrm{T}}$$

例9.1（续）

k	$\boldsymbol{u}_k^{\mathrm{T}}$（规范化向量）	$\max(\boldsymbol{v}_k)$
0	(1,1,1)	
1	(0.909 1, 0.818 2, 1)	2.750 000 0
5	(0.765 1, 0.667 4, 1)	2.558 791 8
10	(0.749 4, 0.650 8, 1)	2.538 002 9
15	(0.748 3, 0.649 7, 1)	2.536 625 6
16	(0.748 3, 0.649 7, 1)	2.536 584 0
17	(0.748 2, 0.649 7, 1)	2.536 559 8
18	(0.748 2, 0.649 7, 1)	2.536 545 6
19	(0.748 2, 0.649 7, 1)	2.536 537 4
20	(0.748 2, 0.649 7, 1)	2.536 532 3

加速方法

$$\boldsymbol{\varepsilon}_k = \sum_{i=2}^{n} a_i \left(\frac{\lambda_i}{\lambda_1}\right)^k \boldsymbol{x}_i$$

- 由前面讨论知道，应用幂法计算$\boldsymbol{A}$的主特征值时，其收敛速度主要由比值$r = \lambda_2/\lambda_1$来决定，但当r接近于1时，收敛可能很慢
  - 一个补救的办法是采用加速收敛的方法

1. 原点平移法
2. Rayleigh商加速法

原点平移法

- 引进矩阵$\boldsymbol{B}=\boldsymbol{A}-p\boldsymbol{I}$，其中$p$为选择参数
  - 设$\boldsymbol{A}$的特征值为$\lambda_1,\lambda_2,\cdots,\lambda_n$，则$\boldsymbol{B}$的相应特征值为$\lambda_1-p,\lambda_2-p,\cdots,\lambda_n-p$，而且$\boldsymbol{A},\boldsymbol{B}$的特征向量相同
- 如果需要计算$\boldsymbol{A}$的主特征值λ_1，就要选择适当的p使λ_1-p仍然是$\boldsymbol{B}$的主特征值，且使

$$\left|\frac{\lambda_2-p}{\lambda_1-p}\right|<\left|\frac{\lambda_2}{\lambda_1}\right|$$

  - 对$\boldsymbol{B}$应用幂法，使得在计算$\boldsymbol{B}$的主特征值λ_1-p的过程中得到加速，这种方法通常称为原点平移法
  - 对于$\boldsymbol{A}$的特征值的某种分布，它是十分有效的

例9.2

- 设$\boldsymbol{A}=\left(a_{ij}\right)_{4\times 4}$有特征值$\lambda_j=15-j\ (j=1,2,3,4)$，比值$r=\lambda_2/\lambda_1\approx 0.9$
  - 作变换

    $$\boldsymbol{B}=\boldsymbol{A}-p\boldsymbol{I}\qquad (p=12)$$

    则$\boldsymbol{B}$的特征值为$\mu_1=2,\mu_2=1,\mu_3=0,\mu_4=-1$

  - 应用幂法计算$\boldsymbol{B}$的主特征值μ_1的收敛速度的比值为

    $$\left|\frac{\mu_2}{\mu_1}\right|=\left|\frac{\lambda_2-p}{\lambda_1-p}\right|=\frac{1}{2}<\left|\frac{\lambda_2}{\lambda_1}\right|\approx 0.9$$

最优的p值

- 虽然常常能够选择有利的p值，使幂法得到加速，但设计一个自动选择适当参数p的过程是困难的
- 下面考虑当$\boldsymbol{A}$的特征值是实数时，怎样选择p使用幂法计算λ_1以得到加速
- 设$\boldsymbol{A}$的特征值满足

$$\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{n-1} > \lambda_n, \qquad (9.2.10)$$

则不管p如何选择，$\boldsymbol{B} = \boldsymbol{A} - p\boldsymbol{I}$的主特征值为$\lambda_1 - p$或$\lambda_n - p$

最优的p值（续）

- 当希望计算λ_1及$\boldsymbol{x}_1$时，应选择p使$|\lambda_1 - p| > |\lambda_n - p|$，且使收敛速度比值尽可能小

$$\omega = \max\left\{\frac{|\lambda_2 - p|}{|\lambda_1 - p|}, \frac{|\lambda_n - p|}{|\lambda_1 - p|}\right\}$$

  - $|\lambda_2 - p|$和$|\lambda_n - p|$之间必然有一个是次大的
  - 显然，当$\lambda_2 - p = -(\lambda_n - p), p = \frac{\lambda_2+\lambda_n}{2} \equiv p^*$时$\omega$最小，这时收敛速度的比值为

$$\frac{\lambda_2 - p^*}{\lambda_1 - p^*} = -\frac{\lambda_n - p^*}{\lambda_1 - p^*} \equiv \frac{\lambda_2 - \lambda_n}{2\lambda_1 - \lambda_2 - \lambda_n}$$

最优的p值（续）

- 因此，当$\boldsymbol{A}$的特征值满足$\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{n-1} > \lambda_n$且$\lambda_2$，$\lambda_n$能初步估计时，就能确定$p^*$的近似值

$$p^* = \frac{\lambda_2 + \lambda_n}{2}$$

- 当希望计算λ_n时，应选择$p = \frac{\lambda_1 + \lambda_{n-1}}{2} = p^*$，使得应用幂法计算$\lambda_n$得到加速

例9.3

- 计算例9.1矩阵$\boldsymbol{A}$的主特征值

$$\boldsymbol{A}=\begin{bmatrix}1.0 & 1.0 & 0.5\\ 1.0 & 1.0 & 0.25\\ 0.5 & 0.25 & 2.0\end{bmatrix}$$

- 作变换$\boldsymbol{B}=\boldsymbol{A}-p\boldsymbol{I}$，取$p=0.75$，则

$$\boldsymbol{B}=\begin{bmatrix}0.25 & 1 & 0.5\\ 1 & 0.25 & 0.25\\ 0.5 & 0.25 & 1.25\end{bmatrix}$$

- 对$\boldsymbol{B}$应用幂法，计算结果如下一页所示
- 迭代10次得$\boldsymbol{B}$的主特征值$\mu_1\approx 1.7865914$，$\boldsymbol{A}$的主特征值 $\lambda_1\approx\mu_1+0.75=2.5365914$

例9.3（续）

- 上述结果比例9.1迭代15次得到的结果还要好
- 若迭代15次，$\mu_1 = 1.7865258$ (相应的$\lambda_1 = 2.5365258$)，所有数值均准确

k	$\boldsymbol{u}_k^{\mathrm{T}}$（规范化向量）	$\max(\boldsymbol{v}_k)$
0	(1,1,1)	
5	(0.7516,0.6522,1)	1.7914011
6	(0.7491,0.6511,1)	1.7888443
7	(0.7488,0.6501,1)	1.7873300
8	(0.7484,0.6499,1)	1.7869152
9	(0.7483,0.6497,1)	1.7866587
10	(0.7482,0.6497,1)	1.7865914

加速方法讨论

- 原点位移的加速方法，是一个矩阵变换方法
  - 这种变换容易计算，又不破坏矩阵$\boldsymbol{A}$的稀疏性
  - 但p的选择依赖于对$\boldsymbol{A}$的特征值分布的大致了解
- 由定理9.4知，对称矩阵$\boldsymbol{A}$的λ_1及λ_n可用Rayleigh商的极值来表示
  - $R(\boldsymbol{x})=\frac{(\boldsymbol{A}\boldsymbol{x},\boldsymbol{x})}{(\boldsymbol{x},\boldsymbol{x})}$为对应于$\boldsymbol{x}$的Rayleigh商
  - $\lambda_n \le R(\boldsymbol{x}) \le \lambda_1$
  - 下面把Rayleigh商应用到用幂法计算实对称矩阵A的主特征值的加速收敛上来

Rayleigh商加速法

- **定理9.7** 设$A \in \mathbf{R}^{n\times n}$为对称阵，特征值满足$|\lambda_1| > |\lambda_2| \ge |\lambda_3| \ge \cdots \ge |\lambda_n|$，对应的特征向量$(\boldsymbol{x}_i, \boldsymbol{x}_j) = \delta_{ij}$，应用幂法计算$\boldsymbol{A}$的主特征值$\lambda_1$，则规范化向量$\boldsymbol{u}_k$的Rayleigh商给出$\lambda_1$的较好的近似，即

$$\frac{(\boldsymbol{A}\boldsymbol{u}_k, \boldsymbol{u}_k)}{(\boldsymbol{u}_k, \boldsymbol{u}_k)} = \lambda_1 + O\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}\right)$$

  - 幂法

$$\begin{cases} \boldsymbol{u}_0 = \boldsymbol{v}_0 \ne \mathbf{0}, \\ \boldsymbol{v}_k = \boldsymbol{A}\boldsymbol{u}_{k-1}, \\ \boldsymbol{u}_k = \dfrac{\boldsymbol{v}_k}{\max(\boldsymbol{v}_k)} \end{cases} \quad (k = 1,2,\cdots) \qquad (9.2.9)$$

定理9.7的证明

- 根据

$$\boldsymbol{A}^k\boldsymbol{u}_0=\boldsymbol{A}^k\boldsymbol{v}_0=\sum_{i=1}^{n}a_i\lambda_i^k\boldsymbol{x}_i$$

$$\boldsymbol{u}_k=\frac{\boldsymbol{A}^k\boldsymbol{u}_0}{\max(\boldsymbol{A}^k\boldsymbol{u}_0)}$$

- 可得 $\dfrac{(\boldsymbol{A}\boldsymbol{u}_k,\boldsymbol{u}_k)}{(\boldsymbol{u}_k,\boldsymbol{u}_k)}=\dfrac{(\boldsymbol{A}^{k+1}\boldsymbol{u}_0,\boldsymbol{A}^k\boldsymbol{u}_0)}{(\boldsymbol{A}^k\boldsymbol{u}_0,\boldsymbol{A}^k\boldsymbol{u}_0)}=\dfrac{\sum_{j=1}^n\alpha_j^2\lambda_j^{2k+1}}{\sum_{j=1}^n\alpha_j^2\lambda_j^{2k}}$

$$=\lambda_1\frac{1+\sum_{j=2}^n(\alpha_j^2/\alpha_1^2)\left(\lambda_j/\lambda_1\right)^{2k+1}}{1+\sum_{j=2}^n(\alpha_j^2/\alpha_1^2)\left(\lambda_j/\lambda_1\right)^{2k}}=\lambda_1+O\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}\right)$$

- 最后一步推导利用$1/(1+x)$的一阶展开

目录

- 引言
- 幂法
- 反幂法
- Householder方法

反幂法

1. 计算矩阵按模最小的特征值及其特征向量

- 设$\boldsymbol{A} \in \mathbf{R}^{n\times n}$为非奇异矩阵，$\boldsymbol{A}$的特征值次序记作$|\lambda_1| \ge |\lambda_2| \ge \cdots \ge |\lambda_n|$，相应的特征向量为$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$，则$\boldsymbol{A}^{-1}$的特征值为$\left|\frac{1}{\lambda_n}\right| \ge \left|\frac{1}{\lambda_{n-1}}\right| \ge \cdots \ge \left|\frac{1}{\lambda_1}\right|$，对应的特征向量为$\boldsymbol{x}_n, \boldsymbol{x}_{n-1}, \cdots, \boldsymbol{x}_1$
- 因此，计算$\boldsymbol{A}$的按模最小的特征值λ_n的问题就是计算$\boldsymbol{A}^{-1}$的按模最大的特征值问题

2. 计算对应于一个给定近似特征值的特征向量

反幂法（续）

- 对$\boldsymbol{A}^{-1}$应用幂法迭代法（称为反幂法），可求得矩阵$\boldsymbol{A}^{-1}$的主特征值$1/\lambda_n$，从而求得$\boldsymbol{A}$的按模最小的特征值λ_n

- 反幂法迭代公式为任取初始向量$\boldsymbol{v}_0=\boldsymbol{u}_0\neq\boldsymbol{0}$，构造向量序列

$$\begin{cases}\boldsymbol{v}_k=\boldsymbol{A}^{-1}\boldsymbol{u}_{k-1},\\ \boldsymbol{u}_k=\dfrac{\boldsymbol{v}_k}{\max(\boldsymbol{v}_k)},\end{cases}\qquad (k=1,2,\cdots)$$

  - 迭代向量$\boldsymbol{v}_k$可以通过解方程组$\boldsymbol{A}\boldsymbol{v}_k=\boldsymbol{u}_{k-1}$求得

反幂法的收敛性

- 定理9.8 设

  1. $\boldsymbol{A}$有n个线性无关的特征向量

  2. $\boldsymbol{A}$为非奇异矩阵且其特征值满足

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

则对任何初始非零向量$\boldsymbol{u}_0 = \boldsymbol{v}_0 (a_n \neq 0)$，由反幂法构造的向量序列$\{\boldsymbol{v}_k\}, \{\boldsymbol{u}_k\}$满足

$$\lim_{k\to\infty} \boldsymbol{u}_k = \frac{\boldsymbol{x}_n}{\max(\boldsymbol{x}_n)}, \quad \lim_{k\to\infty} \max(\boldsymbol{v}_k) = \frac{1}{\lambda_n}$$

收敛速度的比值为$\left|\dfrac{\lambda_n}{\lambda_{n-1}}\right|$

反幂法的加速

- 在反幂法中也可以用原点平移法来加速迭代过程或求其他特征值及特征向量
  - 如果矩阵$(\boldsymbol{A}-p\boldsymbol{I})^{-1}$存在，显然其特征值为$\frac{1}{\lambda_1-p}$，$\frac{1}{\lambda_2-p},\cdots,\frac{1}{\lambda_n-p}$，对应的特征向量仍然是$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$
  - 对 $(\boldsymbol{A}-p\boldsymbol{I})^{-1}$应用幂法，得到加速反幂法的迭代公式

$$\begin{cases} \boldsymbol{v}_0=\boldsymbol{u}_0\neq \boldsymbol{0}\ (\text{初始向量}) \\ \boldsymbol{v}_k=(\boldsymbol{A}-p\boldsymbol{I})^{-1}\boldsymbol{u}_{k-1} \\ \boldsymbol{u}_k=\dfrac{\boldsymbol{v}_k}{\max(\boldsymbol{v}_k)} \end{cases} \quad (k=1,2,\cdots) \qquad (9.2.12)$$

反幂法的加速（续）

- 如果p是$\boldsymbol{A}$的特征值λ_j的一个近似值，且$|\lambda_j - p| < |\lambda_i - p|\ (i \neq j)$，即$\frac{1}{\lambda_j - p}$是$(\boldsymbol{A} - p\boldsymbol{I})^{-1}$的主特征值，可用反幂法式(9.2.12)计算其特征量及特征向组

- 设$\boldsymbol{A} \in \mathbf{R}^{n\times n}$有$n$个线性无关的特征向量$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$，则

$$\boldsymbol{u}_0 = \sum_{i=1}^{n} a_i \boldsymbol{x}_i (a_i \neq 0)$$

$$\boldsymbol{v}_k = \frac{(\boldsymbol{A} - p\boldsymbol{I})^{-k}\boldsymbol{u}_0}{\max((\boldsymbol{A} - p\boldsymbol{I})^{-(k-1)}\boldsymbol{u}_0)} \quad \boldsymbol{u}_k = \frac{(\boldsymbol{A} - p\boldsymbol{I})^{-k}\boldsymbol{u}_0}{\max((\boldsymbol{A} - p\boldsymbol{I})^{-k}\boldsymbol{u}_0)}$$

其中$(\boldsymbol{A} - p\boldsymbol{I})^{-k}\boldsymbol{u}_0 = \sum_{i=1}^{n} a_i(\lambda_i - p)^{-k}\boldsymbol{x}_i$

加速反幂法的收敛性

- **定理9.9** 设
  1. $\boldsymbol{A} \in \mathbf{R}^{n\times n}$有$n$个线性无关的特征向量，$\boldsymbol{A}$的特征值及对应的特征向量记作$\lambda_i$及$\boldsymbol{x}_i \ (i = 1,2,\dots,n)$
  2. p是$\boldsymbol{A}$的特征值λ_j的近似值，$(\boldsymbol{A} - p\boldsymbol{I})^{-1}$存在，且$\left|\lambda_j - p\right| < |\lambda_i - p| \ (i \neq j)$
  3. $\boldsymbol{u}_0 = \sum_{i=1}^{n} a_i \boldsymbol{x}_i \neq \boldsymbol{0}$ 为给定的初始向量$\left(a_j \neq 0\right)$

  则由反幂法迭代公式(9.2.12)构造的向量序列$\{\boldsymbol{v}_k\},\{\boldsymbol{u}_k\}$满足

加速反幂法的收敛性（续）

$$\lim_{k\to\infty} \boldsymbol{u}_k = \frac{\boldsymbol{x}_j}{\max(\boldsymbol{x}_j)}$$

$$\lim_{k\to\infty} \max(\boldsymbol{v}_k) = \frac{1}{\lambda_j - p},\ 即\ p + \frac{1}{\max(\boldsymbol{v}_k)} \to \lambda_j\ （当 k \to \infty）$$

且收敛速度由比值$r = \max\limits_{i\neq j}\left|\frac{\lambda_j - p}{\lambda_i - p}\right|$确定

- 对$\boldsymbol{A} - p\boldsymbol{I}(p \approx \lambda_j)$应用反幂法，可计算特征向量$\boldsymbol{x}_j$
- 只要选择的p是λ_j的一个较好的近似且特征值分离情况较好，一般r很小，常常只要迭代一两次就可完成特征向量的计算

反幂法的计算加速

- 反幂法迭代公式中的$\boldsymbol{v}_k$是通过解方程组$(\boldsymbol{A}-p\boldsymbol{I})\boldsymbol{v}_k=\boldsymbol{u}_{k-1}$求得的
  - 为了节省工作量，可以先将$(\boldsymbol{A}-p\boldsymbol{I})$进行三角分解，即

$$\boldsymbol{P}(\boldsymbol{A}-p\boldsymbol{I})=\boldsymbol{L}\boldsymbol{U}$$

其中$\boldsymbol{P}$为某个置换矩阵

- 求$\boldsymbol{v}_k$相当于解两个三角形方程组

$$\boldsymbol{L}\boldsymbol{y}_k=\boldsymbol{P}\boldsymbol{u}_{k-1}$$

$$\boldsymbol{U}\boldsymbol{v}_k=\boldsymbol{y}_k$$

反幂法的计算加速（续）

- 反幂法迭代公式可写为

$$\begin{cases} \boldsymbol{L}\boldsymbol{y}_k = \boldsymbol{P}\boldsymbol{u}_{k-1} \\ \boldsymbol{U}\boldsymbol{v}_k = \boldsymbol{y}_k \\ \boldsymbol{u}_k = \dfrac{\boldsymbol{v}_k}{\max(\boldsymbol{v}_k)} \end{cases} \quad (k = 1,2,\cdots) \qquad (9.2.13)$$

- 实验表明，按下述方法$\boldsymbol{v}_0 = \boldsymbol{u}_0$是较好的：选$\boldsymbol{u}_0$使

$$\boldsymbol{U}\boldsymbol{v}_1 = \boldsymbol{y}_1 = \boldsymbol{L}^{-1}\boldsymbol{P}\boldsymbol{u}_0 = (1,1,\cdots,1)^{\mathrm{T}} \quad (9.2.14)$$

  - 用回代求解式(9.2.14) 即得$\boldsymbol{v}_1$，然后再按式(9.2.13)迭代，因此并不需要计算$\boldsymbol{u}_0$

例9.4

- 用反幂法求

$$\boldsymbol{A}=\begin{bmatrix}2&1&0\\1&3&1\\0&1&4\end{bmatrix}$$

的对应计算特征值$\lambda=1.2679$（精确特征值为$\lambda_3=3-\sqrt{3}$）的特征向量（用5位浮点数进行运算）

- 用部分选主元的三角分解将$\boldsymbol{A}-p\boldsymbol{I}$（其中$p=1.2679$）分解为 $\boldsymbol{P}(\boldsymbol{A}-p\boldsymbol{I})=\boldsymbol{L}\boldsymbol{U}$

$$\boldsymbol{P}=\begin{bmatrix}0&1&0\\0&0&1\\1&0&0\end{bmatrix}\qquad \boldsymbol{L}=\begin{bmatrix}1&0&0\\0&1&0\\0.7321&-0.26807&1\end{bmatrix}$$

例9.4（续）

$$\boldsymbol{U} = \begin{bmatrix} 1 & 1.7321 & 1 \\ 0 & 1 & 2.7321 \\ 0 & 0 & 0.29405 \times 10^{-3} \end{bmatrix}$$

- 由$\boldsymbol{U}\boldsymbol{v}_1 = (1,1,1)^{\mathrm{T}}$, 得

$$\boldsymbol{v}_1 = (12692, -9290.3, 3400.8)^{\mathrm{T}}$$

进而 $\boldsymbol{u}_1 = (1, -0.73198, 0.26795)^{\mathrm{T}}$

- 由$\boldsymbol{L}\boldsymbol{U}\boldsymbol{v}_2 = \boldsymbol{P}\boldsymbol{u}_1$, 得

$$\boldsymbol{v}_2 = (20404, -14937, 5467.4)^{\mathrm{T}}$$

$$\boldsymbol{u}_2 = (1, -0.73206, 0.26796)^{\mathrm{T}}$$

非常接近

- λ_3对应的特征向量是

$$\boldsymbol{x}_3 = (1, 1-\sqrt{3},\ 2-\sqrt{3})^{\mathrm{T}} \approx (1, -0.73205, 0.26795)^{\mathrm{T}}$$

目录

- 引言
- 幂法
- 反幂法
- Householder方法

研究动机

- 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

- 若要求求出所有特征值及其特征向量，应该用什么方法呢？

- 下面将讨论的以正交相似变换为基础的一类方法即是解决这类问题的方法

- 首先，讨论对于一般实矩阵$\boldsymbol{A} \in \mathbf{R}^{n\times n}$利用正交相似变换约化到什么程度的问题

正交相似变换

- **定理9.10** 设$\boldsymbol{A} \in \mathbf{R}^{n\times n}$，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

- 每一个一阶对角块即为$\boldsymbol{A}$的实特征值
- 每一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》

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
  1. 用正交相似变换约化一般实矩阵为上Hessenberg阵
  2. 用正交相似变换约化对称阵为对称三对角阵
  - 这样，就变成求转换后矩阵的特征值问题

初等反射阵

- **定义9.3** 设向量$\boldsymbol{w}$满足$\|\boldsymbol{w}\|_2 = 1$，矩阵$\boldsymbol{H} = \boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}$称为初等反射阵，记作$\boldsymbol{H}(\boldsymbol{w})$，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，$\boldsymbol{w} = (w_1, w_2, \cdots, w_n)^{\mathrm{T}}$

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

Householder方法

- 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$\boldsymbol{H}$使$\boldsymbol{H}\boldsymbol{a} = \sigma \boldsymbol{e}_1$，这种约化矩阵的方法称为Householder方法

- **推论** 设向量$\boldsymbol{x} \in \mathbf{R}^n$ $(\boldsymbol{x} \neq \boldsymbol{0})$，$\sigma = \pm\|\boldsymbol{x}\|_2$，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$\boldsymbol{H}\boldsymbol{x} = -\sigma \boldsymbol{e}_1, \boldsymbol{u} = \boldsymbol{x} + \sigma \boldsymbol{e}_1, \rho = \|\boldsymbol{u}\|_2^2/2$

用正交相似变换约化矩阵

- 设

$$\boldsymbol{A}=\left[\begin{array}{c:ccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \hdashline a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{array}\right] \equiv \begin{bmatrix} a_{11} & \boldsymbol{A}_{12}^{(1)} \\ \boldsymbol{a}_{21}^{(1)} & \boldsymbol{A}_{22}^{(1)} \end{bmatrix}$$

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

- 根据钱、前一页的推论，上述思路可行

用正交相似变换约化矩阵（续）

- 令$\boldsymbol{U}_1=\begin{bmatrix}\boldsymbol{I} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{R}_1\end{bmatrix}$，则

$$\boldsymbol{A}_2=\boldsymbol{U}_1\boldsymbol{A}_1\boldsymbol{U}_1=\begin{bmatrix} a_{11} & \boldsymbol{A}_{12}^{(1)}\boldsymbol{R}_1\\ \boldsymbol{R}_1\boldsymbol{a}_{21}^{(1)} & \boldsymbol{R}_1\boldsymbol{A}_{22}^{(1)}\boldsymbol{R}_1\end{bmatrix}$$

$$=\begin{bmatrix} \boldsymbol{A}_{11}^{(2)} & \boldsymbol{a}_{21}^{(2)} & \boldsymbol{A}_{13}^{(2)}\\ \boldsymbol{O} & \boldsymbol{a}_{22}^{(2)} & \boldsymbol{A}_{23}^{(2)}\end{bmatrix}$$

其中

$\boldsymbol{A}_{11}^{(2)}\in\mathbf{R}^{2\times1}$， $\boldsymbol{a}_{22}^{(2)}\in\mathbf{R}^{n-2}$， $\boldsymbol{A}_{23}^{(2)}\in\mathbf{R}^{(n-2)\times(n-2)}$

用正交相似变换约化矩阵（续）

- 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

$$
\boldsymbol{A}_k=\boldsymbol{U}_{k-1}\boldsymbol{A}_{k-1}\boldsymbol{U}_{k-1}=\left[\begin{array}{cccc:c:ccc}
a_{11}^{(1)} & a_{12}^{(2)} & \cdots & & a_{1k}^{(k)} & a_{1,k+1}^{(k)} & \cdots & a_{1n}^{(k)} \\
-\sigma_1 & a_{22}^{(2)} & \cdots & & a_{2k}^{(k)} & a_{2,k+1}^{(k)} & \cdots & a_{2n}^{(k)} \\
 & \ddots & \ddots & & \vdots & \vdots & & \vdots \\
 & & -\sigma_{k-1} & & a_{kk}^{(k)} & a_{k,k+1}^{(k)} & \cdots & a_{k,n}^{(k)} \\
\hdashline
 & & & & {\color{red}a_{k+1,k}^{(k)}} & a_{k+1,k+1}^{(k)} & \cdots & a_{k+1,n}^{(k)} \\
 & & & & {\color{red}\vdots} & \vdots & & \vdots \\
 & & & & {\color{red}a_{nk}^{(k)}} & a_{n,k+1}^{(k)} & \cdots & a_{nn}^{(k)}
\end{array}\right]
$$

用正交相似变换约化矩阵（续）

- 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

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

其中

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

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

用正交相似变换约化矩阵（续）

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

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

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

- 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

用正交相似变换约化矩阵（续）

- 重复这过程，直到

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

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

用正交相似变换约化矩阵（续）

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

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

特征值求解

- 由于$\boldsymbol{U}_k$都是正交阵，所以$\boldsymbol{A}_1 \sim \boldsymbol{A}_2 \sim \cdots \sim \boldsymbol{A}_{n-1}$，求$\boldsymbol{A}$的特征值问题，就转化为求上Hessenberg阵$\boldsymbol{C}$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$\lambda$的特征向量

总结

- 引言
  - 特征多项式、特征值、特征向量
- 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
- 反幂法
  - 反法的流程、反幂法的收敛性、反幂法的加速
- Householder方法
  - 正交相似变换、上Hessenberg阵
  - 初等反射阵、Householder方法