

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 | 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 $A = (a_{ij})_{n \times n}$ ，求代数方程

$$\varphi(\lambda) = \det(\lambda I - A) = 0 \quad (9.1.1)$$

的根， $\varphi(\lambda)$ 称为 A 的**特征多项式**

特征值与特征向量 | 引言

- (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 A 的**特征值**

矩阵特征向量

- 设 λ 为 A 的特征值，要求相应的齐次方程组

$$(\lambda I - A)x = 0$$

的非零解（即求 $Ax = \lambda x$ 的非零解）

- 上式的非零解 x 称为矩阵 A 的对应于 λ 的**特征向量**

特征值与特征向量 | 特征值的性质

定理9.1 如果 λ_i ($i = 1, \dots, n$) 是矩阵 A 的特征值, 则有

- 矩阵的迹

$$\sum_{i=1}^n \lambda_i = \sum_{i=1}^n a_{ii} = \text{tr}(A)$$

- 矩阵的行列式

$$\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$$

定理9.2 设 A 与 B 为相似矩阵 (即存在非奇异阵 T 使 $B = T^{-1}AT$), 则

- A 与 B 有相同的特征值
- 若 x 是 B 的一个特征向量, 则 Tx 是 A 的特征向量

特征值与特征向量 | 特征值的性质

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

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

- 设 λ 为 A 的任意一个特征值 , x 为对应的特征向量 , 即

$$(\lambda I - A)x = 0$$

- 记 $x = (x_1, x_2, \dots, x_n)^T \neq \mathbf{0}$ 及 $|x_i| = \max_k |x_k|, x_i \neq 0$

特征值与特征向量 | 特征值的性质

- 从 $(\lambda I - A)x = 0$ 的第 i 个方程

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

及 $|x_j/x_i| \leq 1$ ($j \neq i$) ,

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

- 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

特征值与特征向量 | 特征值的性质

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

定理9.4 设 $A \in \mathbf{R}^{n \times n}$ 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 x_1, x_2, \cdots, x_n 组成规范化正交组，即 $(x, x_j) = \delta_{ij}$ ），则

$$\lambda_n \leq \frac{(Ax, x)}{(x, x)} \leq \lambda_1 \quad (\text{对于任何非零 } x \in \mathbf{R}^n)$$

$$\lambda_1 = \max_{\substack{x \in \mathbf{R}^n \\ x \neq 0}} \frac{(Ax, x)}{(x, x)}, \quad \lambda_n = \min_{\substack{x \in \mathbf{R}^n \\ x \neq 0}} \frac{(Ax, x)}{(x, x)}$$

特征值与特征向量 | 特征值的性质

证明结论1：

- 设 $x \neq 0$ 为 \mathbf{R}^n 中任一向量，则有展开式

$$x = \sum_{i=1}^n a_i x_i, \quad (x, x) = \sum_{i=1}^n a_i^2 \neq 0$$

- 因此

$$\begin{aligned} \frac{(Ax, x)}{(x, x)} &= \frac{(\sum_{i=1}^n a_i Ax_i, x)}{(x, x)} = \frac{(\sum_{i=1}^n a_i \lambda_i x_i, x)}{(x, x)} \\ &= \frac{(\sum_{i=1}^n a_i \lambda_i x_i, \sum_{i=1}^n a_i x_i)}{(x, x)} = \frac{\sum_{i=1}^n a_i^2 \lambda_i}{\sum_{i=1}^n a_i^2} \end{aligned}$$

- 显然

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

特征值与特征向量 | 数值算法

关于计算矩阵 A 的特征值问题

- 当 $n = 2, 3$ 时，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，工作量非常大不切实际
- 需要研究求 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）

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 | 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 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 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

特征值与特征向量 | 幂法

设实矩阵 $A = (a_{ij})_{n \times n}$ 有一个完全的特征向量组，其特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$ ，相应的特征向量为 x_1, x_2, \dots, x_n

- 已知的主特征值是实根，且满足条件

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

幂法的基本思想是任取一个非零的初始向量 v_0 ，由矩阵 A 构造一向量序列

$$\begin{cases} v_1 = Av_0 \\ v_2 = Av_1 = A^2v_0 \\ \vdots \\ v_{k+1} = Av_k = A^{k+1}v_0 \end{cases} \quad (9.2.2)$$

特征值与特征向量 | 幂法

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

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

- 容易得到

$$\begin{aligned} \boldsymbol{v}_k &= A \boldsymbol{v}_{k-1} = A^k \boldsymbol{v}_0 \\ &= a_1 \lambda_1^k \boldsymbol{x}_1 + a_2 \lambda_2^k \boldsymbol{x}_2 + \cdots + a_n \lambda_n^k \boldsymbol{x}_n \\ &= \lambda_1^k \left[a_1 \boldsymbol{x}_1 + \sum_{i=2}^n a_i \left(\frac{\lambda_i}{\lambda_1} \right)^k \boldsymbol{x}_i \right] \\ &= \lambda_1^k (a_1 \boldsymbol{x}_1 + \boldsymbol{\varepsilon}_k) \end{aligned}$$

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

特征值与特征向量 | 幂法

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

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

- 从而

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

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

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

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

特征值与特征向量 | 幂法

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 $\boldsymbol{\varepsilon}_k \rightarrow \mathbf{0} \ (k \rightarrow \infty)$, 得

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

即两相邻迭代向量分量的比值收敛到主特征值

特征值与特征向量 | 幂法

下面再考虑这种由已知非零向量 v_0 及矩阵 A 的乘幂 A^k 构造向量序列 $\{v_k\}$ 以计算 A 的主特征值 λ_1 及相应特征向量 x_1 的方法称为**幂法**

$$v_k \approx a_1 \lambda_1^k x_1 \quad (9.2.5)$$

$$\lim_{k \rightarrow \infty} \frac{(v_{k+1})_i}{(v_k)_i} = \lambda_1 \frac{a_1 (x_1)_i}{a_1 (x_1)_i} = \lambda_1 \quad (9.2.7)$$

根据(9.2.6) , $(v_{k+1})_i / (v_k)_i \rightarrow \lambda_1$ 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

$$\frac{(v_{k+1})_i}{(v_k)_i} = \lambda_1 \frac{a_1 (x_1)_i + (\varepsilon_{k+1})_i}{a_1 (x_1)_i + (\varepsilon_k)_i} \quad (9.2.6)$$

$$\varepsilon_k = \sum_{i=2}^n a_i \left(\frac{\lambda_i}{\lambda_1} \right)^k x_i$$

特征值与特征向量 | 幂法的收敛性

定理9.5 设 $A \in R^{n \times n}$ 有 n 个线性无关的特征向量，主特征值 λ_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 $v_0 (a_1 \neq 0)$ ， $\lim_{k \rightarrow \infty} \frac{v_k}{\lambda_1^k} = a_1 x_1$ 、 $\lim_{k \rightarrow \infty} \frac{(v_{k+1})_i}{(v_k)_i} = \lambda_1$ 成立

当 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 A 的主特征值为实重根，即 $\lambda_1 = \lambda_2 = \cdots = \lambda_r$ ，且 $|\lambda_r| > |\lambda_{r+1}| \geq \cdots \geq |\lambda_n|$

特征值与特征向量 | 幂法的收敛性

- 设 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 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r$
- 重复 \mathbf{v}_k 的推导, 易得

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

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

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

特征值与特征向量 | 幂法的不足

应用幂法计算 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 v_k 的各个不等于零的分量将随 $k \rightarrow \infty$ 而趋于无穷（或趋于零），这样在用计算机计算时就可能“溢出”

$$v_k = a_1 \lambda_1^k x_1 + a_2 \lambda_2^k x_2 + \cdots + a_n \lambda_n^k x_n$$

怎么办？

- **Normalization**：设有一向量 $v \neq 0$ ，将其规范化得到向量 $u = \frac{v}{\max(v)}$ ，其中 $\max(v)$ 表示向量 v 的绝对值最大的分量

特征值与特征向量 | 规范化幂法

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

$$\boldsymbol{v}_1 = \boldsymbol{A}\boldsymbol{u}_0 = \boldsymbol{A}\boldsymbol{v}_0, \quad \boldsymbol{u}_1 = \frac{\boldsymbol{v}_1}{\max(\boldsymbol{v}_1)} = \frac{\boldsymbol{A}\boldsymbol{v}_0}{\max(\boldsymbol{A}\boldsymbol{v}_0)},$$

$$\boldsymbol{v}_2 = \boldsymbol{A}\boldsymbol{u}_1 = \frac{\boldsymbol{A}^2\boldsymbol{v}_0}{\max(\boldsymbol{A}\boldsymbol{v}_0)}, \quad \boldsymbol{u}_2 = \frac{\boldsymbol{v}_2}{\max(\boldsymbol{v}_2)} = \frac{\boldsymbol{A}^2\boldsymbol{v}_0}{\max(\boldsymbol{A}^2\boldsymbol{v}_0)},$$

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

- 根据前面的推导，

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

特征值与特征向量 | 规范化幂法

- 可得

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

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

$$\mathbf{u}_k \rightarrow \frac{\mathbf{x}_1}{\max(\mathbf{x}_1)} \quad (k \rightarrow \infty)$$

特征值与特征向量 | 规范化幂法

- 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

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

特征值与特征向量 | 规范化幂法

定理9.6 设 $A \in R^{n \times n}$ 有 n 个线性无关的特征向量，主特征值 λ_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

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

有

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

特征值与特征向量 | 规范化幂法

例9.1 用幂法计算 $A = \begin{bmatrix} 1.0 & 1.0 & 0.5 \\ 1.0 & 1.0 & 0.25 \\ 0.5 & 0.25 & 2.0 \end{bmatrix}$ 的主特征值和相应的特征向量

- λ_1 和相应的特征向量真值 (8 位数字) 为

$$\lambda_1 = 2.5365258, \quad \tilde{\mathbf{x}}_1 = (0.74822116, 0.64966116, 1)^T$$

- 可迭代计算得到

$$\lambda \approx 2.5365323$$

及相应特征向量 $(0.7482, 0.6497, 1)^T$

- 计算过程如下一页

特征值与特征向量 | 规范化幂法

k	\mathbf{u}_k^T (规范化向量)	$\max(\mathbf{v}_k)$
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

特征值与特征向量 | 加速方法

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

由前面讨论知道，应用幂法计算 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

尝试加速

- 原点平移法
- Rayleigh商加速法

特征值与特征向量 | 加速方法：原点平移法

引进矩阵 $B = A - pI$, 其中 p 为选择参数

- 设 A 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 则 B 的相应特征值为 $\lambda_1 - p, \lambda_2 - p, \dots, \lambda_n - p$, 而且 A, B 的特征向量相同

如果需要计算 A 的主特征值 λ_1 , 就要选择适当的 p 使 $\lambda_1 - p$ 仍然是 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 B 应用幂法 , 使得在计算 B 的主特征值 $\lambda_1 - p$ 的过程中得到加速 , 这种方法通常称为原点平移法
- 对于 A 的特征值的某种分布 , 它是十分有效的

特征值与特征向量 | 加速方法：原点平移法

例9.2 设 $A = (a_{ij})_{4 \times 4}$ 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

$$B = A - pI \quad (p = 12)$$

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

- 应用幂法计算 B 的主特征值 μ_1 的收敛速度的比值为

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

特征值与特征向量 | 加速方法：原点平移法

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 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 A 的特征值满足

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

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

- **当希望计算 λ_1 及 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

特征值与特征向量 | 加速方法：原点平移法

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 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

- 显然，当 $\lambda_2 - p = -(\lambda_n - p), p = \frac{\lambda_2 + \lambda_n}{2} \equiv p^*$ 时 ω 最小，这时收敛速度的比值为

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

- 因此，当 A 的特征值满足 $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_{n-1} > \lambda_n$ 且 λ_2, λ_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 λ_n 得到加速

特征值与特征向量 | 加速方法：原点平移法

例9.3 计算例9.1矩阵 A 的主特征值

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

- 作变换 $B = A - pI$, 取 $p = 0.75$, 则

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

- 对 B 应用幂法 , 迭代 10 次得 B 的主特征值 $\mu_1 \approx 1.7865914$, A 的主特征值

$$\lambda_1 \approx \mu_1 + 0.75 = 2.5365914$$

特征值与特征向量 | 加速方法：原点平移法

- 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	\mathbf{u}_k^T (规范化向量)	$\max(\mathbf{v}_k)$
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

特征值与特征向量 | 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 A 的稀疏性
- 但 p 的选择依赖于对 A 的特征值分布的大致了解

由定理9.4知，对称矩阵 A 的 λ_1 及 λ_n 可用Rayleigh商的极值来表示

- $R(\mathbf{x}) = \frac{(\mathbf{Ax}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})}$ 为对应于 \mathbf{x} 的Rayleigh商
- $\lambda_n \leq R(\mathbf{x}) \leq \lambda_1$
- 可将把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

特征值与特征向量 | 加速方法讨论

定理9.7 设 $A \in \mathbf{R}^{n \times n}$ 为对称阵，特征值满足 $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$ ，对应的特征向量 $(\mathbf{x}_i, \mathbf{x}_j) = \delta_{ij}$ ，应用幂法计算 A 的主特征值 λ_1 ，则规范化向量 \mathbf{u}_k 的Rayleigh商给出 λ_1 的较好的近似，即

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

- 幂法

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

特征值与特征向量 | 加速方法讨论

- 根据

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

- 可得

$$\begin{aligned} \frac{(A\mathbf{u}_k, \mathbf{u}_k)}{(\mathbf{u}_k, \mathbf{u}_k)} &= \frac{(A^{k+1}\mathbf{u}_0, A^k\mathbf{u}_0)}{(A^k\mathbf{u}_0, A^k\mathbf{u}_0)} = \frac{\sum_{j=1}^n \alpha_j^2 \lambda_j^{2k+1}}{\sum_{j=1}^n \alpha_j^2 \lambda_j^{2k}} \\ &= \lambda_1 \frac{1 + \sum_{j=2}^n (\alpha_j^2 / \alpha_1^2) (\lambda_j / \lambda_1)^{2k+1}}{1 + \sum_{j=2}^n (\alpha_j^2 / \alpha_1^2) (\lambda_j / \lambda_1)^{2k}} = \lambda_1 + O\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}\right) \end{aligned}$$

- 最后一步推导利用 $1/(1+x)$ 的一阶展开

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 | 反幂法

计算矩阵按模最小的特征值及其特征向量，设 $A \in \mathbb{R}^{n \times n}$ 为非奇异矩阵

- A 的特征值次序记作 $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$ ，相应的特征向量为 $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ ；
- A^{-1} 的特征值为 $\left| \frac{1}{\lambda_n} \right| \geq \left| \frac{1}{\lambda_{n-1}} \right| \geq \cdots \geq \left| \frac{1}{\lambda_1} \right|$ ，对应的特征向量为 $\mathbf{x}_n, \mathbf{x}_{n-1}, \cdots, \mathbf{x}_1$
- 计算 A 的按模最小的特征值 λ_n 的问题就是计算 A^{-1} 的按模最大的特征值问题

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

特征值与特征向量 | 反幂法

反幂法迭代公式为任取初始向量 $v_0 = u_0 \neq 0$, 构造向量序列

$$\begin{cases} v_k = A^{-1}u_{k-1}, \\ u_k = \frac{v_k}{\max(v_k)}, \end{cases} \quad (k = 1, 2, \dots)$$

- 迭代向量 v_k 可以通过解方程组 $Av_k = u_{k-1}$ 求得

特征值与特征向量 | 反幂法

定理9.8 设 1) A 有 n 个线性无关的特征向量；2) 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 $\mathbf{u}_0 = \mathbf{v}_0 (a_n \neq 0)$ ，由反幂法构造的向量序列 $\{\mathbf{v}_k\}, \{\mathbf{u}_k\}$ 满足

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

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

特征值与特征向量 | 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 $(A - pI)^{-1}$ 存在，显然其特征值为 $\frac{1}{\lambda_1 - p}, \frac{1}{\lambda_2 - p}, \dots, \frac{1}{\lambda_n - p}$ ，对应的特征向量仍然是 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$
- 对 $(A - pI)^{-1}$ 应用幂法，得到加速反幂法的迭代公式

$$\begin{cases} \mathbf{v}_0 = \mathbf{u}_0 \neq \mathbf{0} \text{ (初始向量)} \\ \mathbf{v}_k = (A - pI)^{-1} \mathbf{u}_{k-1} \\ \mathbf{u}_k = \frac{\mathbf{v}_k}{\max(\mathbf{v}_k)} \end{cases} \quad (k = 1, 2, \dots) \quad (9.2.12)$$

特征值与特征向量 | 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 p 是 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 $(A - pI)^{-1}$ 的主特征值, 可用反幂法式(9.2.12)计算其特征量及特征向量组
- 设 $A \in \mathbf{R}^{n \times n}$ 有 n 个线性无关的特征向量 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, 则

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

$$\mathbf{v}_k = \frac{(A - pI)^{-k} \mathbf{u}_0}{\max((A - pI)^{-(k-1)} \mathbf{u}_0)} \quad \mathbf{u}_k = \frac{(A - pI)^{-k} \mathbf{u}_0}{\max((A - pI)^{-k} \mathbf{u}_0)}$$

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

特征值与特征向量 | 加速反幂法的收敛性

定理9.9 设

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

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

$$\lim_{k \rightarrow \infty} \mathbf{u}_k = \frac{\mathbf{x}_j}{\max(\mathbf{x}_j)}, \quad \lim_{k \rightarrow \infty} \max(\mathbf{v}_k) = \frac{1}{\lambda_j - p}, \quad p + \frac{1}{\max(\mathbf{v}_k)} \rightarrow \lambda_j \quad (\text{当 } k \rightarrow \infty)$$

特征值与特征向量 | 加速反幂法的收敛性

定理9.9 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

$$\lim_{k \rightarrow \infty} \boldsymbol{u}_k = \frac{\boldsymbol{x}_j}{\max(\boldsymbol{x}_j)}, \quad \lim_{k \rightarrow \infty} \max(\boldsymbol{v}_k) = \frac{1}{\lambda_j - p}, \quad p + \frac{1}{\max(\boldsymbol{v}_k)} \rightarrow \lambda_j \quad (\text{当 } k \rightarrow \infty)$$

- 收敛速度由比值 $r = \max_{i \neq j} \left| \frac{\lambda_j - p}{\lambda_i - p} \right|$ 确定
- 对 $A - pI$ ($p \approx \lambda_j$) 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

特征值与特征向量 | 反幂法的计算加速

反幂法迭代公式中的 v_k 是通过解方程组 $(A - pI)v_k = u_{k-1}$ 求得的

- 为了节省工作量，可以先将 $(A - pI)$ 进行三角分解，即

$$P(A - pI) = LU$$

其中 P 为某个置换矩阵

- 求 v_k 相当于解两个三角形方程组

$$Ly_k = Pu_{k-1}$$

$$Uv_k = y_k$$

特征值与特征向量 | 反幂法的计算加速

反幂法迭代公式可写为

$$\begin{cases} Ly_k = Pu_{k-1} \\ Uv_k = y_k \\ u_k = \frac{v_k}{\max(v_k)} \end{cases} \quad (k = 1, 2, \dots) \quad (9.2.13)$$

Initialization : 实验表明 , 按下述方法 $v_0 = u_0$ 是较好的 : 选 u_0 使

$$Uv_1 = y_1 = L^{-1}Pu_0 = (1, 1, \dots, 1)^T \quad (9.2.14)$$

- 用回代求解式(9.2.14) 即得 v_1 , 然后再按式(9.2.13)迭代 , 因此并不需要计算 u_0

特征值与特征向量 | 反幂法的计算加速

例9.4 用反幂法求

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

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 $A - pI$ (其中 $p = 1.2679$) 分解为

$$P(A - pI) = LU$$

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

特征值与特征向量 | 反幂法的计算加速

例9.4 用反幂法求

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

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 $A - pI$ (其中 $p = 1.2679$) 分解为

$$P(A - pI) = LU$$

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.7321 & -0.26807 & 1 \end{bmatrix} \quad U = \begin{bmatrix} 1 & 1.7321 & 1 \\ 0 & 1 & 2.7321 \\ 0 & 0 & 0.29405 \times 10^{-3} \end{bmatrix}$$

特征值与特征向量 | 反幂法的计算加速

- 由 $Uv_1 = (1, 1, 1)^T$, 得

$$v_1 = (12692, -9290.3, 3400.8)^T$$

进而

$$u_1 = (1, -0.73198, 0.26795)^T$$

- 由 $LUv_2 = Pu_1$, 得

$$v_2 = (20404, -14937, 5467.4)^T$$

$$u_2 = (1, -0.73206, 0.26796)^T$$

- λ_3 对应的特征向量是

$$x_3 = (1, 1 - \sqrt{3}, 2 - \sqrt{3})^T \approx (1, -0.73205, 0.26795)^T$$