

第 5 章 矩阵特征值计算

与线性方程组的求解问题一样,矩阵特征值与特征向量的计算也是数值线性代数的重要内容。理论上,矩阵的特征值是特征多项式方程的根,因此特征值的计算可转换为单个多项式方程的求解。然而,对于高阶矩阵,这种转换并不能使问题得到简化,而且在实际应用中还会引入严重的数值误差。因此,正如第 2 章指出的,人们一般将多项式方程求解转换为矩阵特征值计算问题,而不是反过来。

本章介绍有关矩阵特征值计算问题的基本理论和算法。与非线性方程求根问题类似,计算矩阵特征值的算法也是迭代方法^①。

5.1 基本概念与特征值分布

本节首先介绍矩阵特征值、特征向量的基本概念和性质,然后讨论对特征值分布范围的简单估计方法。

5.1.1 基本概念与性质

定义 5.1: 矩阵 $\mathbf{A} = (a_{kj}) \in \mathbb{C}^{n \times n}$,

(1) 称

$$\varphi(\lambda) = \det(\lambda \mathbf{I} - \mathbf{A}) = \lambda^n + c_1 \lambda^{n-1} + \cdots + c_{n-1} \lambda + c_n$$

为 $\mathbf{A}$ 的特征多项式(characteristic polynomial); n 次代数方程

$$\varphi(\lambda) = 0$$

为 $\mathbf{A}$ 的特征方程(characteristic equation), 它的 n 个根 $\lambda_1, \lambda_2, \dots, \lambda_n$ 被称为 $\mathbf{A}$ 的特征值(eigenvalue)。此外, 常用 $\lambda(\mathbf{A})$ 表示 $\mathbf{A}$ 的全体特征值的集合, 也称为特征值谱(spectrum of eigenvalue)。

(2) 对于矩阵 $\mathbf{A}$ 的一个给定特征值 λ , 相应的齐次线性方程组

$$(\lambda \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0} \quad (5.1)$$

有非零解(因为系数矩阵奇异), 其解向量 $\mathbf{x}$ 称为矩阵 $\mathbf{A}$ 对应于 λ 的特征向量(eigenvector)。

根据式(5.1)得出矩阵特征值与特征向量的关系, 即

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad (5.2)$$

第 3 章的定义 3.5 就利用式(5.2)对矩阵特征值和特征向量进行了定义, 它与定义 5.1 等价。另外, 同一个特征值对应的特征向量一定不唯一, 它们构成线性子空间, 称为特征子空间(eigenspace)。

我们一般讨论实矩阵的特征值问题。注意, 实矩阵的特征值和特征向量不一定是实数

^① 如果用有限次运算能求得一般矩阵的特征值, 则多项式方程求根问题也可用有限次运算解决, 这与阿贝尔证明的“高于 4 次的多项式并不都有用初等运算表示的求根公式”的理论矛盾。

和实向量,但实特征值一定对应于实特征向量(式(5.1)的解),而一般的复特征值对应的特征向量一定不是实向量。此外,若特征值不是实数,则其复共轭也一定是特征值(由于特征方程为实系数方程)。定理 3.3 表明,实对称矩阵 $A \in \mathbb{R}^{n \times n}$ 的特征值均为实数,存在 n 个线性无关且正交的实特征向量,即存在由特征值组成的对角阵 Λ 和由特征向量组成的正交阵 Q ,使得

$$A = Q \Lambda Q^T \quad (5.3)$$

例 5.1(弹簧-质点系统): 考虑图 5-1 的弹簧-质点系统,其中包括 3 个质量分别为 m_1, m_2, m_3 的物体,由 3 个弹性系数分别为 k_1, k_2, k_3 的弹簧相连,3 个物体的位置均为时间的函数,这里考查 3 个物体偏离平衡位置的位移,分别记为 $y_1(t), y_2(t), y_3(t)$ 。因为物体在平衡状态所受的重力已经和弹簧伸长的弹力平衡,所以物体的加速度只和偏离平衡位置引起的弹簧伸长相关。根据牛顿第二定律以及胡克定律(即弹簧的弹力与拉伸长度成正比)可列出如下微分方程组^①:

$$M y''(t) + K y(t) = 0$$

其中 $y(t) = [y_1(t) \ y_2(t) \ y_3(t)]^T$,

$$M = \begin{bmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{bmatrix}, \quad K = \begin{bmatrix} k_1 + k_2 & -k_2 & 0 \\ -k_2 & k_2 + k_3 & -k_3 \\ 0 & -k_3 & k_3 \end{bmatrix}。$$

一般情况下,这个系统会以自然频率 ω 做谐波振动,而 y 的通解包含如下分量:

$$y_j(t) = x_j e^{i\omega t}, \quad (j = 1, 2, 3),$$

其中, $i = \sqrt{-1}$, 根据它可求解振动的频率 ω 及振幅 x_j 。由这个式子可得出

$$y_j''(t) = -\omega^2 x_j e^{i\omega t}, \quad (j = 1, 2, 3),$$

代入微分方程,可得代数方程

$$-\omega^2 M x + K x = 0,$$

或

$$A x = \lambda x,$$

其中, $A = M^{-1}K, \lambda = \omega^2$ 。通过求解矩阵 A 的特征值可求出这个弹簧-质点系统的自然频率(有多个)。再结合初始条件可确定这 3 个位移函数,它们可能按某个自然频率振动(简正振动),也可能是若干个简正振动的线性叠加。 ■

例 5.2(根据定义计算特征值、特征向量): 求矩阵

$$A = \begin{bmatrix} 5 & -1 & -1 \\ 3 & 1 & -1 \\ 4 & -2 & 1 \end{bmatrix}$$

的特征值和特征向量。

【解】 矩阵 A 的特征方程为

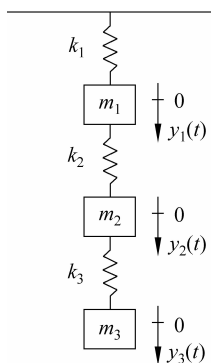


图 5-1 弹簧-质点系统

^① 本书第 8 章将介绍这种常微分方程组的数值求解方法。

$$\det(\lambda \mathbf{I} - \mathbf{A}) = \begin{vmatrix} \lambda - 5 & 1 & 1 \\ -3 & \lambda - 1 & 1 \\ -4 & 2 & \lambda - 1 \end{vmatrix} = (\lambda - 3)(\lambda - 2)^2 = 0,$$

故 $\mathbf{A}$ 的特征值为 $\lambda_1 = 3, \lambda_2 = 2$ (二重特征值)。

当 $\lambda = \lambda_1 = 3$ 时, 由 $(\lambda \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$, 得到方程

$$\begin{bmatrix} -2 & 1 & 1 \\ -3 & 2 & 1 \\ -4 & 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

它有无穷多个解, 若假设 $x_1 = 1$, 则求出解为 $\mathbf{x} = [1, 1, 1]^T$, 记为 $\mathbf{x}_1$, 则 $\mathbf{x}_1$ 是 λ_1 对应的一个特征向量。

当 $\lambda = \lambda_2 = 2$ 时, 由 $(\lambda \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$, 得到方程

$$\begin{bmatrix} -3 & 1 & 1 \\ -3 & 1 & 1 \\ -4 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

它有无穷多个解, 若假设 $x_1 = 1$, 则求出解为 $\mathbf{x} = [1, 1, 2]^T$, 记为 $\mathbf{x}_2$, 则 $\mathbf{x}_2$ 是 λ_2 对应的一个特征向量。 ■

下面概括地介绍有关矩阵特征值、特征向量的一些性质, 它们可根据定义 5.1 以及式(5.2)证明。

定理 5.1: 设 $\lambda_j (j=1, 2, \dots, n)$ 为 n 阶矩阵 $\mathbf{A}$ 的特征值, 则

$$(1) \sum_{j=1}^n \lambda_j = \sum_{j=1}^n a_{jj} = \text{tr}(\mathbf{A}).$$

$$(2) \prod_{j=1}^n \lambda_j = \det(\mathbf{A}).$$

这里, $\text{tr}(\mathbf{A})$ 表示矩阵对角线上元素之和, 称为矩阵的迹(trace)。

从上述结论(2)也可以看出, 非奇异矩阵特征值均不为 0, 而 0 一定是奇异矩阵的特征值。

定理 5.2: 矩阵转置不改变特征值, 即 $\lambda(\mathbf{A}) = \lambda(\mathbf{A}^T)$ 。

定理 5.3: 若矩阵 $\mathbf{A}$ 为对角阵或上(下)三角阵, 则其对角线元素即矩阵的特征值。

定理 5.4: 若矩阵 $\mathbf{A}$ 为分块对角阵, 或分块上(下)三角阵, 例如

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1m} \\ & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2m} \\ & & \ddots & \vdots \\ & & & \mathbf{A}_{mm} \end{bmatrix},$$

其中, 每个对角块 $\mathbf{A}_{jj}$ 均为方阵, 则矩阵 $\mathbf{A}$ 的特征值为各对角块矩阵特征值的合并, 即 $\lambda(\mathbf{A}) = \bigcup_{j=1}^m \lambda(\mathbf{A}_{jj})$ 。

定理 5.5: 矩阵的相似变换(similarity transformation)不改变特征值。设矩阵 $\mathbf{A}$ 和 $\mathbf{B}$ 为相似矩阵, 即存在非奇异矩阵 $\mathbf{X}$ 使得 $\mathbf{B} = \mathbf{X}^{-1}\mathbf{A}\mathbf{X}$, 则

(1) 矩阵 $\mathbf{A}$ 和 $\mathbf{B}$ 的特征值相等, 即 $\lambda(\mathbf{A}) = \lambda(\mathbf{B})$ 。

(2) 若 $\mathbf{y}$ 为 $\mathbf{B}$ 的特征向量, 则相应地, $\mathbf{Xy}$ 为 $\mathbf{A}$ 的特征向量。

通过相似变换并不总能把矩阵转换为对角阵, 或者说矩阵 $\mathbf{A}$ 并不总是可对角化的 (diagonalizable)。下面给出特征值的代数重数、几何重数和亏损矩阵的概念, 以及几个定理。

定义 5.2: 设矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 有 m 个 ($m \leq n$) 不同的特征值 $\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_m$, 若 $\tilde{\lambda}_j$ 是特征方程的 n_j 重根, 则称 n_j 为 $\tilde{\lambda}_j$ 的代数重数 (algebraic multiplicity), 并称 $\tilde{\lambda}_j$ 对应的特征子空间 ($\mathbb{C}^n$ 的子空间) 的维数为 $\tilde{\lambda}_j$ 的几何重数 (geometric multiplicity)。

定理 5.6: 设矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 的 m 个不同的特征值为 $\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_m$, 特征值 $\tilde{\lambda}_j$ ($j=1, 2, \dots, m$) 的代数重数为 n_j , 几何重数为 k_j , 则

(1) $\sum_{j=1}^m n_j = n$, 且任一个特征值的几何重数都不大于代数重数, 即 $\forall j, n_j \geq k_j$ 。

(2) 不同特征值的特征向量线性无关, 并且将所有特征子空间的 $\sum_{j=1}^m k_j$ 个基 (特征向量) 放在一起, 它们构成一组线性无关向量。

(3) 若每个特征值的代数重数等于几何重数, 则总共可得 n 个线性无关的特征向量, 它们是全空间 $\mathbb{C}^n$ 的基。

定义 5.3: 若矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 的某个代数重数为 k 的特征值对应的线性无关特征向量数目少于 k (即几何重数小于代数重数), 则称 $\mathbf{A}$ 为亏损矩阵 (defective matrix), 否则称其为非亏损矩阵 (nondefective matrix)。

定理 5.7: 矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 可对角化的充要条件是 $\mathbf{A}$ 为非亏损矩阵。若 $\mathbf{A}$ 可对角化, 即存在非奇异矩阵 $\mathbf{X} \in \mathbb{C}^{n \times n}$ 使得

$$\mathbf{X}^{-1}\mathbf{A}\mathbf{X} = \Lambda,$$

其中, $\Lambda \in \mathbb{C}^{n \times n}$ 为对角矩阵, 则 Λ 的对角线元素为矩阵 $\mathbf{A}$ 的特征值, 而矩阵 $\mathbf{X}$ 的列向量为 n 个线性无关的特征向量。

定理 5.7 中方程的等价形式为 $\mathbf{A} = \mathbf{X}\Lambda\mathbf{X}^{-1}$, 它被称为特征值分解, 也被称为谱分解 (spectrum decomposition)。特征值分解存在的充要条件是 $\mathbf{A}$ 为非亏损矩阵。但现实中还有很多矩阵是亏损矩阵, 如例 5.2 中的矩阵, 它的特征值 2 的代数重数为 2, 而几何重数仅为 1。这种矩阵不能相似变换为对角矩阵, 但存在下面的约当分解 (Jordan decomposition)。

定理 5.8: 设矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$, 存在非奇异矩阵 $\mathbf{X} \in \mathbb{C}^{n \times n}$ 使得

$$\mathbf{A} = \mathbf{X}\mathbf{J}\mathbf{X}^{-1},$$

矩阵 $\mathbf{J}$ 为形如 $\begin{bmatrix} \mathbf{J}_1 & & \\ & \ddots & \\ & & \mathbf{J}_p \end{bmatrix}$ 的分块对角矩阵 (称为约当标准型), 其中 $\mathbf{J}_s = \begin{bmatrix} \lambda_s & 1 & & \\ & \lambda_s & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_s \end{bmatrix}$,

$1 \leq s \leq p$, 称为约当块, 其对角线元素为矩阵 $\mathbf{A}$ 的特征值。设矩阵 $\mathbf{A}$ 有 m 个不同的特征值为 $\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_m$, 特征值 $\tilde{\lambda}_j$ ($j=1, 2, \dots, m$) 的代数重数为 n_j , 几何重数为 k_j , 则 $p = \sum_{j=1}^m k_j$, $\tilde{\lambda}_j$ 对应 k_j 个约当块 (对角元均为 $\tilde{\lambda}_j$), 其阶数之和等于 n_j 。

在约当分解中, 如果所有约当块都是 1 阶的, 则 $\mathbf{J}$ 为对角矩阵, 这种分解就是特征值分

解,相应的矩阵为非亏损矩阵。约当分解是很有用的理论工具,利用它还可证明下面关于矩阵运算结果的特征值的定理。

定理 5.9: 设 $\lambda_j (j=1,2,\cdots,n)$ 为 n 阶矩阵 $\mathbf{A}$ 的特征值,则

- (1) 矩阵 $c\mathbf{A}$ (c 为常数) 的特征值为 $c\lambda_1, c\lambda_2, \cdots, c\lambda_n$ 。
- (2) 矩阵 $\mathbf{A}+c\mathbf{I}$ (c 为常数) 的特征值为 $\lambda_1+c, \lambda_2+c, \cdots, \lambda_n+c$ 。
- (3) 矩阵 $\mathbf{A}^k$ (k 为正整数) 的特征值为 $\lambda_1^k, \lambda_2^k, \cdots, \lambda_n^k$ 。
- (4) 设 $p(t)$ 为一多项式函数, 则矩阵 $p(\mathbf{A})$ 的特征值为 $p(\lambda_1), p(\lambda_2), \cdots, p(\lambda_n)$ 。
- (5) 若 $\mathbf{A}$ 为非奇异矩阵, 则 $\lambda_j \neq 0 (j=1,2,\cdots,n)$, 且矩阵 $\mathbf{A}^{-1}$ 的特征值为 $\lambda_1^{-1}, \lambda_2^{-1}, \cdots, \lambda_n^{-1}$ 。

5.1.2 特征值分布范围的估计

估计特征值的分布范围或它们的界,无论在理论上或实际应用上,都有重要意义。例如,本书前面的内容曾涉及两个问题。

- (1) 计算矩阵的 2-条件数: $\text{cond}(\mathbf{A})_2 = \frac{\sqrt{\lambda_{\max}(\mathbf{A}^T \mathbf{A})}}{\sqrt{\lambda_{\min}(\mathbf{A}^T \mathbf{A})}}$ 。

- (2) 考查 1 阶定常迭代法 $\mathbf{x}^{(k+1)} = \mathbf{B}\mathbf{x}^{(k)} + \mathbf{f}$ 的收敛性、收敛速度,收敛的判据是谱半径 $\rho(\mathbf{B}) = \max_{1 \leq j \leq n} |\lambda_j(\mathbf{B})| < 1$, 收敛速度为 $R = -\lg \rho(\mathbf{B})$ 。

其中,都需要了解矩阵特征值的分布范围。

第 4 章的定理 4.4 说明谱半径的大小不超过任何一种算子范数,即

$$\rho(\mathbf{A}) \leq \|\mathbf{A}\|,$$

这是关于特征值的上界的一个重要结论。

下面先给出定义 5.4,再介绍有关特征值的界的另一个重要结论。

定义 5.4: 设 $\mathbf{A}=(a_{kj}) \in \mathbb{C}^{n \times n}$, 记 $r_k = \sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}| (k=1,2,\cdots,n)$, 则集合 $D_k = \{z: |z -$

$a_{kk}| \leq r_k, z \in \mathbb{C}\} (k=1,2,\cdots,n)$ 在复平面为以 a_{kk} 为圆心、 r_k 为半径的圆盘,称为 $\mathbf{A}$ 的 Gerschgorin(格什戈林)圆盘。

图 5-2 显示了一个 3×3 复矩阵的格什戈林圆盘。

定理 5.10(圆盘定理): 设 $\mathbf{A}=(a_{kj}) \in \mathbb{C}^{n \times n}$, 则

- (1) $\mathbf{A}$ 的每个特征值必属于 $\mathbf{A}$ 的格什戈林圆盘中, 即对任一特征值 λ , 必定存在 $k (1 \leq k \leq n)$, 使得

$$|\lambda - a_{kk}| \leq \sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}| \quad (5.4)$$

用集合的关系说明, 这意味着 $\lambda(\mathbf{A}) \subseteq \bigcup_{k=1}^n D_k$ 。

- (2) 若 $\mathbf{A}$ 的格什戈林圆盘中有 m 个圆盘组成一连通并集 S , 且 S 与余下的 $n-m$ 个圆盘分离, 则 S 内恰好包含 $\mathbf{A}$ 的 m 个特征值(重特征值按重数计)。

对如图 5-2 所示的例子,定理 5.10 的第(2)个结论的含义是: D_1 中只包含一个特征

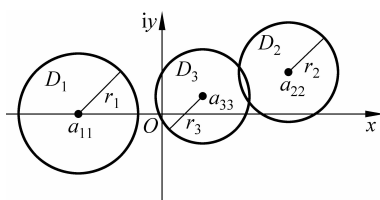


图 5-2 复坐标平面,以及 3×3 矩阵 $\mathbf{A}$ 的格什戈林圆盘

值,而另外两个特征值在 D_2, D_3 的并集中。下面对定理 5.10 的结论(1)进行证明,结论(2)的证明超出了本书的范围。

【证明】 设 λ 为 $\mathbf{A}$ 的任一特征值,则有 $\mathbf{A}\mathbf{x}=\lambda\mathbf{x}$, $\mathbf{x}$ 为非零向量。设 $\mathbf{x}$ 中第 k 个分量最大,即

$$|x_k| = \max_{1 \leq j \leq n} |x_j| > 0,$$

考虑式(5.2)中第 k 个方程

$$\sum_{j=1}^n a_{kj}x_j = \lambda x_k,$$

将其中与 x_k 有关的项移到等号左边,其余项移到右边,再两边取模得

$$|\lambda - a_{kk}| |x_k| = \left| \sum_{\substack{j=1 \\ j \neq k}}^n a_{kj}x_j \right| \leq \sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}| |x_j| \leq |x_k| \sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}| \quad (5.5)$$

最后一个不等式的推导利用了“ $\mathbf{x}$ 中第 k 个分量最大”的假设。将不等式(5.5)除以 $|x_k|$,得到式(5.4),因此证明了定理 5.10 的结论(1)。上述证明过程还说明,若某个特征向量的第 k 个分量的模最大,则相应的特征值必属于第 k 个圆盘。■

根据定理 5.2,还可以按照矩阵的每列元素定义 n 个圆盘,对于它们定理 5.10 仍然成立。下面的定理是圆盘定理的重要推论,其证明留给感兴趣的读者。

定理 5.11: 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$, 且 $\mathbf{A}$ 的对角元均大于 0, 则

(1) 若 $\mathbf{A}$ 严格对角占优, 则 $\mathbf{A}$ 的特征值的实部都大于 0。

(2) 若 $\mathbf{A}$ 为对角占优的对称矩阵, 则 $\mathbf{A}$ 一定是对称半正定矩阵。若同时 $\mathbf{A}$ 非奇异, 则 $\mathbf{A}$ 为对称正定矩阵。

例 5.3(圆盘定理的应用): 试估计矩阵

$$\mathbf{A} = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & -4 \end{bmatrix}$$

的特征值范围。

【解】 直接应用圆盘定理,该矩阵的 3 个圆盘如下:

$$D_1: |\lambda - 4| \leq 1, \quad D_2: |\lambda| \leq 2, \quad D_3: |\lambda + 4| \leq 2.$$

D_1 与其他圆盘分离,则它仅含一个特征值,且必定为实数(若为虚数,则其共轭也是特征值,这与 D_1 仅含一个特征值矛盾)。所以,对矩阵特征值的范围的估计是

$$3 \leq \lambda_1 \leq 5, \quad \lambda_2, \lambda_3 \in D_2 \cup D_3.$$

再对矩阵 $\mathbf{A}^T$ 应用圆盘定理,则可以进一步优化上述结果。矩阵 $\mathbf{A}^T$ 对应的 3 个圆盘为

$$D'_1: |\lambda - 4| \leq 2, \quad D'_2: |\lambda| \leq 2, \quad D'_3: |\lambda + 4| \leq 1.$$

这说明 D'_3 中存在一个特征值,且为实数,它属于区间 $[-5, -3]$, 经过综合分析可知 3 个特征值均为实数,它们的范围是

$$\lambda_1 \in [3, 5], \quad \lambda_2 \in [-2, 2], \quad \lambda_3 \in [-5, -3].$$

事实上,使用 MATLAB 的 eig 命令可求出矩阵 $\mathbf{A}$ 的特征值为 4.2030, -0.4429, -3.7601。■

根据定理 5.5,还可以对矩阵 $\mathbf{A}$ 做简单的相似变换,如取 $\mathbf{X}$ 为对角阵,然后再应用圆盘定理估计特征值的范围。

例 5.4(特征值范围的估计): 选取适当的矩阵 $\mathbf{X}$, 应用定理 5.5 和定理 5.10 估计例 5.3 中矩阵的特征值范围。

【解】 取

$$\mathbf{X}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.9 \end{bmatrix}$$

则

$$\mathbf{A}_1 = \mathbf{X}^{-1} \mathbf{A} \mathbf{X} = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 0 & -10/9 \\ 0.9 & 0.9 & -4 \end{bmatrix}$$

的特征值与 $\mathbf{A}$ 的相同。对 $\mathbf{A}_1$ 应用圆盘定理, 得到 3 个分离的圆盘, 它们分别包含一个实特征值, 由此得到特征值的范围估计

$$\lambda_1 \in [3, 5], \quad \lambda_2 \in \left[-\frac{19}{9}, \frac{19}{9}\right], \quad \lambda_3 \in [-5.8, -2.2]$$

此外, 还可进一步估计 $\rho(\mathbf{A})$ 的范围, 即 $3 \leq \rho(\mathbf{A}) \leq 5.8$ 。 ■

例 5.4 表明, 综合运用圆盘定理和矩阵特征值的性质(如定理 5.2、定理 5.5), 可对特征值的范围进行一定的估计。对具体例子, 可适当设置相似变换矩阵, 尽可能让圆盘相互分离, 从而提高估计的有效性。

5.2 幂法与反幂法

幂法是一种计算矩阵最大的特征值及其对应特征向量的方法。本节介绍幂法、反幂法以及加快幂法迭代收敛的技术。

5.2.1 幂法

定义 5.5: 在矩阵 $\mathbf{A}$ 的特征值中, 模最大的特征值称为主特征值, 也称为“第一特征值”, 它对应的特征向量称为主特征向量。

应注意的是, 主特征值有可能不唯一, 因为模相同的复数可以有很多。例如, 模为 5 的特征值可能是 5, -5 , $3+4i$, $3-4i$, 等等。另外, 请注意谱半径和主特征值的区别。

如果矩阵 $\mathbf{A}$ 有唯一的主特征值, 则通过幂法能方便地计算出主特征值及其对应的特征向量。对于实矩阵, 这个唯一的主特征值显然是实数, 但不排除它是重特征值的情况。幂法(power iteration)的计算过程是: 首先随机选一非零向量 $\mathbf{v}_0 \in \mathbb{R}^n$, 然后进行迭代计算

$$\mathbf{v}_k = \mathbf{A} \mathbf{v}_{k-1}, \quad (k = 1, 2, \dots)$$

得到向量序列 $\{\mathbf{v}_k\}$, 根据它即可求出主特征与特征向量。下面用定理说明。

定理 5.12: 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$, 其主特征值唯一, 记为 λ_1 , 随机选一非零向量 $\mathbf{v}_0 \in \mathbb{R}^n$ 按迭代公式进行计算:

$$\mathbf{v}_k = \mathbf{A} \mathbf{v}_{k-1}, \quad (k = 1, 2, \dots),$$

则有

(1) 当 $k \rightarrow \infty$ 时, $\mathbf{v}_k$ 趋近于 λ_1 的特征向量。

$$(2) \quad \lim_{k \rightarrow \infty} \frac{(\mathbf{v}_{k+1})_j}{(\mathbf{v}_k)_j} = \lambda_1, \quad (5.6)$$

其中 $(\mathbf{v}_k)_j$ 表示向量 $\mathbf{v}_k$ 的第 j 个分量, 且 j 为 $\mathbf{v}_k$ 、 $\mathbf{v}_{k+1}$ 绝对值最大元素的下标。

【证明】 这里只考虑矩阵 $\mathbf{A}$ 为非亏损矩阵的情况。设 $\mathbf{A}$ 的主特征值为 $p-1$ 重特征值, 全部 n 个特征值按模从大到小排列为: $|\lambda_1| = \cdots = |\lambda_{p-1}| > |\lambda_p| \geq \cdots \geq |\lambda_n|$, 它们对应于一组线性无关的单位特征向量 $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \cdots, \hat{\mathbf{x}}_n$ 。向量 $\mathbf{v}_0$ 可写成这些特征向量的线性组合:

$$\mathbf{v}_0 = \alpha_1 \hat{\mathbf{x}}_1 + \cdots + \alpha_n \hat{\mathbf{x}}_n,$$

一般地, $\alpha_1 \neq 0$, 则

$$\begin{aligned} \mathbf{v}_k &= \mathbf{A} \mathbf{v}_{k-1} = \mathbf{A}^k \mathbf{v}_0 = \alpha_1 \lambda_1^k \hat{\mathbf{x}}_1 + \alpha_2 \lambda_2^k \hat{\mathbf{x}}_2 + \cdots + \alpha_n \lambda_n^k \hat{\mathbf{x}}_n \\ &= \lambda_1^k \left[\sum_{j=1}^{p-1} \alpha_j \hat{\mathbf{x}}_j + \sum_{j=p}^n \alpha_j \left(\frac{\lambda_j}{\lambda_1} \right)^k \hat{\mathbf{x}}_j \right] \\ &= \lambda_1^k (\sum_{j=1}^{p-1} \alpha_j \hat{\mathbf{x}}_j + \boldsymbol{\varepsilon}_k), \end{aligned} \quad (5.7)$$

其中 $\boldsymbol{\varepsilon}_k = \sum_{j=p}^n \alpha_j \left(\frac{\lambda_j}{\lambda_1} \right)^k \hat{\mathbf{x}}_j$ 。由于 $\left| \frac{\lambda_j}{\lambda_1} \right| < 1, (j=p, p+1, \cdots, n)$, 则

$$\lim_{k \rightarrow \infty} \boldsymbol{\varepsilon}_k = \mathbf{0} \Rightarrow \lim_{k \rightarrow \infty} \frac{\mathbf{v}_k}{\lambda_1^k} = \sum_{j=1}^{p-1} \alpha_j \hat{\mathbf{x}}_j。$$

式(5.7)中, $\sum_{j=1}^{p-1} \alpha_j \hat{\mathbf{x}}_j$ 是主特征值 λ_1 的一个特征向量。由于特征向量放大、缩小任意倍数后仍是特征向量, 则随 k 的增大, $\mathbf{v}_k$ 越来越趋近于 λ_1 的特征向量。

更进一步, 特征向量为非零向量, 其绝对值最大元素一定不为 0, 所以当 k 足够大时, $(\mathbf{v}_k)_j \neq 0, (\mathbf{v}_{k+1})_j \neq 0$ 。因此, 根据式(5.7), 有

$$\frac{(\mathbf{v}_{k+1})_j}{(\mathbf{v}_k)_j} = \lambda_1 \frac{(\sum_{j=1}^{p-1} \alpha_j \hat{\mathbf{x}}_j + \boldsymbol{\varepsilon}_{k+1})_j}{(\sum_{j=1}^{p-1} \alpha_j \hat{\mathbf{x}}_j + \boldsymbol{\varepsilon}_k)_j}$$

由于 $\lim_{k \rightarrow \infty} \boldsymbol{\varepsilon}_k = \mathbf{0}$, 随 k 的增大, 上式等号右边趋于一个常数 λ_1 , 这就证明了定理的结论。■

若矩阵 $\mathbf{A}$ 为亏损矩阵, 则可利用矩阵的约当分解证明式(5.6), 这里略去。在这种情况下, 式(5.6)的收敛速度可能很慢。

关于定理 5.12, 再说明下列两点。

(1) 理论上讲, 定理 5.12 成立的前提条件应包括 $\alpha_1 \neq 0$ 。特别地, 若初始向量 $\mathbf{v}_0$ 为某个非主特征值对应的特征向量, 则式(5.6)求出的就是那个特征值。然而, 在实践中, 若随机选取 $\mathbf{v}_0$, 且由于计算的舍入误差, 就不会出现此情况。

(2) 式(5.6)的含义是相邻迭代向量分量的比值收敛到主特征值。其中 j 在实际计算时取 $\mathbf{v}_k$ 中最大的分量对应的 j 。

直接使用幂法, 还存在如下两方面问题。

(1) 溢出: 由于 $\mathbf{v}_k \approx \lambda_1^k \mathbf{x}_1$, 则

$|\lambda_1| > 1$ 时, 实际计算 $\mathbf{v}_k$ 会出现上溢出(当 k 很大时)。

$|\lambda_1| < 1$ 时, 实际计算 $\mathbf{v}_k$ 会出现下溢出(当 k 很大时)。

(2) 可能收敛速度很慢。由于 $\epsilon_k = \sum_{j=p}^n \alpha_j \left(\frac{\lambda_j}{\lambda_1}\right)^k \mathbf{x}_j$, 因此 $\epsilon_k \rightarrow \mathbf{0}$ 的速度取决于求和式中衰减最慢的因子 $\left|\frac{\lambda_p}{\lambda_1}\right|$, 也就是说, 当绝对值第 2 大的特征值其绝对值接近主特征值的绝对值时, 幂法的收敛速度就会很慢。

下面采用规格化向量的技术防止溢出, 导出实用的幂法。关于加速收敛技术的讨论, 见 5.2.2 节。

定义 5.6: 记 $\overline{\max}(\mathbf{v})$ 为向量 $\mathbf{v} \in \mathbb{R}^n$ 的绝对值最大的分量, 即 $\overline{\max}(\mathbf{v}) = v_j$, 其中 j 满足 $|v_j| = \max_{1 \leq k \leq n} |v_k|$, 若 j 的值不唯一, 则取最小的那个, 并且称 $\mathbf{u} = \mathbf{v} / \overline{\max}(\mathbf{v})$ 为向量 $\mathbf{v}$ 的规格化向量(normalized vector)。

例 5.5(规格化向量): 设 $\mathbf{v} = [3, -5, 0]^T$, $\overline{\max}(\mathbf{v}) = -5$, 对应的规格化向量为 $\mathbf{u} = \left[-\frac{3}{5}, 1, 0\right]^T$ 。

根据定义 5.6, 容易得出规格化向量的两条性质。

定理 5.13: 定义 5.6 中的规格化向量满足如下两条性质。

(1) 若 $\mathbf{u}$ 为规格化向量, 则 $\|\mathbf{u}\|_\infty = 1$, 并且 $\overline{\max}(\mathbf{u}) = 1$ 。

(2) 设向量 $\mathbf{v}_1$ 和 $\mathbf{v}_2$ 的规格化向量分别为 $\mathbf{u}_1$ 和 $\mathbf{u}_2$, 若 $\mathbf{v}_1 = \alpha \mathbf{v}_2$, 实数 $\alpha \neq 0$, 则 $\mathbf{u}_1 = \mathbf{u}_2$ 。

在幂法的每一步增加向量规格化的操作可解决溢出问题。先看第一步, $\mathbf{v}_1 = \mathbf{A} \mathbf{v}_0$, 此时计算 $\mathbf{v}_1$ 的规格化向量

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{\overline{\max}(\mathbf{v}_1)} = \frac{\mathbf{A} \mathbf{v}_0}{\overline{\max}(\mathbf{A} \mathbf{v}_0)}$$

然后使用规格化向量计算 $\mathbf{v}_2$

$$\mathbf{v}_2 = \mathbf{A} \mathbf{u}_1 = \frac{\mathbf{A}^2 \mathbf{v}_0}{\overline{\max}(\mathbf{A} \mathbf{v}_0)} \quad (5.8)$$

再进行向量规格化操作,

$$\mathbf{u}_2 = \frac{\mathbf{v}_2}{\overline{\max}(\mathbf{v}_2)} = \frac{\mathbf{A}^2 \mathbf{v}_0}{\overline{\max}(\mathbf{A}^2 \mathbf{v}_0)} \quad (5.9)$$

式(5.9)的推导利用了式(5.8)和定理 5.13 的结论(2)。依此类推, 得到

$$\begin{cases} \mathbf{v}_k = \mathbf{A} \mathbf{u}_{k-1} = \frac{\mathbf{A}^k \mathbf{v}_0}{\overline{\max}(\mathbf{A}^{k-1} \mathbf{v}_0)} \\ \mathbf{u}_k = \frac{\mathbf{v}_k}{\overline{\max}(\mathbf{v}_k)} = \frac{\mathbf{A}^k \mathbf{v}_0}{\overline{\max}(\mathbf{A}^k \mathbf{v}_0)} \end{cases}, \quad (k = 1, 2, \dots) \quad (5.10)$$

根据定理 5.12 的证明过程(为了表述简单, 这里不妨设主特征值 λ_1 不是重特征值),

$$\begin{aligned} \mathbf{A}^k \mathbf{v}_0 &= \lambda_1^k \left[\alpha_1 \hat{\mathbf{x}}_1 + \sum_{j=2}^n \alpha_j \left(\frac{\lambda_j}{\lambda_1}\right)^k \hat{\mathbf{x}}_j \right] \\ \Rightarrow \mathbf{u}_k &= \frac{\mathbf{A}^k \mathbf{v}_0}{\overline{\max}(\mathbf{A}^k \mathbf{v}_0)} = \frac{\alpha_1 \hat{\mathbf{x}}_1 + \sum_{j=2}^n \alpha_j \left(\frac{\lambda_j}{\lambda_1}\right)^k \hat{\mathbf{x}}_j}{\overline{\max} \left[\alpha_1 \hat{\mathbf{x}}_1 + \sum_{j=2}^n \alpha_j \left(\frac{\lambda_j}{\lambda_1}\right)^k \hat{\mathbf{x}}_j \right]} \xrightarrow{k \rightarrow \infty} \frac{\mathbf{x}_1}{\overline{\max}(\mathbf{x}_1)}, \end{aligned}$$

即 $\mathbf{u}_k$ 逐渐逼近规格化的主特征向量。同理,

因此,根据定理 5.13 的结论(1),有

基于上述推导,得到如下定理,以及如算法 5.1 描述的实用幂法。应注意的是,由于 $\mathbf{x}_1 \neq \mathbf{0}$, 则 $\overline{\max(\mathbf{x}_1)} \neq 0$ 。

其中, $\mathbf{x}_1$ 为主特征向量。

$$x_1 := u \quad \{\text{规格化的主特征向量}\}$$

• 157 •