计算方法

令 丹

数学与统计学院

邮箱：danling@xjtu.edu.cn

第八章
矩阵特征值的数值计算

主要内容

1. 基本性质和概念

特征值和特征向量

定义 1 A 为 $n \times n$ 阶矩阵, x 为 n 维非零向量. 若存在数 λ 使得 $Ax = \lambda x$ 成立, 则称 λ 为 A 的特征值, x 为对应于 λ 的特征向量.

特征值和特征向量

定义 1 A 为 $n \times n$ 阶矩阵, x 为 n 维非零向量. 若存在数 λ 使得 $Ax = \lambda x$ 成立, 则称 λ 为 A 的特征值, x 为对应于 λ 的特征向量.

定义 2 若 λ 满足 $p_A(\lambda) = \det(\lambda I - A) = 0$, 则称 $p_A(\lambda)$ 为矩阵 A 的特征多项式, 称 $\det(\lambda I - A) = 0$ 为特征方程.

特征值和特征向量

定义 1 A 为 $n\times n$ 阶矩阵, x 为 n 维非零向量. 若存在数 λ 使得 $Ax=\lambda x$ 成立, 则称 λ 为 A 的特征值, x 为对应于 λ 的特征向量.

定义 2 若 λ 满足 $p_A(\lambda)=\det(\lambda I-A)=0$, 则称 $p_A(\lambda)$ 为矩阵 A 的特征多项式, 称 $\det(\lambda I-A)=0$ 为特征方程.

定义 3 若 λ_i 为特征方程的 k 重根, 称 k 为特征值 λ_i 的代数重数. 齐次线性方程组 $(\lambda_i I-A)x=0$ 的基础解系所含的向量个数 (即 λ_i 的特征子空间的维数) 称为特征值 λ_i 的几何重数.

特征值和特征向量

性质 1 设 n 阶矩阵 A 的全部特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 则有

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = \mathrm{tr}(A), \quad \lambda_1 \lambda_2 \cdots \lambda_n = \det(A).$$

特征值和特征向量

性质 1 设 n 阶矩阵 A 的全部特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 则有

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = \mathrm{tr}(A), \quad \lambda_1 \lambda_2 \cdots \lambda_n = \det(A).$$

性质 2 设 λ 为矩阵 A 的一个特征值, 则对任何多项式

$$p(x) = a_0 + a_1 x + \cdots + a_m x^m,$$

$p(\lambda)$ 为矩阵 $p(A)$ 的一个特征值.

特征值和特征向量

性质 1 设 n 阶矩阵 A 的全部特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 则有

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = \mathrm{tr}(A), \quad \lambda_1 \lambda_2 \cdots \lambda_n = \det(A).$$

性质 2 设 λ 为矩阵 A 的一个特征值, 则对任何多项式

$$p(x) = a_0 + a_1 x + \cdots + a_m x^m,$$

$p(\lambda)$ 为矩阵 $p(A)$ 的一个特征值.

性质 3 设 λ 为 n 阶可逆矩阵 A 的一个特征值, 则 $\lambda \neq 0$ 且 $\frac{1}{\lambda}$ 为 A 的逆矩阵 A^{-1} 的一个特征值, $\frac{\det(A)}{\lambda}$ 为 A 的伴随矩阵 A^* 的一个特征值.

性质 4　设 $\lambda_1, \lambda_2, \cdots, \lambda_m$ 为矩阵 A 的互异的特征值, x_i 为 A 的对应特征值 λ_i $(i = 1, 2, \cdots, m)$ 的特征向量, 则 $x_1, x_2, \cdots, x_m$ 线性无关.

性质 4　设 $\lambda_1, \lambda_2, \cdots, \lambda_m$ 为矩阵 A 的互异的特征值, x_i 为 A 的对应特征值 λ_i $(i = 1, 2, \cdots, m)$ 的特征向量, 则 $x_1, x_2, \cdots, x_m$ 线性无关.

性质 5　矩阵 A 的任意特征值的几何重数不大于其代数重数.

2. 乘幂法和反幂法

乘幂法

乘幂法：用于求模 (绝对值) 最大的特征值.

设 A 的特征值按模大小排列为

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

并且对应的特征向量 $\xi_1, \xi_2, \cdots, \xi_n$ 线性无关. 此时任一非零向量 z_0 可用 $\xi_1, \xi_2, \cdots, \xi_n$ 线性表示

$$z_0 = c_1\xi_1 + c_2\xi_2 + \cdots + c_n\xi_n$$

且 $c_1, c_2, \cdots, c_n$ 不全为0.

乘幂法

作向量序列 $z_k = A^k z_0$, 则

$$\begin{aligned} z_k = A^k z_0 &= c_1 A^k \xi_1 + c_2 A^k \xi_2 + \cdots + c_n A^k \xi_n \\ &= c_1 \lambda_1^k \xi_1 + c_2 \lambda_2^k \xi_2 + \cdots + c_n \lambda_n^k \xi_n \\ &= \lambda_1^k \Big(c_1 \xi_1 + c_2 \big(\frac{\lambda_2}{\lambda_1}\big)^k \xi_2 + \cdots + c_n \big(\frac{\lambda_n}{\lambda_1}\big)^k \xi_n \Big) \end{aligned}$$

由此可见, 若 $\lambda_1 \neq 0$, 则当 $k \to \infty$ 时,

$$\big(\frac{\lambda_j}{\lambda_1}\big)^k \to 0, \quad j = 2, 3, \cdots, n.$$

所以当 k 充分大时, $z_k \approx c_1 \lambda_1^k \xi_1$. z_k 可近似看作 λ_1 对应的特征向量. z_k 与 z_{k-1} 的非零分量之比约为 λ_1.

乘幂法

问题：

① $|\lambda_1| > 1$ 时, $\lambda_1^k \to \infty,\ k \to \infty.$ “上溢”

② $|\lambda_1| < 1$ 时, $\lambda_1^k \to 0,\ k \to \infty.$ “下溢”

$\Longrightarrow$ z_k 单位化

据此得到求模最大的特征值 λ_1 和对应特征向量 ξ_1 的算法.

(1) 任取非零向量 z_0, 给定 $\varepsilon > 0, \delta > 0$. 令 $k = 1$.

(2) (i) $y_k := Az_{k-1}$

(ii) 求 y_k 绝对值最大的分量 $m_k := \max(y_k)$.

(iii) $z_k := y_k/m_k$.

(iv) 若 $|m_k - m_{k-1}| < \delta$ 或 $\|z_k - z_{k-1}\| < \varepsilon$, 则

取 $\lambda_1 \approx m_k, \xi_1 \approx z_k$, 停止计算. 否则, $k := k+1$, 转 (i).

乘幂法

例 1：设

$$A=\begin{pmatrix}-1 & 2 & 1\\ 2 & -4 & 1\\ 1 & 1 & -6\end{pmatrix}$$

求 A 按模最大的特征值 λ_1 和对应的特征向量 ξ_1, 误差限为 $\|z_{k+1}-z_k\|\leqslant 10^{-8}$.

乘幂法

解　取 $z_0 = (1,1,1)^{\mathrm{T}}$, 用乘幂法计算结果如下

k	m_k	$(z_k)_1$	$(z_k)_2$	$(z_k)_3$
1	-4.00000000	-0.50000000	0.25000000	1.00000000
2	-6.25000000	-0.32000000	0.16000000	1.00000000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
61	-6.42106658	-0.04614550	-0.37492108	1.00000000
62	-6.42106659	-0.04614550	-0.37492109	1.00000000
63	-6.42106659	-0.04614549	-0.37492110	1.00000000
64	-6.42106660	-0.04614549	-0.37492110	1.00000000

乘幂法的收敛性

定理

乘幂法定义的序列 $\{m_k\}$ 和向量序列 $\{z_k\}$ 满足

$$\lim_{k\to\infty} m_k = \lambda_1, \qquad \lim_{k\to\infty} z_k = \frac{\xi_1}{\max(\xi_1)}.$$

乘幂法的收敛性

定理

乘幂法定义的序列 $\{m_k\}$ 和向量序列 $\{z_k\}$ 满足

$$\lim_{k\to\infty} m_k = \lambda_1, \qquad \lim_{k\to\infty} z_k = \frac{\xi_1}{\max(\xi_1)}.$$

证 由于

$$z_k = \frac{y_k}{m_k} = \frac{Az_{k-1}}{m_k} = \frac{A^2 z_{k-2}}{m_k m_{k-1}} = \cdots = \frac{A^k z_0}{m_k \cdot m_{k-1} \cdots m_1},$$

即 z_k 与 $A^k z_0$ 的方向相同或相反，而 z_k 的绝对值最大分量为 1.

乘幂法的收敛性

记 $\varepsilon_k = \sum_{j=2}^{n} c_j \left(\frac{\lambda_j}{\lambda_1}\right)^k \xi_j$, 则

$$z_k = \frac{A^k z_0}{\max(A^k z_0)} = \frac{\lambda_1^k \left(c_1\xi_1 + \varepsilon_k\right)}{\max\left[\lambda_1^k (c_1\xi_1 + \varepsilon_k)\right]} = \frac{c_1\xi_1 + \varepsilon_k}{\max(c_1\xi_1 + \varepsilon_k)},$$

$$m_k = \frac{\max(A^k z_0)}{\max(A^{k-1} z_0)} = \frac{\max(c_1\xi_1 + \varepsilon_k)}{\max(c_1\xi_1 + \varepsilon_{k-1})} = \lambda_1 \left(1 + O(\left|\frac{\lambda_2}{\lambda_1}\right|^k)\right).$$

由于 $|\lambda_1| > |\lambda_2|$, 于是

$$\lim_{k\to\infty} m_k = \lambda_1, \qquad \lim_{k\to\infty} z_k = \frac{c_1\xi_1}{\max(c_1\xi_1)} = \frac{\xi_1}{\max(\xi_1)}.$$

乘幂法的收敛性

另一方面

$$\lim_{k\to\infty}\frac{|m_{k+1}-\lambda_1|}{|m_k-\lambda_1|}=\left|\frac{\lambda_2}{\lambda_1}\right|\neq 0$$

因此线性收敛. $\left|\frac{\lambda_2}{\lambda_1}\right|$ 称为收敛率. 收敛率越小, 收敛越快.

当 $|\lambda_1|$ 与 $|\lambda_2|$ 接近时, 收敛慢. 可用 Aitken 方法进行加速

$$m_k^*=m_k-\frac{(m_k-m_{k-1})^2}{m_k-2m_{k-1}+m_{k-2}}.$$

乘幂法

例 2：用 Aitken 方法对例 1 的算法进行加速.

解 取 $z_0 = (1,1,1)^{\mathrm{T}}$, 用加速的乘幂法计算结果如下

k	m_k	$(z_k)_1$	$(z_k)_2$	$(z_k)_3$
1	-4.00000000	-0.50000000	0.25000000	1.00000000
2	-6.16346153	-0.22019166	-0.04635613	1.00929959
3	-6.31175629	-0.10766956	-0.25996439	1.00583497
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
15	-6.42106642	-0.04614558	-0.37492093	1.00000000
16	-6.42106655	-0.04614551	-0.37492106	1.00000000
17	-6.42106659	-0.04614549	-0.37492110	1.00000000

原点位移法

若 A 的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 则对任意常数 p, 矩阵 $B = A - pI$ 的特征值为 $\lambda_i - p \ (i = 1, 2, \cdots, n)$, 且 A, B 有相同的特征向量.

乘幂法的加速技术

原点位移法

若 A 的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 则对任意常数 p, 矩阵 $B = A - pI$ 的特征值为 $\lambda_i - p\ (i = 1, 2, \cdots, n)$, 且 A, B 有相同的特征向量.

若能选择合适的 p, 使得

$$|\lambda_1 - p| > |\lambda_2 - p| \geqslant \cdots \geqslant |\lambda_n - p| \quad \text{且} \quad \left|\frac{\lambda_2 - p}{\lambda_1 - p}\right| < \left|\frac{\lambda_2}{\lambda_1}\right|.$$

再对 B 用乘幂法求出 $\lambda_1 - p$ 的近似值 m_k, 则 $\lambda_1 \approx m_k + p$.

原点位移法

若 A 的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 则对任意常数 p, 矩阵 $B = A - pI$ 的特征值为 $\lambda_i - p\ (i = 1, 2, \cdots, n)$, 且 A, B 有相同的特征向量.

若能选择合适的 p, 使得

$$|\lambda_1 - p| > |\lambda_2 - p| \geqslant \cdots \geqslant |\lambda_n - p| \quad \text{且} \quad \left|\frac{\lambda_2 - p}{\lambda_1 - p}\right| < \left|\frac{\lambda_2}{\lambda_1}\right|.$$

再对 B 用乘幂法求出 $\lambda_1 - p$ 的近似值 m_k, 则 $\lambda_1 \approx m_k + p$.

p 的选取有赖于对 A 的特征值分布有大致的了解, 因此选取合适的 p 比较困难, 实际计算中并不常用.

例 3：取 $p=-2, z_0=(1,1,1)^{\mathrm{T}}$, 用原点位移法计算例 1.

解　计算结果如下

k	m_k+p	$(z_k)_1$	$(z_k)_2$	$(z_k)_3$
1	2.00000000	1.00000000	0.25000000	-0.50000000
2	1.25000000	0.30769230	0.30769230	1.000000000
3	-5.38461538	-0.56818181	-0.29545454	1.00000000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
40	-6.42106658	-0.04614550	-0.37492109	1.00000000
41	-6.42106659	-0.04614549	-0.37492110	1.00000000
42	-6.42106660	-0.04614549	-0.37492111	1.00000000

反幂法

用来求 A 的按模最小的特征值和对应的特征向量

设 A 可逆, 则对 A^{-1} 作乘幂法, 称为反幂法.

若 A 的特征值满足

$$|\lambda_1| \geqslant |\lambda_2| \geqslant \cdots \geqslant |\lambda_{n-1}| > |\lambda_n|.$$

根据反幂法有

$$m_k \to \frac{1}{\lambda_n}, \quad z_k \to \frac{\xi_n}{\max(\xi_n)},$$

且收敛率为 $\left|\frac{\lambda_n}{\lambda_{n-1}}\right|$.

$$y_k = A^{-1} z_{k-1} \quad \Longrightarrow \quad z_{k-1} = A y_k$$

y_k 的计算需要解线性方程组.

反幂法

若已知 A 的某个特征值 λ_m 的近似值 $\widetilde{\lambda}_m$, 则可用原点平移的反幂法来求 λ_m 对应的特征向量 ξ_m, 并在计算过程中改进 $\widetilde{\lambda}_m$ 的精度.

由于 $\widetilde{\lambda}_m$ 接近 λ_m, 一般应有

$$0 < |\lambda_m - \widetilde{\lambda}_m| \ll |\lambda_j - \widetilde{\lambda}_m|, \quad j \neq m.$$

则 $\left(\lambda_j - \widetilde{\lambda}_m\right)^{-1}$ 是 $(A - \widetilde{\lambda}_m I)^{-1}$ 的按模最大的特征值, ξ_m 是对应的特征向量.

反幂法

对 $A-\widetilde{\lambda}_m I$ 应用反幂法, 则当 $k\to\infty$ 时,

$$z_k \to \frac{\xi_m}{\max(\xi_m)}, \quad m_k \to \frac{1}{\lambda_m-\widetilde{\lambda}_m}$$

$$\Longrightarrow \quad \widetilde{\lambda}_m + \frac{1}{m_k} \to \lambda_m, \quad k\to\infty.$$

收敛率 $\left|\frac{\lambda_m-\widetilde{\lambda}_m}{\lambda_j-\widetilde{\lambda}_m}\right| \ll 1$, 迭代很快, 一般只经过一两次迭代就可达到较高的精度.

反幂法

(1) 给定 $\delta > 0, \varepsilon > 0$.

(2) 对 $A - \widetilde{\lambda}_m$ 作三角分解 $A - \widetilde{\lambda}_m = LR$. 令 $k = 1$.

(3) 解线性方程组

 (i) $k = 1$, 取 $u = (1, 1, \cdots, 1)^{\mathrm{T}}$.

 $k \neq 1$ 时, $Lu = z_{k-1}$.

 (ii) $Ry_k = u$.

(4) $m_k = \max(y_k),\ z_k = y_k / m_k$.

(5) 若 $|m_k - m_{k-1}| < \delta$ 或 $\|z_k - z_{k-1}\| < \varepsilon$, 则停止计算,

取 $\lambda_m \approx \widetilde{\lambda}_m + \frac{1}{m_k}, \xi_m \approx z_k$. 否则, $k := k + 1$ 转 (3).

反幂法

例 4：已知矩阵

$$A = \begin{pmatrix} -1 & 2 & 1 \\ 2 & -4 & 1 \\ 1 & 1 & -6 \end{pmatrix}$$

在 -6.42 附近有一个特征值, 用反幂法求解该特征值和对应的特征向量 ξ_1, 误差限为 $\|z_{k+1} - z_k\| \leqslant 10^{-8}$.

反幂法

例 4：已知矩阵

$$A = \begin{pmatrix} -1 & 2 & 1 \\ 2 & -4 & 1 \\ 1 & 1 & -6 \end{pmatrix}$$

在 -6.42 附近有一个特征值, 用反幂法求解该特征值和对应的特征向量 ξ_1, 误差限为 $\|z_{k+1} - z_k\| \leqslant 10^{-8}$.

解　按照反幂法先对矩阵 $A + 6.42I$ 作三角分解, 结果如下

$$A + 6.42I = LU, \quad L = \begin{pmatrix} 1 & & \\ 0.36900369 & 1 & \\ 0.18450184 & 0.37514808 & 1 \end{pmatrix}$$

反幂法

$$U = \begin{pmatrix} 5.42 & 2 & 1 \\ & 1.68199262 & 0.63099631 \\ & & -0.00121890 \end{pmatrix}.$$

k	$\widetilde{\lambda}_m + \frac{1}{m_k}$	$(z_k)_1$	$(z_k)_2$	$(z_k)_3$
1	-6.42121890	-0.04602829	-0.37587276	1.00000000
2	-6.42106628	-0.04614571	-0.37492057	1.00000000
3	-6.42106661	-0.04614548	-0.37492113	1.00000000
4	-6.42106661	-0.04614548	-0.37492113	1.00000000

3. 特征值的其他计算方法

用于求矩阵的全部特征值与特征向量

(1) 基本 QR 方法

令 $A_0 = A$, 对 $K = 1, 2, \cdots$, 定义矩阵序列

$$\begin{cases} A_{k-1} = Q_{k-1}R_{k-1}, \\ A_k = R_{k-1}Q_{k-1}, \end{cases}$$

其中 Q_{k-1} 为正交矩阵, R_{k-1} 为上三角矩阵.

由于 A_k 相似于 A, 从而与 A 有相同的特征值.

QR 方法

Schur 分解　对一个给定的矩阵 $A \in \mathbb{R}^{n\times n}$, 存在一个正交矩阵 $Q \in \mathbb{R}^{n\times n}$ 使得

$$Q^{\mathrm{T}}AQ = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ & R_{22} & \cdots & R_{2n} \\ & & \ddots & \vdots \\ & & & R_{nn} \end{pmatrix},$$

其中对角子块 R_{ii} 为 1×1 或 2×2 矩阵.

QR 方法

基本收敛 当 $k \to \infty$ 时, A_k 的对角线元素均收敛, 其中严格下三角部分收敛到零, 则称A_k 基本收敛到上三角矩阵.

收敛结论 基本 QR 方法产生的序列 $\{A_k\}$ 基本收敛到分块上三角矩阵, 其对角块的特征值收敛到 A 的特征值. 具体地,若对角块为 1×1 的矩阵, 则收敛到 A 的特征值; 若对角块为 2×2 的矩阵, 其特征值为一对共轭复数, 收敛到 A 的一对共轭特征值.

QR 方法的每一步运算量为 $O(n^3)$. 为使更高效, 可先将 A 通过正交相似变换化为拟上三角 (上 Hessenberg) 形式, 更接近 Schur 形式, 再进行迭代, 可减少计算量.

(2) 带位移的 QR 方法 (加速方法)

令 $A_0 = A$, 对 $K = 1, 2, \cdots$, 定义矩阵序列

$$\begin{cases} A_{k-1} - \mu_{k-1} I = Q_{k-1} R_{k-1}, \\ A_k = R_{k-1} Q_{k-1} + \mu_{k-1} I, \end{cases}$$

其中 μ_{k-1} 为位移量, 通常取为 A 的某一特征值的近似. 实际计算中可取 A_{k-1} 的右下角元素 $a_{nn}^{(k-1)}$ 或右下角 2×2 矩阵的特征值中最接近 $a_{nn}^{(k-1)}$ 的.

Arnoldi 方法

用于求大型稀疏矩阵的特征值

在 m 维 Krylov 子空间里寻找特征向量的最佳逼近. 设 Krylov 子空间为

$$K_m = \text{span}\left\{z_0, Az_0, \cdots, A^{m-1}z_0\right\},$$

正交化过程得到正交向量 $z_1, z_2, \cdots, z_m$. 记 $Q_m = (z_1, \cdots, z_m)$, 则

$$AQ_m = Q_{m+1}H_{m+1,m} = Q_{m+1}\begin{pmatrix} H_m \\ h_{m+1,m}e_{m+1} \end{pmatrix}.$$

其中 $e_{m+1} = (0, \cdots, 0, 1) \in \mathbb{R}^{m+1}$. 如果 (λ, ξ) 是 H_m 的特征值和特征向量, 则 (λ, ξ) 接近 A 的特征值和特征向量.

Jacobi 方法

用于求实对称矩阵的全部特征值和特征向量

通过一系列正交相似变换, 将 A 化为对角矩阵

$$\Lambda = \mathrm{diag}\big(\lambda_1, \lambda_2, \cdots, \lambda_n\big),$$

即令

$$\begin{cases} A_0 = A, \\ A_k = R_k A_{k-1} R_k^{\mathrm{T}}, \quad R_k \text{ 为正交矩阵}, \quad k = 1, 2, \cdots \end{cases}$$

对于某个 m, A_m 为对角矩阵, 即

$$\Lambda = R_m R_{m-1} \cdots R_1 A R_1^{\mathrm{T}} R_2^{\mathrm{T}} \cdots R_m^{\mathrm{T}}.$$

记 $P_m = R_1^{\mathrm{T}} R_2^{\mathrm{T}} \cdots R_m^{\mathrm{T}}$, 则 P_m 为正交矩阵, 且

$$P_m^{\mathrm{T}} A P_m = \Lambda, \quad A P_m = P_m \Lambda.$$

记 P_m 的列向量为 $\xi_1, \xi_2, \cdots, \xi_n$, 则

$$A\xi_i = \lambda_1 \xi_1, \qquad i = 1, 2, \cdots, n$$

因此 Λ 的对角元为 A 的特征值, P_m 的列向量为 A 的特征向量.

R_k的取法: Givens 变换.

Givens 方法

用于求实对称三对角矩阵的特征值

$$A = \begin{pmatrix} d_1 & b_1 & & & & \\ b_1 & d_2 & b_2 & & & \\ & b_2 & d_3 & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & b_{n-2} & d_{n-1} & b_{n-1} \\ & & & & b_{n-1} & d_n \end{pmatrix}, \quad b_i \neq 0 \ (\forall i)$$

令 A_i 为 A 的 i 阶首主子式, 定义 Strum 序列

$$p_i(x) = \det(A_i - xI_i), \quad i = 1, 2, \cdots, n$$

Givens 方法

则有

$$
\begin{aligned}
p_0(x) &= 1, \\
p_1(x) &= d_1 - x, \\
p_2(x) &= (d_2 - x)(d_1 - x) - b_1^2 = (d_2 - x)p_1(x) - b_1^2 p_0(x), \\
&\ \vdots \\
p_i(x) &= (d_i - x)p_{i-1}(x) - b_{i-1}^2 p_{i-2}(x), \quad i = 3, 4, \cdots, n
\end{aligned}
$$

可证 $p_n(x)$ 是 A 的特征值多项式. 对任意给定的 $\mu \in \mathbb{R}$,

$$
S_\mu = \{p_0(\mu), p_1(\mu), \cdots, p_n(\mu)\}
$$

符号变化数 $S(\mu)$ 恰是 A 的小于 μ 的特征值的个数, 因此可用二分法确定特征值的近似值.

Lanczos 方法

用于求大型对称稀疏矩阵的最大和最小特征值

Krylov 子空间

$$\left\{z_0, Az_0, \cdots, A^{m-1}z_0\right\}$$

正交化过程产生三对角矩阵 H_m, 得到迭代序列 $\{H_m\}$.

H_m 的最大和最小特征值收敛于 A 的最大和最小特征值.

4. 奇异值和广义特征值问题

奇异值分解

定义 设 $A \in \mathbb{C}^{m\times n}$, 存在酉矩阵 $U \in \mathbb{C}^{m\times m}, V \in \mathbb{C}^{n\times n}$, 使得

$$U^{\mathrm{H}}AV = \begin{pmatrix} \Sigma_p & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{m\times n}, \qquad \Sigma_p = \mathrm{diag}\big(\sigma_1, \cdots, \sigma_p\big),$$

其中 $p = \min(m, n)$ 且 $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_p \geqslant 0$, 则上式称为 A 的奇异值分解 (SVD). σ_i 称为 A 的奇异值, V 的列 v_i 称为 A 的对应 σ_i 的右奇异向量, V 的列 v_i 称为 A 的对应 σ_i 的左奇异向量.

奇异值分解

定义 设 $A \in \mathbb{C}^{m\times n}$, 存在酉矩阵 $U \in \mathbb{C}^{m\times m}, V \in \mathbb{C}^{n\times n}$, 使得

$$U^{\mathrm{H}}AV = \begin{pmatrix} \Sigma_p & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{m\times n}, \qquad \Sigma_p = \mathrm{diag}\big(\sigma_1, \cdots, \sigma_p\big),$$

其中 $p = \min(m, n)$ 且 $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_p \geqslant 0$, 则上式称为 A 的奇异值分解 (SVD). σ_i 称为 A 的奇异值, V 的列 v_i 称为 A 的对应 σ_i 的右奇异向量, V 的列 v_i 称为 A 的对应 σ_i 的左奇异向量.

如果 A 为实矩阵, 则

$$\sigma_i(A) = \sqrt{\lambda_i(A^{\mathrm{T}}A)}, \quad i = 1, 2, \cdots, p.$$

A 的奇异值分解可通过计算 $A^{\mathrm{T}}A$ 的特征值和特征向量得到.

广义特征值问题

定义 1　设 $A, B \in \mathbb{C}^{n\times n}$ 为两个给定的矩阵, 则广义特征值问题定义为: 求 $\lambda \in \sigma(A, B)$ 和非零向量 $x \in \mathbb{C}^n$ 满足

$$Ax = \lambda Bx$$

其中广义特征值 $\sigma(A, B)$ 定义为

$$\sigma(A, B) = \left\{\lambda \middle| \operatorname{det}(A - \lambda B) = 0\right\}$$

当 $B = I$ 时, 即为标准的矩阵特征值问题.

定义 2　称 $p(\lambda) = \operatorname{det}(A - \lambda B)$ 为特征多项式, $\operatorname{det}(A - \lambda B) = 0$ 称为特征方程. 如果 $\operatorname{det}(A - \lambda B) \neq 0$, 则称 (A, B) 是正则的, 否则为奇异的.

广义特征值问题

(A, B) 有 n 个特征值的充要条件是 B 非奇异. 此时广义特征值问题转化为

$$Cx = \lambda x, \quad \text{其中 } C \text{ 满足 } BC = A$$

若 B 良态可用 QR 方法求解.

更实用的求解方法为基于广义 Schur 分解的 QZ 方法. 如果 A, B 对称且 B 正定, 则可用 QR-Cholesky 算法.

广义特征值与广义逆

本章总结

- 乘幂法
  - 按模最大的特征值
  - 加速方法: 原点平移法 (需大致了解) A 的特征值分布
- 反幂法
  - 按模最大的特征值: 对 A^{-1} 应用乘幂法
  - 原点平移反幂法: 修正近似特征值, 求相应的特征向量
- 其他方法
  - QR 方法: 全部特征值和特征向量
  - Arnoldi 方法: 大型稀疏矩阵的特征值
  - Jacobi 方法: 实对称矩阵的全部特征值和特征向量
  - Givens 方法: Strum 序列, 实对称三对角矩阵的特征值
  - Lanczos 方法: 大型稀疏对称矩阵的最小和最大特征值
- 奇异值问题: 可转化为 $A^{\mathrm{T}}A$ 的特征值问题
- 广义特征值问题: $Ax = \lambda Bx$