矩阵特征值问题

- 矩阵特征值与特征向量的计算
- 矩阵特征值=特征多项式方程的解 ⟹ 一定是迭代解法
- 内容
  - 基本概念与特征值的分布
  - 幂法与反幂法
  - 计算所有特征值的QR迭代算法

矩阵特征值有关概念

基本概念

- 矩阵$\boldsymbol{A}$的**特征值**与**特征向量**，($\boldsymbol{A}$为n×n方阵)

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}\,, \quad \boldsymbol{x} \neq \boldsymbol{0}$$

- 特征值λ是**特征方程**的根, 复数域内有n个(含重根)

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix} = 0$$

- **特征值谱** $\lambda(\boldsymbol{A})$: $\boldsymbol{A}$ 的全体特征值的集合
- 给定一特征值λ, **特征向量**是方程$(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$的**非零解**
- 对任一特征值, 特征向量都不唯一, 构成**特征子空间**
- 假设$\boldsymbol{A} \in \mathbb{R}^{n\times n}$
  - 特征值不一定是实数;
  - 实特征值一定可对应实特征向量;
  - 非实特征值的共轭也是特征值, 其对应的特征向量一定不是实向量.

例（根据定义计算特征值、特征向量）：求矩阵的特征值和特征向量。

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

【解】矩阵 $\boldsymbol{A}$ 的特征方程为

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

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

- 当 $\lambda=\lambda_1=3$ 时, 由 $(\lambda\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=0$, 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 $\boldsymbol{x}=[1,1,1]^{\mathrm{T}}$, 记为 $\boldsymbol{x}_1$, 则 $\boldsymbol{x}_1$ 是 λ_1 对应的一个特征向量。

- 当 $\lambda=\lambda_2=2$ 时, 由 $(\lambda\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$, 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 $\boldsymbol{x}=[1,1,2]^{\mathrm{T}}$, 记为 $\boldsymbol{x}_2$, 则 $\boldsymbol{x}_2$ 是 λ_2 对应的一个特征向量。

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

(1) $\sum_{j=1}^{n} \lambda_j = \sum_{j=1}^{n} a_{jj} = \operatorname{tr}(\boldsymbol{A})$ 。

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

$\operatorname{tr}(\boldsymbol{A})$ 表示矩阵对角线上元素之和, 称为矩阵的**迹 (trace)**。

从结论 (2) 也可以得出:

- 非奇异矩阵特征值均不为 0
- 0一定是奇异矩阵的特征值

特征值的有关性质 $\boldsymbol{Ax}=\lambda\boldsymbol{x}$， $\boldsymbol{x}\neq\boldsymbol{0}$

- 非奇异矩阵的特征值$\neq 0$; 0一定是奇异矩阵的特征值
- $\lambda(\boldsymbol{A})=\lambda(\boldsymbol{A}^T)$ $\boxed{\det(\lambda\boldsymbol{I}-\boldsymbol{A})=\det\left((\lambda\boldsymbol{I}-\boldsymbol{A})^T\right)}$
- 若$\boldsymbol{A}$为对角阵或上(下)三角阵, 则其特征值为其对角元
- 若$\boldsymbol{A}$为分块对角阵或分块上(下)三角阵(对角块为方阵), 则 $\lambda(\boldsymbol{A})=\bigcup_{j=1}^{m}\lambda\left(\boldsymbol{A}_{jj}\right)$

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

- 矩阵运算结果的特征值: 设λ_j为$\boldsymbol{A}$的特征值，则

$\lambda(c\boldsymbol{A})=\{c\lambda_j\}$; $\lambda(\boldsymbol{A}+c\boldsymbol{I})=\{\lambda_j+c\}$; $\lambda(\boldsymbol{A}^k)=\{\lambda_j^k\}$; $\lambda(\boldsymbol{A}^{-1})=\{\lambda_j^{-1}\}$;

矩阵的相似变换 (similarity transformation) 不改变特征值。

设矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为相似矩阵, 即存在非奇异矩阵 $\boldsymbol{X}$ 使得 $\boldsymbol{B}=\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}$, 则

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

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

特征值的有关性质

- 设$\boldsymbol{A} \in \mathbb{R}^{n\times n}$有m个(m≤n)不同的特征值$\tilde{\lambda}_1, \cdots, \tilde{\lambda}_m$, 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**, $\tilde{\lambda}_j$的特征子空间的维数是其**几何重数**.
- 设n阶实方阵$\boldsymbol{A}$的m个不同的特征值为$\tilde{\lambda}_1, \cdots, \tilde{\lambda}_m$, $\tilde{\lambda}_j$的代数重数为n_j, 几何重数为k_j, 则
  - $\sum_j n_j = n$;
  - $\forall j$, $n_j \geq k_j$
  - 不同特征值的特征向量线性无关, 所有特征子空间的$\sum_{j=1}^{m} k_j$个基形成一组线性无关向量 (可能少于n个)
  - 若$\forall j$, $n_j = k_j$, 这种矩阵$\boldsymbol{A}$为**非亏损阵**, 否则为**亏损阵**. 非亏损阵有n个特征向量构成全空间的基 ($\mathbb{C}^n$)

特征值的有关性质

- $\boldsymbol{A}$为非亏损阵 $\Leftrightarrow \exists \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X} = \boldsymbol{\Lambda}$, $\boldsymbol{\Lambda}$为对角阵

非奇异矩阵 $\boldsymbol{X} \in \mathbb{C}^{n\times n}$列向量为 n 个线性无关的特征向量

$$\boldsymbol{A} = \boldsymbol{X}\boldsymbol{\Lambda}\boldsymbol{X}^{-1} \quad \text{(特征值分解)}$$

- (**Jordan分解**):矩阵 $A \in \mathbb{R}^{n\times n}$

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

$p=\sum k_j$, k_j为特征值的几何重数;

$\tilde{\lambda}_j$对应k_j个约当块, 其阶数之和$= n_j$

亏损阵

$$\boldsymbol{J}=\begin{bmatrix} \boldsymbol{J}_1 & & \\ & \ddots & \\ & & \boldsymbol{J}_p \end{bmatrix}, \qquad J_s=\begin{bmatrix} \lambda_s & 1 & & \\ & \lambda_s & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_s \end{bmatrix}$$

(约当标准型)

$1 \leqslant s \leqslant p$, 称为**约当块**, 其对角线元素为矩阵 $\boldsymbol{A}$ 的特征值

Jordan矩阵

$$\begin{bmatrix} -2 & 1 & \\ & -2 & 1 \\ & & -2 \end{bmatrix}, \begin{bmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}, \begin{bmatrix} 1-i & 1 & \\ & 1-i & 1 \\ & & 1-i \end{bmatrix}$$

约当块

$$\begin{bmatrix} \boxed{1} & & \\ & \boxed{2} & \\ & & \boxed{1-i} \end{bmatrix}, \left[\begin{array}{cc:c} 2 & 1 & \\ & 2 & \\ \hdashline & & 3 \end{array}\right], \left[\begin{array}{cc:ccc} 1 & 1 & & & \\ & 1 & & & \\ \hdashline & & i & 1 & \\ & & & i & 1 \\ & & & & i \end{array}\right], \begin{bmatrix} \boxed{\begin{matrix} 1 & 1 \\ & 1 \end{matrix}} & & \\ & \boxed{-2} & \\ & & \boxed{\begin{matrix} -2 & 1 \\ & -2 \end{matrix}} \end{bmatrix}$$

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

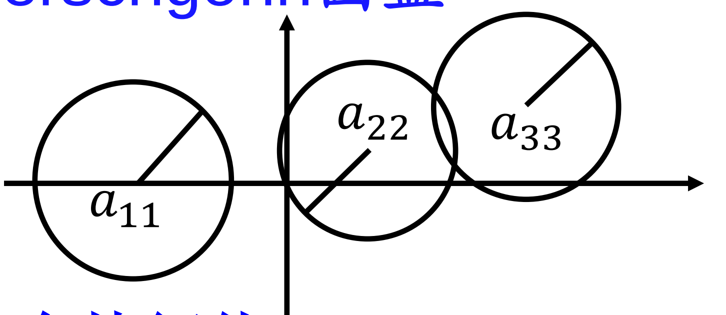
实对称阵为非亏损阵 (可正交对角化)

实对称矩阵$\boldsymbol{A} \in R^{n\times n}$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$\mathbf{\Lambda}$和由特征向量组成的正交阵$\mathbf{Q}$，使得

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathrm{T}}$$

特征值的分布估计

- 估计特征值的分布范围或它们的界有重要意义
  - 迭代法$\boldsymbol{x}^{(k+1)}=\boldsymbol{B}\boldsymbol{x}^{(k)}+\boldsymbol{f}$的收敛性, 看$\rho(\boldsymbol{B})=\max_j\left|\lambda_j(\boldsymbol{B})\right|$
  - 矩阵的2-条件数: $\mathrm{cond}(\boldsymbol{A})_2=\sqrt{\dfrac{\lambda_{\max}(\boldsymbol{A}^T\boldsymbol{A})}{\lambda_{\min}(\boldsymbol{A}^T\boldsymbol{A})}}$
- $\rho(\boldsymbol{A})\le\|\boldsymbol{A}\|_v$
- 定义5.4 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$, 设$r_k=\sum_{j=1,j\neq k}^{n}\left|a_{kj}\right|$, 在复平面上以$a_{kk}$为圆心、$r_k$为半径的圆, 称为$\boldsymbol{A}$的Gerschgorin圆盘
- Th5.10 (圆盘定理)
  - $\boldsymbol{A}$的特征值必在n个圆盘并集上
  - 若n个圆盘中有m个连通, 且与其他分离, 则这m个圆盘中正好含m个特征值



结论(1)的证明:

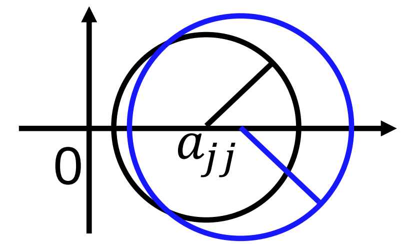
$\boldsymbol{Ax} = \lambda \boldsymbol{x}$, 设$\boldsymbol{x}$的第k个分量最大 $|x_k| = \max\limits_{1 \leqslant j \leqslant n} |x_j| > 0$,

$$\sum_{j=1}^{n} a_{kj} x_j = \lambda x_k, \qquad (\lambda - a_{kk}) x_k = \sum_{j \neq k} a_{kj} x_j$$

$$|\lambda - a_{kk}||x_k| = \left| \sum_{\substack{j=1 \\ j \neq k}}^{n} a_{kj} x_j \right| \leqslant \sum_{\substack{j=1 \\ j \neq k}}^{n} |a_{kj}||x_j| \leqslant |x_k| \sum_{\substack{j=1 \\ j \neq k}}^{n} |a_{kj}|$$

$$|\lambda - a_{kk}| \leq \sum_{j \neq k} |a_{kj}|$$

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。



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

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

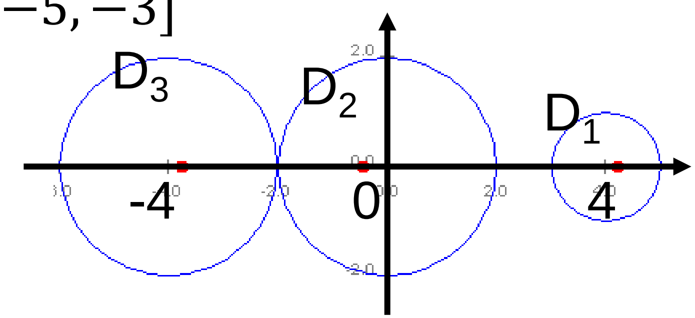
  (2) 若 $\boldsymbol{A}$ 为对角占优的对称矩阵,则 $\boldsymbol{A}$ 一定是对称半正定矩阵。若同时 $\boldsymbol{A}$ 非奇异,则 $\boldsymbol{A}$ 为对称正定矩阵。

圆盘定理的应用

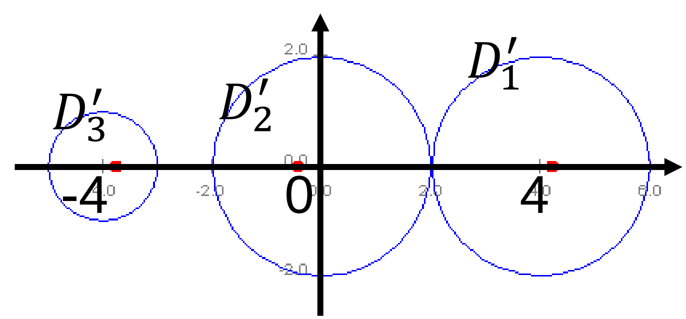
例: 估计矩阵$A=\begin{bmatrix}4 & 1 & 0\\1 & 0 & -1\\1 & 1 & -4\end{bmatrix}$的特征值分布，　$A^T=\begin{bmatrix}4 & 1 & 1\\1 & 0 & 1\\0 & -1 & -4\end{bmatrix}$

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）

D_3' 中存在一个特征值,且为实数,它属于区间 $[-5,-3]$



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



准确特征值为: 4.2030, -3.7601, -0.4429

幂法与反幂法

计算最大的特征值、特征向量

- 模最大的特征值称为主特征值, 也叫“第一特征值”,记为λ_1，它对应的特征向量称为主特征向量

  主特征值可能不唯一, 例如$5, -5, 3+4i, 3-4i$的模都是5

  谱半径$\neq$主特征值

- 幂法(power iteration): 取任意的非零向量$\boldsymbol{v}_0$, 计算$\boldsymbol{v}_k = \boldsymbol{A}\boldsymbol{v}_{k-1}, (k = 1,2,\cdots)$, 得到向量序列$\{\boldsymbol{v}_k\}$

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

(2) $\lim\limits_{k\to\infty} \frac{(v_{k+1})_j}{(v_k)_j} = \lambda_1$， j为$\boldsymbol{v}_k, \boldsymbol{v}_{k+1}$绝对值最大元素的下标

幂法

- **定理5.12** 若矩阵$\boldsymbol{A}$有唯一的主特征值λ_1 ($\neq$0), 取随机非零向量$\boldsymbol{v}_0$, 计算$\boldsymbol{v}_k=\boldsymbol{A}\boldsymbol{v}_{k-1}, (k=1,2,\cdots)$, 则:

$$k\to\infty,\quad \boldsymbol{v}_k\to\lambda_1\text{的某个特征向量}$$

证明: 考虑$\boldsymbol{A}$为非亏损阵的情况

注意: 需要$\boldsymbol{v}_0$对应$\alpha_1\neq 0$

设$\boldsymbol{A}$的线性无关的特征向量为$\hat{\boldsymbol{x}}_1,\cdots,\hat{\boldsymbol{x}}_n$, $\boldsymbol{v}_0=\alpha_1\hat{\boldsymbol{x}}_1+\cdots+\alpha_n\hat{\boldsymbol{x}}_n$

$$\boldsymbol{v}_k=\boldsymbol{A}^k\boldsymbol{v}_0=\alpha_1\lambda_1^k\hat{\boldsymbol{x}}_1+\cdots+\alpha_n\lambda_n^k\hat{\boldsymbol{x}}_n=\lambda_1^k\left[\sum_{j=1}^{s}\alpha_j\hat{\boldsymbol{x}}_j+\sum_{j=s+1}^{n}\alpha_j\left(\frac{\lambda_j}{\lambda_1}\right)^k\hat{\boldsymbol{x}}_j\right]$$

模<1

$\therefore\ \boldsymbol{v}_k$趋向于某主特征向量的倍数

(设λ_1为s重特征值)

若j为$\boldsymbol{v}_k$, $\boldsymbol{v}_{k+1}$绝对值最大元素的下标, $\lim\limits_{k\to\infty}\dfrac{(\boldsymbol{v}_{k+1})_j}{(\boldsymbol{v}_k)_j}=\lambda_1$

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

- 关于幂法的说明
  - □ $\boldsymbol{A}$为亏损矩阵的情况, 可以用矩阵的Jordan分解进行证明
  - □ $\{\boldsymbol{v}_k\}$相邻项的第j分量比值→主特征值, 选最大分量≠0
  - □ 在实际应用时随机选取$\boldsymbol{v}_0$, 避免出现特殊情况
- 幂法的问题

$$\boldsymbol{v}_k \approx \lambda_1^k \boldsymbol{x}_1 \ ,$$

k很大时, 可能出现上溢或下溢

$$\boldsymbol{v}_k = \lambda_1^k \left[\sum_{j=1}^{s} \alpha_j \hat{\boldsymbol{x}}_j + \sum_{j=s+1}^{n} \alpha_j \left(\frac{\lambda_j}{\lambda_1} \right)^k \hat{\boldsymbol{x}}_j \right],$$

收敛速度主要取决于$\left| \frac{\tilde{\lambda}_2}{\lambda_1} \right|$

$\tilde{\lambda}_2$为仅小于λ_1的特征值

- 实用的幂法: 用规格化向量的技术防止溢出

规格化向量

- 定义5.6 记$\overline{\max}(\boldsymbol{v})$为向量$\boldsymbol{v}$的绝对值最大分量, 若不唯一则取编号最小的那个. 称$\boldsymbol{u}=\boldsymbol{v}/\overline{\max}(\boldsymbol{v})$为向量$\boldsymbol{v}$的规格化向量

例: $\boldsymbol{v}=[3,-5,0]^T, \overline{\max}(\boldsymbol{v})=-5$, 规格化向量为$\boldsymbol{u}=\left[-\frac{3}{5},1,0\right]^T$

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

- 向量$\boldsymbol{v}_1,\boldsymbol{v}_2$的规格化向量分别为$\boldsymbol{u}_1,\boldsymbol{u}_2$, 若$\boldsymbol{v}_1=\alpha\boldsymbol{v}_2$, 则$\boldsymbol{u}_1=\boldsymbol{u}_2$

规格化也可以使用其他范数代替$\overline{\max}(\boldsymbol{v})$

实用的幂法

- 在幂法的每步增加向量规格化, 避免上溢或下溢

$$\boldsymbol{v}_1 = \boldsymbol{A}\boldsymbol{v}_0 \quad \Longrightarrow \quad \text{规格化向量}\boldsymbol{u}_1 = \frac{\boldsymbol{v}_1}{\overline{\max}(\boldsymbol{v}_1)} = \frac{\boldsymbol{A}\boldsymbol{v}_0}{\overline{\max}(\boldsymbol{A}\boldsymbol{v}_0)}$$

$$\boldsymbol{v}_2 = \boldsymbol{A}\boldsymbol{u}_1 = \frac{\boldsymbol{A}^2\boldsymbol{v}_0}{\overline{\max}(\boldsymbol{A}\boldsymbol{v}_0)} \Longrightarrow \text{规格化向量}\boldsymbol{u}_2 = \frac{\boldsymbol{v}_2}{\overline{\max}(\boldsymbol{v}_2)} = \frac{\boldsymbol{A}^2\boldsymbol{v}_0}{\overline{\max}(\boldsymbol{A}^2\boldsymbol{v}_0)}$$

$$\boldsymbol{v}_k = \boldsymbol{A}\boldsymbol{u}_{k-1} = \frac{\boldsymbol{A}^k\boldsymbol{v}_0}{\overline{\max}(\boldsymbol{A}^{k-1}\boldsymbol{v}_0)} \Longrightarrow \boldsymbol{u}_k = \frac{\boldsymbol{v}_k}{\overline{\max}(\boldsymbol{v}_k)} = \frac{\boldsymbol{A}^k\boldsymbol{v}_0}{\overline{\max}(\boldsymbol{A}^k\boldsymbol{v}_0)} \qquad k = 1, 2, \cdots$$

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

$\boldsymbol{x}_1$为某个主特征向量

实用的幂法

- 算法5.1:计算主特征值λ_1和主特征向量$\boldsymbol{x}_1$的实用幂法

输入： $\boldsymbol{A}$；输出： $\boldsymbol{x}_1, \lambda_1$.
$\boldsymbol{u} := $ 随机向量；
While 不满足判停准则 do
  $\boldsymbol{v} := \boldsymbol{A}\boldsymbol{u}$；
  $\lambda_1 := \overline{\max}(\boldsymbol{v})$； {主特征值近似值}
  $\boldsymbol{u} := \boldsymbol{v}/\lambda_1$； {规格化}
End
$\boldsymbol{x}_1 := \boldsymbol{u}$. {规格化的主特征向量}

每步的主要计算是算一次矩阵与向量乘法

对稀疏阵A很友好

- 可用相邻两步得到的λ_1之差作判停准则
- 规格化保证数值不上/下溢, 绝对值最大分量收敛到λ_1
- 前提条件: 主特征值唯一

迭代步数少时可不做规格化,
规格化也可改为$\boldsymbol{v}/\|\boldsymbol{v}\|$

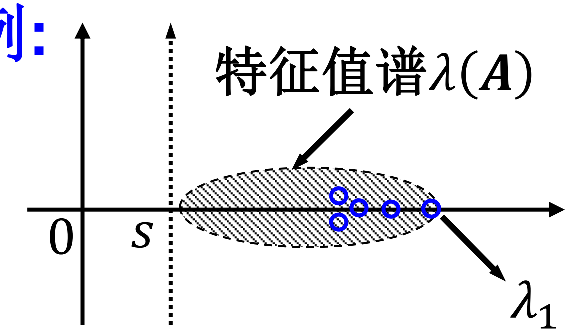
例：用实用的幂法求矩阵的主特征值，$\boldsymbol{A}=\begin{bmatrix}3 & 1\\1 & 3\end{bmatrix}$

取初始向量为 $\boldsymbol{v}_0=\boldsymbol{u}_0=[0 \quad 1]^{\mathrm{T}}$。迭代过程的计算结果：

k	$\boldsymbol{v}_k^{\mathrm{T}}$		$\overline{max}(\boldsymbol{v}_k)$	$\boldsymbol{u}_k^{\mathrm{T}}$	
0				0.000	1.0
1	1.0	3.000	3.000	0.333	1.0
2	2.0	3.333	3.333	0.600	1.0
3	2.800	3.600	3.600	0.778	1.0
4	3.333	3.778	3.778	0.882	1.0
5	3.647	3.882	3.882	0.939	1.0
6	3.818	3.939	3.939	0.969	1.0
7	3.908	3.969	3.969	0.984	1.0
8	3.953	3.984	3.984	0.992	1.0
9	3.977	3.992	3.992	0.996	1.0

加速幂法的收敛

例:



- 原点位移技术
  - $\boldsymbol{B}=\boldsymbol{A}-s\boldsymbol{I}$的特征值为$\boldsymbol{A}$的特征值$-s$
    对$\boldsymbol{B}$应用幂法可能加快收敛
  - 如图中$\boldsymbol{A}$的特征值分布，$\left|\frac{\tilde{\lambda}_2-s}{\lambda_1-s}\right|<\left|\frac{\tilde{\lambda}_2}{\lambda_1}\right|$, 更快收敛
- 瑞利商(Rayleigh quotient)加速
  - 实对称阵$\boldsymbol{A}$的瑞利商: $R(\boldsymbol{x})=\frac{\langle \boldsymbol{A}\boldsymbol{x},\boldsymbol{x}\rangle}{\langle \boldsymbol{x},\boldsymbol{x}\rangle}=\frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}$
  - 对实对称阵$\boldsymbol{A}$, $\lambda_n \le R(\boldsymbol{x}) \le \lambda_1$, $\lambda_1(\lambda_n)$为最大(最小)特征值;
    若$\boldsymbol{x}$为相应特征向量, 不等号变等号 “实对称阵可正交对角化”
  - 实对称阵的主特征值λ_1唯一, 则幂法中规格化向量$\boldsymbol{u}_k$满足
    $R(\boldsymbol{u}_k)=\lambda_1+O\left(\left(\tilde{\lambda}_2/\lambda_1\right)^{2k}\right)$ 注意:幂法, $\sim \lambda_1+O\left(\left(\tilde{\lambda}_2/\lambda_1\right)^{k}\right)$
  - 在幂法基础上, 每步对$\boldsymbol{u}_k$计算$R(\boldsymbol{u}_k)\approx\lambda_1$

例 对于前例中的问题, 采用瑞利商加速技术的幂法计算主特征值。

k	$\overline{max}(\boldsymbol{v}_k)$	$\boldsymbol{u}_k^{\mathrm{T}}$		$\boldsymbol{u}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u}_k / \boldsymbol{u}_k^{\mathrm{T}}\boldsymbol{u}_k$
0		0.000	1.0	3.000
1	3.000	0.333	1.0	3.600
2	3.333	0.600	1.0	3.882
3	3.600	0.778	1.0	3.969
4	3.778	0.882	1.0	3.992
5	3.882	0.939	1.0	3.998
6	3.939	0.969	1.0	4.000

反幂法

- A^{-1}的特征值是A特征值的倒数, 对A^{-1}应用幂法得最小特征值

算法5.2 算最小特征值/特征向量的反幂法
输入：A；输出：x_n, λ_n.
$\boldsymbol{u} := $ 随机向量;
While 不满足判停准则 do
  $\boldsymbol{v} := A^{-1}\boldsymbol{u}$； {求解线性方程组}
  $\lambda_n := 1/\overline{\max}(\boldsymbol{v})$； {最小特征值的近似值}
  $\boldsymbol{u} := \lambda_n \boldsymbol{v}$； {规格化}
End
$\boldsymbol{x}_n := \boldsymbol{u}$. {规格化的特征向量}

求解线性方程组, 计算量可能比幂法大很多

- 适用范围: A按模最小的特征值唯一
- 与原点位移技术结合: 若已知某个特征值$\lambda_j \approx p$, 则$\lambda_j - p$是$B = A - pI$按模最小的特征值, 对B使用反幂法

例 用反幂法求前例中矩阵的按模最小特征值。

k	$\boldsymbol{u}_k^{\mathrm{T}}$		$\overline{max}(\boldsymbol{v}_k)$
0	0.000	1.0	
1	-0.333	1.0	0.375
2	-0.600	1.0	0.417
3	-0.778	1.0	0.450
4	-0.882	1.0	0.472
5	-0.939	1.0	0.485
6	-0.969	1.0	0.492
7	-0.984	1.0	0.496
8	-0.992	1.0	0.498
9	-0.996	1.0	0.499

小结

- 实用幂法可能失败的情况
  - 矩阵A的主特征值不唯一，例如某个实矩阵，其模最大的特征值有不止一个数 (相反数、成对虚数)
  - 反幂法的情况类似
- 说明几点
  - 按幂法迭代计算，若前后两次迭代向量成比例，则它一定就是特征向量，也相应求出特征值
  - 加速幂法的方法还有Aitken外推等算法
  - 反幂法与位移技术结合，是很重要的技术

QR迭代算法

计算矩阵的所有特征值

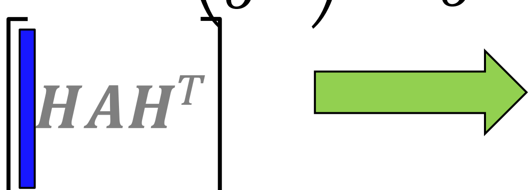
- 两个问题
  - 什么样的矩阵易于求所有特征值?
  - 对矩阵做怎样的变换能保持特征值不变?

$X^{-1}AX$

Q^TAQ

- 思路：用正交相似变换化矩阵为三角阵或分块三角阵
- 收缩技术(deflation)
  - 若已知A的一个特征向量x (比如通过幂法/反幂法)
  - 用正交变换将x消元: $Hx=\sigma e_1$, 则e_1是HAH^T的特征向量

$$HAH^Te_1=HA\left(\frac{1}{\sigma}x\right)=\frac{1}{\sigma}HAx=\frac{1}{\sigma}H\lambda x_1=\frac{\lambda}{\sigma}(\sigma e_1)=\lambda e_1$$



$$HAH^T=\begin{bmatrix}\lambda & r_1^T\\ \mathbf{0} & A_1\end{bmatrix}$$

问题变小!

计算矩阵的所有特征值

- 收缩技术的例子: 已知 $\boldsymbol{A}=\begin{bmatrix}3 & -1 & 1\\ 2 & 0 & 1\\ 1 & -1 & 2\end{bmatrix}$的一个特征值为 $\lambda_1=2$, 对应特征向量为$\boldsymbol{x}_1=[1,1,0]^T$, 求其他特征值

解: 用Householder变换对$\boldsymbol{x}_1$消元, 相应的$\sigma=\sqrt{2}=1.4142$

构造矩阵$\boldsymbol{H}$的$\boldsymbol{v}$向量为: $\boldsymbol{v}=\boldsymbol{x}_1+\sigma\boldsymbol{e}_1=\begin{bmatrix}2.4142\\ 1\\ 0\end{bmatrix}$

$$\Rightarrow \boldsymbol{H}=\begin{bmatrix}-0.7072 & -0.7072 & 0\\ -0.7072 & 0.7072 & 0\\ 0 & 0 & 1\end{bmatrix},\ \boldsymbol{HAH}^T=\begin{bmatrix}2 & 3 & -1.4142\\ 0 & 1 & 0\\ 0 & -1.4142 & 2\end{bmatrix}$$

$$\boldsymbol{A}_1=\begin{bmatrix}1 & 0\\ -1.4142 & 2\end{bmatrix}$$

易知$\boldsymbol{A}_1$的特征值为1, 2，得到$\boldsymbol{A}$的所有特征值

- 不形成$\boldsymbol{H}$计算$\boldsymbol{HAH}^T$: 先算$\boldsymbol{B}=\boldsymbol{HA}^T$, 再$\boldsymbol{HAH}^T=\boldsymbol{HB}^T$

用收缩技术+幂法效率不高，而且会误差累积，使结果不准。

QR迭代算法

1959~1961, John G.F. Francis和 Vera N. Kublanovskaya各自发明

- 理论基础
  - **Th5.21**(实**Schur**分解): $\forall \boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\exists$正交阵$\boldsymbol{Q}$使$\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{S}$, $\boldsymbol{S}$为**拟上三角阵**
  - **拟上三角阵**即**实Schur型:** 对角块为1阶或2阶的分块上三角阵

    2阶对角块的特征值是*A*的两个共轭复特征值
  - 思路：用一系列正交相似变换$\boldsymbol{B} = \boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{Q}$, 逐渐将$\boldsymbol{A}$化为上三角或对角块阶数$\leq$ **2**的分块上三角矩阵
- **QR算法** *("二十世纪十大算法"之一)*

$$\begin{cases} \text{对}\boldsymbol{A}_k\text{做QR分解: } \boldsymbol{A}_k = \boldsymbol{Q}_k\boldsymbol{R}_k \\ \boldsymbol{A}_{k+1} = \boldsymbol{R}_k\boldsymbol{Q}_k, (k = 0, 1, \cdots) \end{cases}$$

生成正交相似矩阵序列$\{\boldsymbol{A}_k\}$

$$\boldsymbol{A}_{k+1} = \boldsymbol{Q}_k^{\mathrm{T}}\boldsymbol{A}_k\boldsymbol{Q}_k$$

QR迭代算法

算法5.4: 计算矩阵特征值的QR算法

输入： $\boldsymbol{A}$；

输出： $\lambda_1,\cdots,\lambda_n$.

While $\boldsymbol{A}$不是拟上三角阵 **do**

计算$\boldsymbol{A}$的QR分解，得到矩阵$\boldsymbol{Q}$和$\boldsymbol{R}$;

$\boldsymbol{A} := \boldsymbol{RQ}$；

End

根据$\boldsymbol{A}$的对角块求特征值$\lambda_1,\cdots,\lambda_n$.

QR迭代算法

基本上是幂法收敛条件的推广
(否则, 对角块数可能>2)

- **Th5.22**: 收敛定理
  - 若矩阵A的等模的特征值为实重特征值或复共轭特征值, … , 则QR迭代所得矩阵序列$\{A_k\}$基本收敛于拟上三角阵

"基本收敛"

$$\begin{bmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & 0 & \times \end{bmatrix}$$

正交相似变换保对称性: 若A对称, 则$Q^T AQ$也对称

对实对称阵做QR迭代, 若满足Th5.22条件, 则极限为对角阵

实用的QR算法

- QR算法的不足之处
  - 每步迭代的计算量很大 *将矩阵化简为上Hessenberg阵*
  - 可能不收敛, 或收敛很慢 *带原点位移的**QR**迭代*
- 如果矩阵是上Hessenberg型 $\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \end{bmatrix}$
  - 对它执行QR迭代算法, 其中QR分解应用Givens旋转, 每步计算由$O(n^3)$降为$O\left(n^2\right)$

设$\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_{n-1}$为Givens旋转阵, $\boldsymbol{A}_k$为上Hessenberg阵

$$\boldsymbol{P}_{n-1} \cdots \boldsymbol{P}_2 \boldsymbol{P}_1 \boldsymbol{A}_k = \boldsymbol{R}_k \Longrightarrow \boldsymbol{Q}_k = (\boldsymbol{P}_{n-1} \cdots \boldsymbol{P}_2 \boldsymbol{P}_1)^T$$

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

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

$\boldsymbol{A}_{k+1}$仍是上Hessenberg阵

实用的QR算法

$Q^T AQ$

$$\begin{bmatrix} \times&\times&\times&\times&\times\\ \times&\times&\times&\times&\times\\ \times&\times&\times&\times&\times\\ \times&\times&\times&\times&\times\\ \times&\times&\times&\times&\times \end{bmatrix} \longrightarrow \begin{bmatrix} \times&\times&\times&\times&\times\\ \times&\times&\times&\times&\times\\ &\times&\times&\times&\times\\ &&\times&\times&\times\\ &&&\times&\times \end{bmatrix}$$

- 将$\boldsymbol{A}$正交相似约化为上Hessenberg
  - 用Householder变换

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & \boldsymbol{r}_1^T \\ \boldsymbol{c}_1 & \boldsymbol{A}_{22}^{(1)} \end{bmatrix} \xrightarrow{\boldsymbol{H}_1=\begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \boldsymbol{H}_1' \end{bmatrix}} \boldsymbol{H}_1\boldsymbol{A}=\begin{bmatrix} a_{11} & \boldsymbol{r}_1^T \\ \sigma_1\boldsymbol{e}_1 & \boldsymbol{H}_1'\boldsymbol{A}_{22}^{(1)} \end{bmatrix} \longrightarrow \boldsymbol{H}_1\boldsymbol{A}\boldsymbol{H}_1=\begin{bmatrix} a_{11} & \boldsymbol{r}_1^T\boldsymbol{H}_1' \\ \sigma_1\boldsymbol{e}_1 & \boldsymbol{H}_1'\boldsymbol{A}_{22}^{(1)}\boldsymbol{H}_1' \end{bmatrix}$$

$$\boldsymbol{A}^{(2)}=\left[\begin{array}{cc|cccc} a_{11} & a_{12}^{(2)} & a_{13}^{(2)} & \cdots & a_{1n}^{(2)} \\ \sigma_1 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ \hline 0 & a_{32}^{(2)} & a_{33}^{(2)} & \cdots & a_{3n}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(2)} & a_{n3}^{(2)} & \cdots & a_{nn}^{(2)} \end{array}\right] \xrightarrow{\boldsymbol{H}_2=\begin{bmatrix} \boldsymbol{I}_2 & \boldsymbol{O}^T \\ \boldsymbol{O} & \boldsymbol{H}_2' \end{bmatrix}} \boldsymbol{H}_2\boldsymbol{A}^{(2)}\boldsymbol{H}_2=\left[\begin{array}{cc|cccc} a_{11} & a_{12}^{(2)} & a_{13}^{(3)} & \cdots & \cdots & a_{1n}^{(3)} \\ \sigma_1 & a_{22}^{(2)} & a_{23}^{(3)} & \cdots & \cdots & a_{2n}^{(3)} \\ \hline 0 & \sigma_2 & a_{33}^{(3)} & \cdots & \cdots & a_{3n}^{(3)} \\ 0 & 0 & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & a_{n3}^{(3)} & \cdots & \cdots & a_{nn}^{(3)} \end{array}\right]$$

…… ……

  - 最终$\boldsymbol{H}_{n-2}\cdots\boldsymbol{H}_1\boldsymbol{A}\boldsymbol{H}_1\cdots\boldsymbol{H}_{n-2}$为上Hessenberg阵

实用的QR算法

- 实对称矩阵A

$$\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{bmatrix} \to \begin{bmatrix} \times & \times & & & \\ \times & \times & \times & & \\ & \times & \times & \times & \\ & & \times & \times & \times \\ & & & \times & \times \end{bmatrix} \to \begin{bmatrix} \times & & & & \\ & \times & & & \\ & & \times & & \\ & & & \times & \\ & & & & \times \end{bmatrix}$$

- 带原点位移的QR迭代(改善收敛)
  - 单位移技术 $\begin{cases} \boldsymbol{Q}_k\boldsymbol{R}_k = \boldsymbol{A}_k - s_k\boldsymbol{I}\ , & \text{(作QR分解)} \\ \boldsymbol{A}_{k+1} = \boldsymbol{R}_k\boldsymbol{Q}_k + s_k\boldsymbol{I}, & k = 0,1,2,\cdots \end{cases}$ 位移因子

$$\boldsymbol{A}_{k+1} = \boldsymbol{R}_k\boldsymbol{Q}_k + s_k\boldsymbol{I}$$
$$= \boldsymbol{Q}_k^T(\boldsymbol{A}_k - s_k\boldsymbol{I})\boldsymbol{Q}_k + s_k\boldsymbol{I} = \boldsymbol{Q}_k^T\boldsymbol{A}_k\boldsymbol{Q}_k \Longrightarrow \{\boldsymbol{A}_k\}\text{仍两两正交相似}$$

  - 简单策略: 取$s_k = \boldsymbol{A}_k(n,n)$, 加速收敛
  - 非对称阵有复特征值, 采用双位移